

Correction succincte du contrôle continu.

Pour les questions de cours, je vous renvoie à votre cours!

Problème:

Démarche bioinformatique:

- Utiliser chacun des produits protéiques codés par les gènes *cupE1* à *cupE6* comme sonde pour une recherche par similarité avec BlastP sur l'ensemble des protéines codées par les génomes complets bactériens disponibles.
- Pour chaque génome dans lequel six protéines homologues aux six protéines de *P. aeruginosa* ont été trouvées, vérifier que l'organisation génétique du système est conservée en analysant chacun des locus.
- Réaliser ensuite quatre alignements multiples (ushers, chaperones, adhésines, pilines).
- Rechercher le modèle évolutif le plus approprié pour construire l'arbre et ceci pour chaque famille de protéines (JModelTest)
- Construire les arbres correspondant en choisissant de préférence la méthode du maximum de vraisemblance PhyML ou alors une méthode de distance (NJ ou bioNJ).

PhyML appartient à la classe des méthodes du maximum de vraisemblance

Le calcul des valeurs de bootstrap sert à évaluer la topologie de l'arbre. C'est une méthode de rééchantillonnage (tirage aléatoire avec remise des positions de l'alignement pour construire un nouvel alignement de même longueur qui servira à construire un arbre). Elle teste individuellement la validité de chaque branche interne de l'arbre en calculant le pourcentage de fois où une branche de départ se retrouve dans les arbres construits par rééchantillonnage. Si la valeur du bootstrap est inférieure à 70%, la branche n'est pas fiable. Il faut alors interpréter la topologie avec précaution.

Les arbres ont été construits à partir de séquences protéiques et non nucléiques car lorsque les espèces sont distantes dans l'évolution, les séquences nucléiques peuvent avoir subi des substitutions multiples qui conduiront à une sous-estimation de leurs distances évolutives. On peut même dans certains cas avoir perdu le signal phylogénétique. Ceci est dû au petit alphabet de ces séquences (4 lettres, les 4 bases). On préfère donc travailler au niveau protéique.

Comparaison des 5 topologies d'arbres :

1) De façon globale, il n'y a pas congruence entre les 5 arbres, donc pas une co-évolution des différentes protéines dans toutes les espèces. Aucun des quatre arbres n'est congruent avec l'arbre réalisé sur les séquences d'ARNr 16S, laissant envisager une évolution complexe de ce système. Si on compare deux à deux les arbres, ceux du usher et de l'adhésine sont les plus congruents, à l'exception de la position de la séquence de *P. aeruginosa*.

2) Nous pouvons toutefois identifier des groupes d'espèces au sein desquels les séquences ont coévoluées et dont la présence du système dans le génome résulterait d'un héritage vertical à l'intérieur du groupe (du à la spéciation). Il y a trois groupes :

- le groupe des Enterobacterales (violet)
- le groupe des Pseudomonadaceae (vert) à l'exception de *P. aeruginosa*
- le groupe des Burkholderiaceae (rouge).

En effet, sur chacun des arbres les séquences correspondant à ces espèces sont toujours retrouvées regroupées ensemble sous un même nœud.

Nous pouvons cependant noter que si chaque système contient 3 pilines, et que l'existence de ces trois gènes semblent due à deux évènements de duplication, ces évènements se sont produits indépendamment dans le groupe Enterobacterales et les groupes Pseudomonadaceae et Burkholderiaceae. En effet, les séquences CupE1 et CupE3 des Pseudomonadaceae et Burkholderiaceae forment deux clusters (donc issus d'une même séquence ancêtre) qui se trouve sous un même nœud, donc une séquence ancêtre se serait dupliquée pour donner les pilines CupE1 et les pilines CupE3 dans un ancêtre commun aux Pseudomonadaceae et Burkholderiaceae. La position des séquences CupE2 de ces deux groupes est plus difficile à expliquer (duplication éventuelle de la séquence ancêtre à CupE1 puis perte de cette séquence dans les Pseudomonadaceae et remplacement par transfert horizontal est une possibilité). Les séquences des Enterobacterales n'étant pas regroupées avec les séquences des Pseudomonadaceae et des Burkholderiaceae ont subi des évènements de duplication indépendants de ceux décrits ci-dessus.

3) Scénario évolutif pour le système Cup de *P. aeruginosa*.

L'adhésine CupE6, la chaperone CupE4 et les deux pilines CupE2 et CupE3 de *P. aeruginosa* forment dans chacun des arbres un cluster avec les protéines du groupe des Enterobacterales (violet). La piline CupE1 ne se trouve regroupée avec aucune autre piline (faible valeur du bootstrap). La séquence du usher CupE5 se retrouve en groupe externe des deux clusters vert et rouge (faible valeur de bootstrap de la branche).

On peut donc supposer que *P. aeruginosa* a acquis le système Cup par transfert horizontal d'une espèce ancêtre commune aux espèces *S. proteamaculans* and *Yersinia*. Les gènes *cupE1* et *cupE5* auraient été remplacés par de nouveaux variants dont l'origine ne pas être établie avec les données actuelles.