

Algorithmique pour l'évolution des interactions géniques

2013-2014

Mots-clés Algorithmique, graphe, combinatoire, programmation dynamique, arbres phylogénétiques, évolution.

Description La connaissance de l'histoire évolutive des êtres vivants est une question cruciale dans la compréhension du fonctionnement des organismes actuels. Pour un gène donné présent dans plusieurs espèces, on peut reconstruire sous la forme d'un arbre enraciné l'histoire des événements ayant conduit à ces génomes actuels (spéciation, duplication, pertes, transferts). On s'intéresse ici à la reconstruction des relations d'adjacence entre les gènes. Une méthode est actuellement implémentée, appelée DeCo, qui partant d'un ensemble d'arbres de gènes et d'un ensemble d'adjacences actuelles entre ces gènes, permet de reconstruire les adjacences ancestrales.

La question que l'on cherche à résoudre dans ce stage consiste à déterminer dans quelle mesure cette méthode de reconstruction peut permettre d'inférer des informations sur les adjacences dans les génomes actuels partiellement ou incorrectement assemblés. En effet, dans ces génomes, certaines adjacences ne sont pas détectées non pas parce qu'elles n'existent pas, mais parce que le génome est fragmenté et que les gènes en question se trouvent dans deux fragments différents.

Le stage se déroulera donc de la façon suivante :

- Modifier le programme existant afin de tenir compte dans les coûts de programmation dynamique d'une information sur la qualité de l'assemblage (mesure de confiance sur les adjacences proposées ou l'absence d'adjacences). Ainsi, si une adjacence est "bien supportée" par le reste des espèces, elle ne sera pas pénalisée du fait qu'elle n'apparaît pas dans le génome mal assemblé.
- Récupérer les données d'adjacences inférées pour proposer un ordre de contigs dans l'espèce mal assemblée. On s'attachera notamment à conserver un caractère de linéarité dans le ou les génomes actuels dont les adjacences ont été complétées.
- Tester l'hypothèse sur des données réelles et proposer une méthodologie pour déterminer les paramètres optimaux de fonctionnement de la méthode (proportion de génomes actuels "de confiance", taux de qualité minimal pour espérer pouvoir reconstruire les adjacences actuelles).

Prérequis Intérêt pour la bioinformatique, Programmation

Contacts Sèverine Bérard (severine.berard@univ-montp2.fr), Vincent Berry (vincent.berry@lirmm.fr), Annie Chateau (annie.chateau@lirmm.fr).