

## Contrôle continu : Evolution Moléculaire (EM7BMAAM) – Octobre 2011

### Questions de cours :

1. Donnez la définition des termes suivants. Accompagnez celles-ci d'un schéma :

- homologie
- orthologie
- paralogie
- xénologie

2. Vous devez réaliser un arbre phylogénétique à partir de séquences d'acides nucléiques d'ARNr 16S. Deux modèles évolutifs sont à votre disposition dans le logiciel que vous utilisez, le modèle de Jukes et Cantor et le modèle de Kimura deux paramètres. Donnez les hypothèses sous-jacentes communes à ces deux modèles et celle(s) qui leur est(ont) spécifique(s). Lequel utiliseriez vous ? Justifiez votre réponse.

3. Le tableau suivant montre la distribution de 4 caractères dans 4 organismes différents.

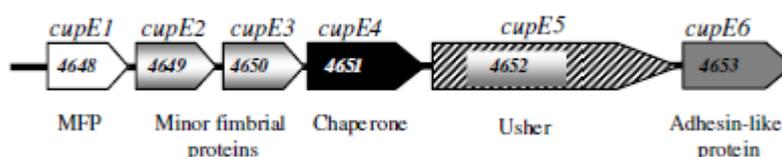
- Quelle méthode de reconstruction d'arbre utiliseriez vous pour établir les relations phylogénétiques existant entre ces espèces?
- Reconstituez cet arbre en expliquant les principes de la construction et en détaillant chaque étape.

|         |            | Caractères |                         |           |                |
|---------|------------|------------|-------------------------|-----------|----------------|
|         |            | Gésier (1) | Membrane nictitante (2) | Plume (3) | Sang chaud (4) |
| Espèces | Grenouille | 0          | 0                       | 0         | 0              |
|         | Chien      | 0          | 0                       | 0         | 1              |
|         | Alligator  | 1          | 1                       | 0         | 0              |
|         | Pie        | 1          | 1                       | 1         | 1              |

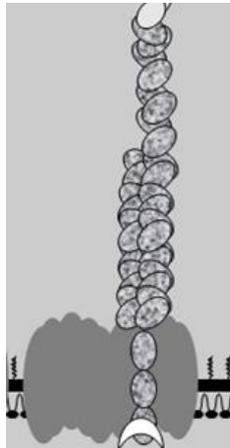
0 signifie que l'état de caractère n'est pas observé (donc ancestral) et 1 que l'état de caractère est observé (donc dérivé).

**Problème** (basé sur les travaux publiés dans Giraud *et al.*, 2011, Environ Microbiol, 13, 666-683)

La voie "chaperonne-usher" (CU) présente chez de nombreuses bactéries à gram négatif permet d'assembler différents types de fibres à leur surfaces et ces structures fibrillaires jouent un rôle important dans la pathogénicité de nombreuses bactéries. *Pseudomonas aeruginosa* possède sur son génome 4 à 5 copies de clusters de gènes codant pour des systèmes de la voie usher, les loci *cupA*, *cupB*, *cupC* et *cupE* présents sur tous les génomes des différentes souches et le locus *cupD* présents uniquement dans certaines souches. Nous nous intéresserons ici au locus *cupE* contenant 6 ORFs formant un opéron comme décrit ci-dessous.



Les trois premiers gènes *cupE1*, *cupE2* et *cupE3* codent pour des sous-unités pilines impliquées dans la formation de la fibre (Figure 1). Le gène *cupE6* code pour une adhésine-like qui se trouve localisée au sommet de la fibre. Le gène *cupE5* code pour le usher localisé à la membrane externe et qui permet la construction de la fibre. Le gène *cupE4* code pour une chaperone qui prend en charge, dans le périplasma, les pilines et l'adhésine pour les "amener" au usher.



**Figure 1 :** Le usher est représenté en gris foncé à la membrane externe (OM). Les sous-unités pilines composant la fibre sont représentées en marbré. L'adhésine est située au sommet de la fibre et la chaperone est représentée sous forme de croissant blanc au bas de la fibre.

