

Guillaume Lahaie  
LAHG04077707

INF4500  
Devoir 2

**Question 1 : Décrivez brièvement l'arbre que vous avez obtenu et sa ressemblance à votre connaissance de la réalité évolutive des espèces. (10 lignes maximum).**

L'arbre obtenu est enregistré dans le fichier devoir2\_tree.pdf. Il s'agit d'un phylogramme, ce qui est expliqué par l'utilisation de Neighbor Joining pour sa construction. J'y remarque surtout deux éléments qui confirment deux informations généralement véhiculée : premièrement, il y a une grande similarité génétique entre l'homme et les singes en général. On peut voir dans l'arbre obtenu que toutes les espèces de primates soient regroupées dans un sous-arbre. On peut aussi remarquer que la souris et le rat sont reliés dans un sous-arbre, ce à quoi on pourrait s'attendre, considérant les similarités morphologiques entre les deux espèces. On peut voir aussi que le poulet semble éloigner des espèces de l'arbre, ce qui est explicable car ce n'est un mammifère, alors que toutes les autres espèces en sont. Ce qui est plus surprenant dans l'arbre est la proximité entre le chien et la vache, et surtout le sous-arbre liant le tatou à l'éléphant.

**Question 2 : Étant donné l'arbre à la page suivante qui correspond à l'arbre consensus actuel des espèces concernés. Voyez-vous des divergences avec l'arbre que vous avez reconstruit ? Si des divergences sont observées, pourriez vous énumérer à votre connaissance les raisons de cette divergence. (10 lignes maximum)**

Afin de comparer les deux arbres plus facilement, j'ai produit, à l'aide de Archaeopteryx, un arbre circulaire (devoir2\_muscle\_curved.pdf). Je remarque une divergence entre l'arbre produit par MUSCLE et l'arbre donné : les sous-arbres souris-rat et éléphant-tatou sont échangé dans l'arbre. Selon l'arbre produit par MUSCLE, on semble obtenir comme résultat une distance phylogénique plus importante entre le rat et la souris et l'humain, alors que l'arbre donné semble plutôt montrer que cette distance est plus petite. De plus, on y voit aussi une topologie d'arbre différente : les sous-arbres sont reliés de façon différente. Par exempl, on remarque qu'il n'y a pas un grand écart entre le tatou et la souris. Ces divergences peuvent être causées par le choix des séquences utilisées pour l'alignement multiple. Il serait utile de produire un arbre à partir d'autres séquences orthologues pour ces espèces.

**Question 3 : Refaite les quatre étapes de l'inférence phylogénétique décrite plus haut en remplaçant seulement la méthode d'alignement par ClustalW. Arrivez vous au même arbre que pour Muscle? Si non! Décrivez les différences. D'après vous quel serait le meilleur arbre? (10 lignes maximum)**

On peut remarquer rapidement que les deux arbres ont des topologies différentes. J'ai d'ailleurs confirmé ce résultat à l'aide de Robinson-Foulds, qui calcule une distance de 4 entre les deux arbres. Toutefois, pour plusieurs espèces, on retrouve des valeurs de distance similaires (l'humain et les autres primates, le chien et la vache). La grande différence est dans le placement des espèces dans les deux arbres. L'arbre produit par l'alignement multiple de MUSCLE place les primates à l'intérieur de l'arbre, tandis que celui produit par ClustalW place les primates aux extrémités de l'arbre. On obtient aussi différents sous-arbres, par exemple le tatou est maintenant plus éloigné de l'éléphant. L'arbre créé par MUSCLE est probablement de meilleure qualité, surtout car l'algorithme de ClustalW est plus généraliste, il n'a pas de fonction spécifique, tandis que l'algorithme d'alignement multiple de MUSCLE semble produire des résultats plus précis (Nucleic Acids Res. 2004; 32(5): 1792–1797.)