

Correction du TD2 Evolution moléculaire

Problème

5) Séquences SHC

Les séquences SHC bactériennes ont été ajoutées à l'analyse pour la construction de l'arbre sur les protéines ERG7 de manière à fournir un groupe externe. Ceci permet donc de connaître l'ancêtre commun (racine) des séquences ERG7 qui est alors localisé sur la branche qui relie le groupe externe aux autres séquences. D'autre part, ici le groupe externe est constitué de plusieurs espèces présentant des distances évolutives étalonnées, ce qui permet de « casser » les longues branches et donc de limiter l'impact de l'artéfact de l'attraction des longues branches auquel sont sujettes les méthodes de reconstruction d'arbre.

Sur les quatre génomes bactériens, trois possèdent un exemplaire de cette séquence SHC. Le génome de *Stigmatella aurantiaca* en est dépourvu.

6) Comparaison arbre ERG7 et arbre des espèces.

La protéine ERG7 étant indispensable à la synthèse des stéroïdes, les organismes qui en sont dépourvus ne peuvent donc pas réaliser cette biosynthèse. En comparant l'arbre obtenu sur ERG7 et celui donnant la phylogénie des eucaryotes, il est donc possible d'identifier les génomes dépourvus de ce gène. Si plusieurs génomes eucaryotes (feuilles de l'arbre des espèces) issus d'un ancêtre commun ne possèdent plus ce gène, c'est que celui-ci a été perdu sur la branche de l'arbre conduisant à cet ancêtre (lui-même ne possédait plus le gène). Les croix marron sur les branches de l'arbre des espèces ci-dessous indiquent les positions dans l'arbre des espèces où la capacité de synthèse des stéroïdes a été perdue.

7) Analyse de l'arbre des protéines ERG7

Duplications sur l'arbre ERG7:

- Une duplication qui a eu lieu dans le génome de *Trypanosoma cruzi*

- Duplications dans les génomes d'*Oriza sativa* et *Arabidopsis thaliana*

- 1^{ère} hypothèse : deux duplications indépendantes dans chacun des génomes

Etant donné la structure de l'arbre il apparaîtrait qu'une duplication se soit produite dans le génome d'*Oriza sativa* avec la séquence NP_001068103 qui a subi une accélération dans sa vitesse d'évolution (branche plus longue).

Concernant la duplication dans le génome d'*Arabidopsis thaliana* cela est moins clair car les deux séquences ne sont pas sous le même nœud ancêtre. Cependant, une branche a un faible bootstrap (54) et la séquence NP_190099 a également subi une accélération de sa vitesse d'évolution. Cette différence de vitesse pourrait expliquer son positionnement et comme la branche est peu supportée par les données, la topologie est à analyser avec précaution. Nous pouvons donc faire l'hypothèse que la duplication s'est bien produite dans le génome d'*Arabidopsis thaliana*.

- 2^{ème} hypothèse : une duplication dans le génome de l'ancêtre commun aux deux espèces (cf. Arbre1).

En effet, les deux branches menant aux deux séquences d'*Oriza sativa* ont également des faibles valeurs de bootstrap (63 et 67 respectivement). Donc avec là aussi la topologie est à analyser avec précaution. Donc étant donné ces faibles valeurs de bootstrap et le fait qu'une des deux copies évoluent plus vite dans chacun des génomes pouvant donc perturber la reconstruction phylogénétique, l'hypothèse d'une duplication du gène dans le génome ancêtre et son acquisition par spéciation ensuite (héritage vertical) ne peut pas être exclue.

- Duplication dans le génome d'*Aspergillus fumigatus*.

Dans ce dernier génome, il y a trois exemplaires de la protéine. Concernant la duplication dans le bas de l'arbre, étant donné la topologie de l'arbre, une copie d'*Aspergillus fumigatus* (XP_747936) externe et l'autre copie (XP_751627) possédant un ancêtre commun avec celle de *Neurospora crassa*, l'hypothèse la plus probable est une duplication du gène dans le génome de l'ancêtre commun aux deux espèces (voir Figure 1) suivi d'une perte d'une des deux copies dans *Neurospora crassa* (celle qui aurait été l'orthologue de XP_747936. Ici les branches ont des valeurs élevées de bootstrap (96 et 97).

Concernant la copie la plus externe XP_751356, deux hypothèses :

- duplication dans le génome de l'organisme ancêtre des champignons et perte d'une des deux copies dans toutes les espèces sauf dans *Aspergillus fumigatus*.

- acquisition de cette séquence par transfert horizontal à partir d'un autre champignon car il a été montré que des transferts horizontaux existent chez les champignons. Cependant ici, pas d'évidence très claire car position externe sur l'arbre quand on regarde le sous-arbre des séquences ERG7 de champignons.

Transferts horizontaux :

La localisation de la séquence bactérienne de *Stigmatella aurantiaca* avec les séquences eucaryotes est une indication de l'acquisition de cette séquence par la bactérie au travers d'un transfert horizontal d'une séquence provenant d'un génome eucaryote. De plus, l'identification de la séquence ERG7 dans un petit nombre de génomes bactériens (quatre) est aussi en faveur de l'acquisition de cette séquence par ces génomes via un transfert horizontal dont la source serait un génome eucaryote. En effet, autrement il faudrait supposer que la séquence du gène ERG7 était présente dans l'ensemble des génomes procaryotes, au moins ceux possédant SHC et qu'ensuite elle ait été perdue par la majorité de ces génomes excepté les quatre génomes en question. Ceci nécessiterait d'envisager un grand nombre de pertes indépendantes, ce qui n'est pas l'hypothèse la plus parcimonieuse et la position de la séquence de *S. aurantiaca* indique clairement l'acquisition horizontale du gène.

8) Les gènes eucaryotes codant pour la protéine ERG7 sont homologues car ils possèdent un ancêtre commun. Ils sont orthologues car ils ont été acquis par héritage vertical donc suite à la spéciation. Cependant dans les cas où des duplications ont eu lieu, ces gènes sont paralogues entre-eux (dans le même génome) et ils forment des groupes d'orthologues quand comparés aux autres espèces.

9) Protéines ERG5

La distribution taxonomique de la protéine ERG5 montre que son gène est identifié dans des génomes UNIKONTS comme par exemple *Saccharomyces cerevisiae* mais aussi dans des génomes BIKONTS comme *Arabidopsis Thaliana*. Le gène codant pour cette enzyme devait donc être présent dans le génome ancestral avant la séparation BIKONTS et UNIKONTS, c'est-à-dire dans le génome de LECA.

Ce gène a subi plusieurs pertes indiquées par des croix vertes sur les branches de la phylogénie des eucaryotes. Nous pouvons également observer des événements de duplication récents dans certains génomes car les feuilles des arbres correspondant aux gènes dupliqués sont regroupées sous un même nœud ancêtre. Un événement de duplication a eu lieu dans le génome d'*Oryza sativa* et dans celui de *Cyanidioschyzon merolae*. Le génome d'*Arabidopsis thaliana* a subi trois événements de duplication. Après le premier événement, chacun des deux gènes s'est de nouveau dupliqué.