Pour connaître la distribution taxonomique de ce nouveau système CU et étudier son évolution, nous nous proposons d'analyser l'ensemble des 1056 génomes bactériens complètement séquencés et de ne retenir que les systèmes qui ont une organisation génétique identique à celle du locus *cupE* de *P. aeruginosa*.

- Décrivez la démarche bioinformatique que vous adopteriez pour réaliser cette étude (de la construction des jeux de données à la construction des arbres).

Le résultat de votre démarche a conduit à l'obtention des quatre arbres présentés dans l'annexe I obtenus à partir des séquences protéiques des différents partenaires du système (adhésine, usher, chaperone, pilines). Une souche a été retenue par espèce. Un arbre de référence a été construit à partir des ARNr 16S des espèces dans lesquelles le locus *cupE* a été identifié pour mettre en évidence les relations évolutives entre ces espèces. (cf. légende des arbres Annexe I) :

- A quelle classe de méthodes appartient PhyML ?
- Quel est l'intérêt de calculer des valeurs de bootstrap ?
- Pourquoi, dans le cas des gènes *cupE1* à *cupE6*, les arbres ont-ils été réalisés à partir des séquences protéiques et non à partir des séquences d'acides nucléiques ?

En comparant les topologies de ces cinq arbres, répondez aux questions suivantes :

- Pour toutes les espèces étudiées, il a-t-il coévolution des différentes protéines impliquées dans le système ? Justifiez votre réponse.
- Si non, pouvez-vous identifier des groupes d'espèces pour lesquelles il y a coévolution et pour lesquelles la transmission des gènes du locus serait due à un héritage vertical (suite à des événements de spéciation).
- Proposez un scénario évolutif pour le cluster *cupE* de *P. aeruginosa*.

## Annexe 1

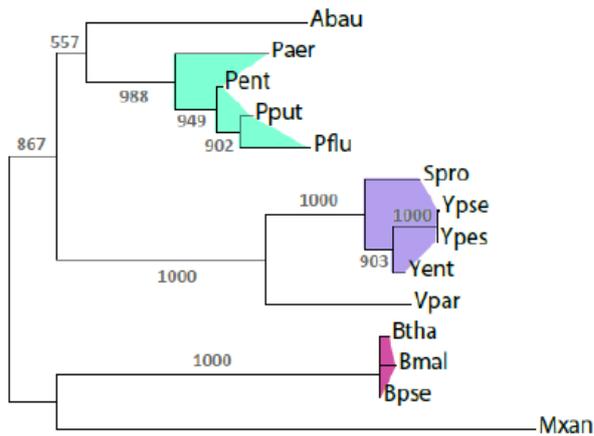


Figure A1 : Arbre phylogénétique de référence construit à partir des séquences d'ARNr 16S

