

École Thématique MuFoPam

(Quimper - 14 – 18 mars 2016)

TP Phylogénie moléculaire

Ce TP a pour objectif de mettre en pratique les méthodes et concepts vus en cours en utilisant les défensines bêta humaines comme exemple.

Les défensines sont de petits peptides cationiques impliqués dans l'immunité innée. Elles font partie de la première ligne de défense contre les invasions de pathogènes. Il existe plusieurs familles de défensines, dont la famille bêta. Les membres de cette famille présentent un agencement particulier de cystéines et de liaisons intracellulaires. Ils sont produits principalement par les cellules épithéliales exposées. Ils présentent des propriétés antimicrobiennes et sont impliqués dans la signalisation. Les séquences protéiques des défensines bêta présentent une forte variabilité suggérant haut degré de spécialisation.

47 gènes codant pour des défensines bêta ont été identifiées chez l'homme (source : <http://www.genenames.org/genefamilies/DEFB>).

Ces gènes sont codés principalement sur le chromosome 8 (locus 8p23.1), sur le chromosome 20 (locus 20p13 et 20q11.21) et le chromosome 6 (locus 6p12.3).

I. Histoire évolutive des défensines bêta humaines

Télécharger le fichier HS_defensin_beta_nuc.fst via le lien http://www.frangun.org/HS_defensin_beta_nuc.fst. Ce fichier contient 46 séquences de gènes codant pour des défensines humaines.

1) Réaliser un alignement des séquences nucléiques avec CLUSTAL0. Traduire les séquences nucléiques en protéines. Refaire la même analyse en alignant les séquences avec MUSCLE. Quelles observations majeures pouvez-vous faire ?

2) Désaligner les séquences nucléiques. Traduire les séquences nucléiques en protéines. Aligner les séquences protéiques avec CLUSTAL0. Afficher les séquences nucléiques correspondantes. Refaire la même analyse avec MUSCLE. Quelles observations majeures pouvez vous faire ?

3) Nettoyer l'alignement obtenu avec MUSCLE à l'aide du logiciel GBLOCKS. Faites varier les paramètres et observez les conséquences au niveau des régions conservées. Enregistrer la sélection effectuée avec les paramètres les moins stringents de GBLOCKS.

4) Reconstruire l'arbre correspondant à vos séquences à l'aide de la méthode du maximum de parcimonie (randomize sequence order 5 times, décocher l'option ignore all gaps sites → traiter les gap comme des caractères inconnus). Combien d'arbres équiparcimonieux sont trouvés ? Combien d'événements évolutifs impliquent-ils ? Refaites l'analyse trois fois. Que remarquez-vous ?

Estimer la robustesse des regroupements obtenus à l'aide de la procédure de bootstrap non-paramétrique implémentée dans SeaView (même paramètre que pour l'analyse précédente + 100 répliqués de l'alignement original). Enregistrer l'arbre obtenu dans le menu file (save to trees menu). Qu'observez-vous ?

5) A l'aide du logiciel IQ-TREE vous allez identifier le modèle d'évolution le plus adapté. Connectez-vous au site <http://iqtree.cibiv.univie.ac.at/>. Sélectionnez l'onglet « Model selection ». Choisir les paramètres adaptés.

Quels sont les modèle le plus adapté à votre jeu de données (BIC criterion) ?

6) Reconstruire l'arbre correspondant à vos séquences à l'aide du modèle le plus approprié proposé par IQ-TREE et la méthode du maximum de vraisemblance. Demander une exploration de l'espace des arbres basée sur les NNI+SPR, optimiser le ratio TS/TV.

Qu'observez-vous ?

7) Comparez les deux arbres reconstruits. Que pouvez-vous en déduire concernant l'évolution de cette famille de gènes chez l'homme ?

II. Comparaison des défensines humaines avec leurs orthologues chez le rat.

Vous Télécharger le fichier de séquences HS_RN_defensin_beta_nuc.fst via le lien http://www.frangun.org/HS_RN_defensin_beta_nuc.fst.

8) Traduire les séquences nucléiques en protéines. Aligner les séquences protéiques avec MUSCLE. Afficher les séquences nucléiques correspondantes. Nettoyer l'alignement à l'aide du logiciel GBLOCKS et les paramètres les moins stringents. Enregistrer la sélection effectuée.

9) Reconstruire l'arbre correspondant à vos séquences à l'aide de la méthode du maximum de parcimonie (randomize sequence order 5 times, décocher l'option ignore all gaps sites → traiter les gap comme des caractères inconnus, 100 réplcats de bootstrap).

Combien d'arbres équiparcimonieux sont trouvés ? Combien d'événements évolutifs impliquent-ils ?

Analysez l'arbre obtenu. Combien de paires d'orthologues Homme/Rat pouvez-vous identifier ?

10) Reconstruire l'arbre correspondant à vos séquences à l'aide du modèle le plus approprié proposé par IQ-TREE et la méthode du maximum de vraisemblance. Demander une exploration de l'espace des arbres basée sur les NNI+SPR, optimiser le ratio TS/TV.

Analysez l'arbre obtenu. Combien de paires d'orthologues Homme/Rat pouvez-vous identifier ? Le résultat est-il en accord avec l'arbre inféré par la méthode du maximum de vraisemblance ?

11) Refaire l'analyse en maximum de vraisemblance en ne prenant pas en compte les séquences de pseudogènes et les séquences trop partielles.

Analysez l'arbre obtenu. Combien de paires d'orthologues Homme/Rat pouvez-vous identifier ? Le résultat est-il en accord avec les arbres précédents ?

Quelle est l'origine probable de ces gènes ? Définir les clusters de gènes ancestraux.