

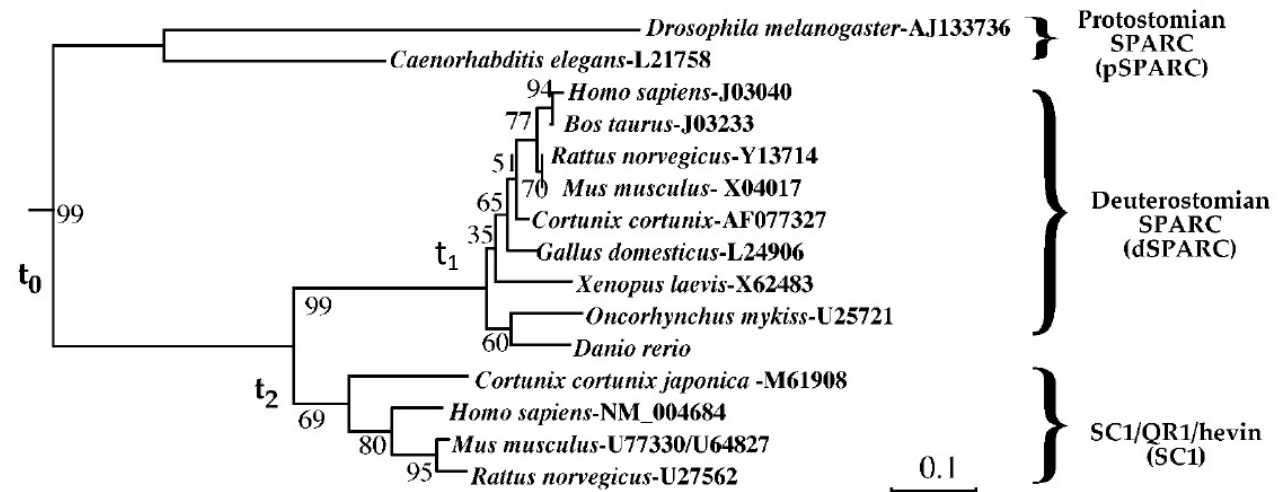
université
PARIS-SACLAY

FACULTÉ
DES SCIENCES
D'ORSAY

TP de biologie évolutive

Marc Girondot

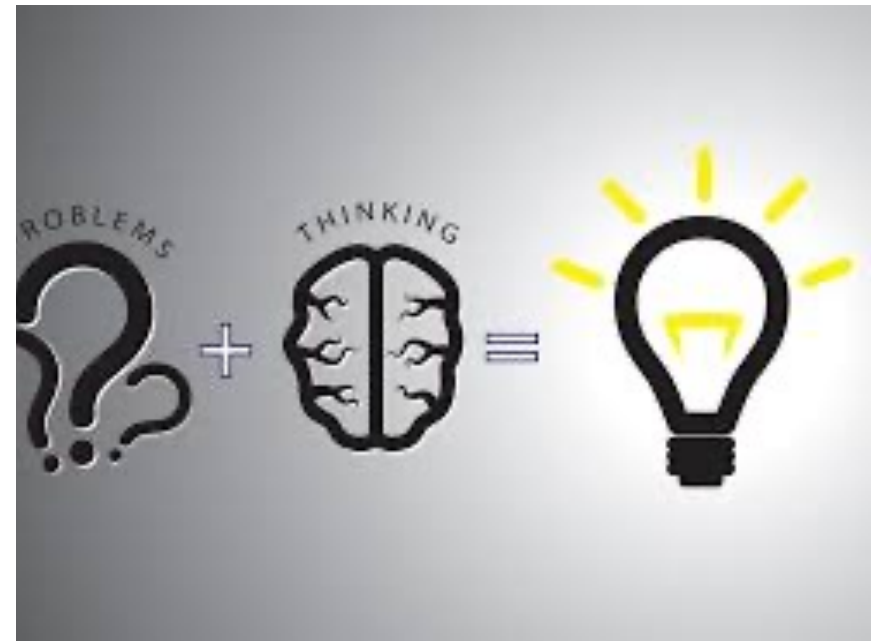
Problématique:
Retracer
l'évolution de
SPARC et SL1



Dans Delgado, S., Casane, D., Bonnaud, L., Laurin, M., Sire, J.-Y., Girondot, M., 2001. Molecular evidence for precambrian origin of amelogenin, the major protein of vertebrate enamel. *Molecular Biology and Evolution* 18, 2146-2153.

But du TP

- Le but du TP est de vous mettre dans une situation où vous cherchez vous-même les outils à utiliser pour répondre à une question: on ne vous prend pas la main; on est là seulement pour débloquer des situations qui vous empêcheraient d'avancer.



Procédure

- Trouver des séquences de SPARC et SPARC-Like 1 permettant de répondre à la problématique
- Aligner les séquences pour que les différentes positions dans la matrice des séquences soient homologues
- Produire une phylogénie
- Enraciner la phylogénie
- Discuter l'évolution des gènes SPARC et SPARC-Like 1
- Faire un compte-rendu d'une page (recto) donnant les principales informations sur les méthodes et le ou les résultats

Matériels et méthodes

- Genbank, GenBank is the NIH genetic sequence database
 - Interrogation par mots-clés ou par recherche de similitudes
- Blast, algorithme de recherche extrêmement rapide permettant de comparer une séquence avec l'ensemble des séquences disponibles dans Genbank
 - Blastn: séquence nucléotidique contre séquences nucléotidiques
 - Bastp: séquence protéique contre séquences protéiques
 - Bastx: séquence protéique contre séquences nucléotiques
 - Megablast: recherche de séquences très proches

Alignement

- Clustal-W: This method works by analyzing the sequences as a whole, then utilizing the UPGMA/Neighbor-joining method to generate a distance matrix. A guide tree is then calculated from the scores of the sequences in the matrix, then subsequently used to build the multiple sequence alignment by progressively aligning the sequences in order of similarity.
- T-Coffee (Tree-based Consistency Objective Function for Alignment Evaluation) is a multiple sequence alignment software using a progressive approach. It generates a library of pairwise alignments to guide the multiple sequence alignment.
- The MUSCLE algorithm proceeds in three stages: the draft progressive, improved progressive, and refinement stages. In the draft progressive stage, the algorithm produces a draft multiple alignment, emphasising speed over accuracy. In the improved progressive stage, the Kimura distance is used to reestimate the binary tree to create the draft alignment, in turn producing a more accurate multiple alignment. The final refinement stage refines the improved alignment made in step two.

Phylogénie

- ME: Minimum evolution is a distance method employed in phylogenetics modeling. It searches for the phylogeny that has the shortest total sum of branch lengths.
 - BioNJ
- ML: The maximum likelihood method uses standard statistical techniques for inferring probability distributions to assign probabilities to particular possible phylogenetic trees. The method requires a substitution model to assess the probability of particular mutations; roughly, a tree that requires more mutations at interior nodes to explain the observed phylogeny will be assessed as having a lower probability.
 - PhyML

Outils en ligne

- Vous trouverez ces outils en ligne à l'adresse:
<http://www.phylogeny.fr>
- Vous trouverez des informations intéressantes pour comprendre les algorithmes sur internet et en particulier en partant de Wikipedia:
https://en.wikipedia.org/wiki/Computational_phylogenetics