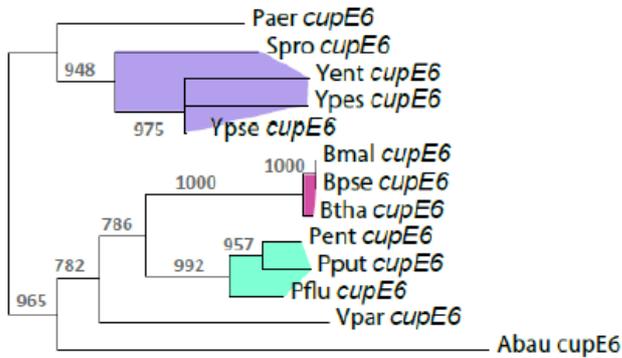


Figure A2 : Arbre obtenu sur les séquences protéiques d'adhésine

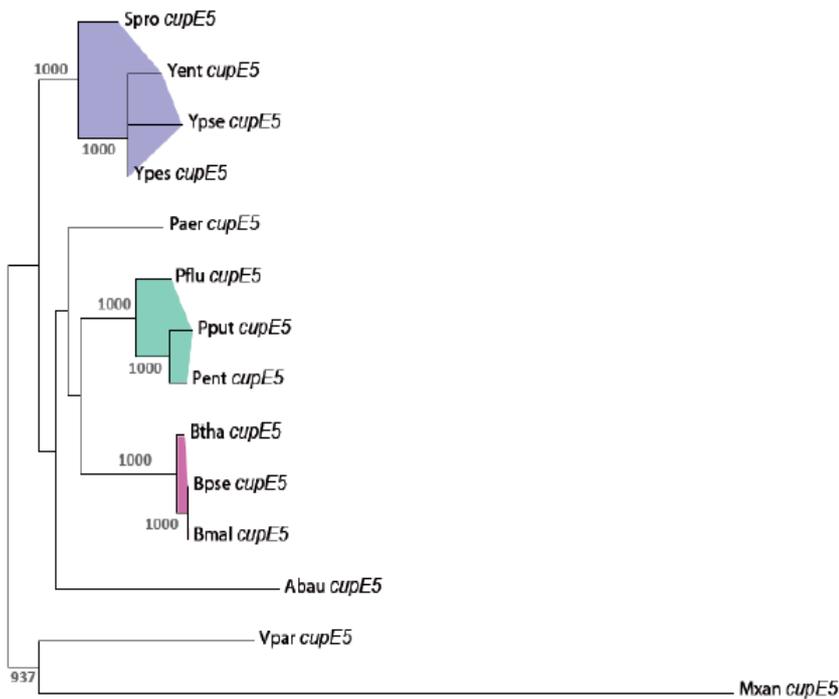


Figure A3 : Arbre obtenu sur les séquences protéiques du usher

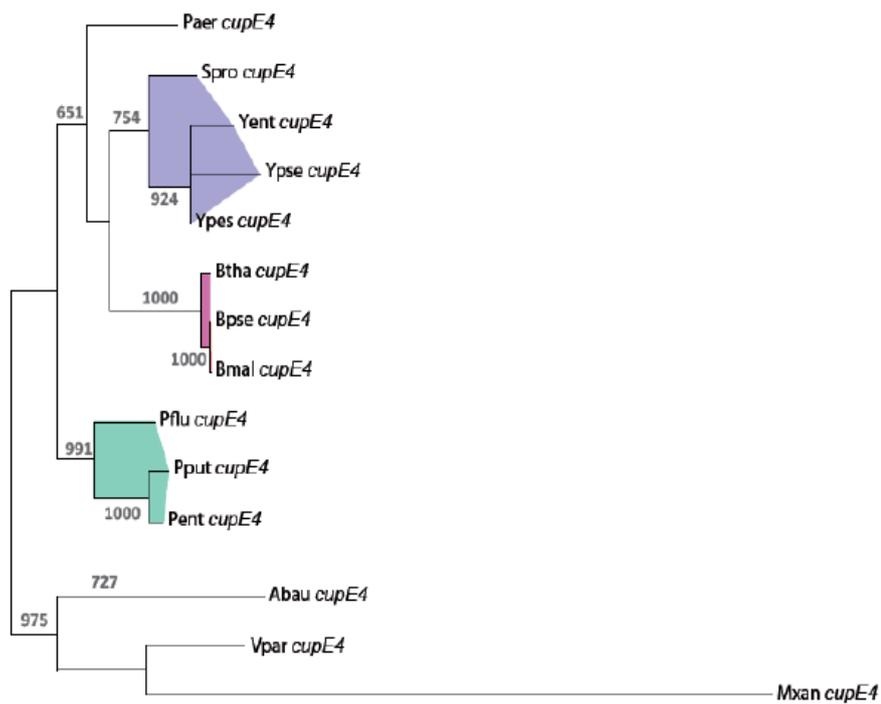


Figure A4 : Arbre obtenu à partir des séquences protéiques de la chaperone

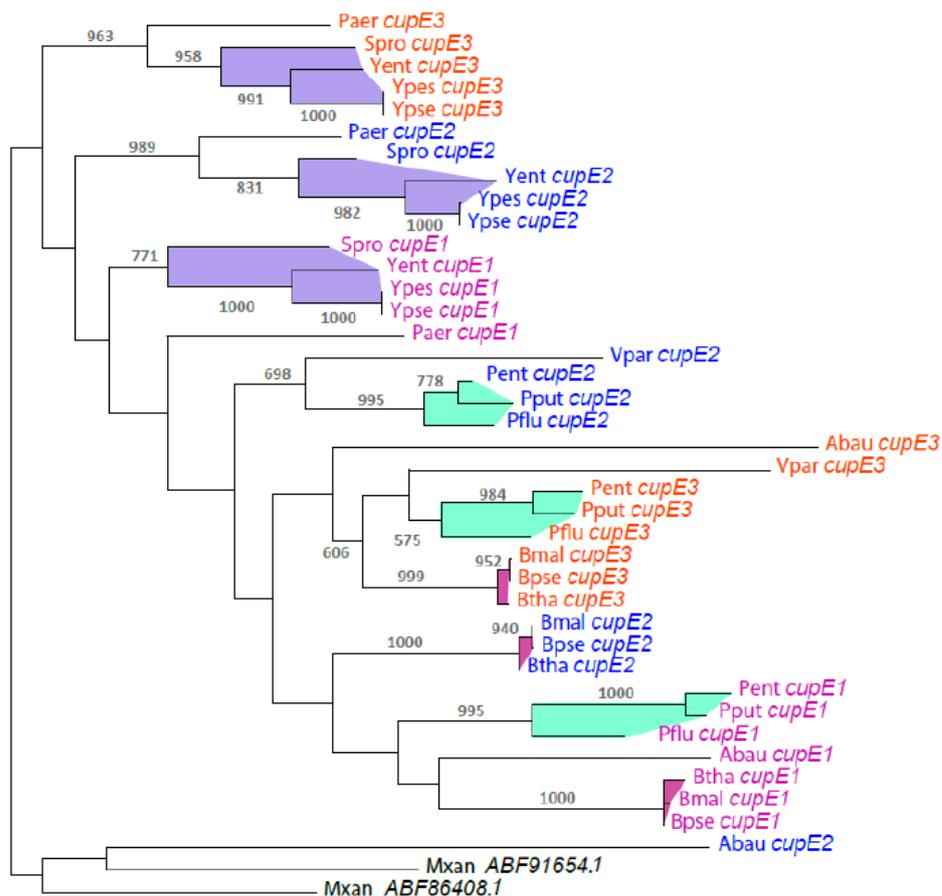


Figure A5 : Arbre obtenu à partir des séquences protéiques des pilines

## Légende communes aux cinq arbres.

Les arbres ont été réalisés avec PhyML et ne sont pas enracinés. Pour les séquences protéiques le modèle évolutif de substitution des acides aminés JTT a été utilisé avec 4 catégories de vitesse d'évolution de sites pour la correction par la loi Gamma. Pour les séquences d'ARNr 16S, le modèle évolutif de Tamura&Nei a été utilisé. 1000 bootstrap ont été réalisés et seules les valeurs supérieures à 500 ont été reportées sur les arbres.

Nomenclature pour les noms des espèces :

Paer : *Pseudomonas aeruginosa*; Pent : *Pseudomonas entomophila*; Pflu : *Pseudomonas fluorescens*; Pput : *Pseudomonas putida*; Spro : *Serratia proteamaculans*; Yent : *Yersinia enterocolitica*; Ypes : *Yersinia pestis*; Ypse : *Yersinia pseudotuberculosis*; Abau : *Acinetobacter baumannii*; Bmal : *Burkholderia mallei*; Bpse : *Burkholderia pseudomallei*; Btha : *Burkholderia thailandensis*;

Les systèmes présents dans *Vibrio parahaemolyticus* (Vpar) et *Myxococcus xanthus* (Mxan) ont été ajoutés à l'analyse malgré un nombre différent de gènes codant pour les pilines et l'absence de gène codant pour l'adhésine dans *M. xanthus* car ils avaient été identifiés et étudiés expérimentalement auparavant.

Les arbres ont été coloriés par rapport à la taxonomie : violet : Enterobacteriales (□ proteobacteria), vert : Pseudomonadaceae (□ proteobacteria), rose : Burkholderiaceae (□ proteobacteria).

*A. baumannii* (Moraxellaceae, □ proteobacteria), *V. parahaemolyticus* (Vibrionaceae, □ proteobacteria), *M. xanthus* (□ proteobacteria).