

Stage de Master 2 recherche : Heuristiques efficaces pour l'échafaudage de génome

Annie Chateau - Rodolphe Giroudeau

2013–2014

En bioinformatique, le problème de **l'échafaudage de génomes** (ou scaffolding) consiste à déterminer, étant donné un ensemble de séquences d'ADN orientées appelées contigs, et un ensemble de relations valuées entre ces contigs, un **ordre et une orientation des contigs** qui maximise le poids total des relations entre ces contigs.

On peut modéliser ce problème sous forme d'un problème d'optimisation dans un graphe particulier, muni d'un couplage parfait initial, appelé graphe d'échafaudage. Le problème consiste alors à trouver un ensemble de cycles et chemins de poids maximum dans ce graphe.

Dans le cas où l'on cherche un seul cycle, on se retrouve avec une adaptation d'un *problème de voyageur de commerce*, version maximisante. Dans le cas où il n'y a pas de contrainte sur le nombre de cycles et chemins, le problème est polynomial. Mais dans ces deux cas extrêmes, la solution produite n'est pas forcément en adéquation avec la structure générale du génome (nombre et forme des chromosomes).

On veut donc à présent considérer le problème suivant : étant donné un graphe de contigs, et une signature sur la structure du génome en termes de cycles et chemins (par exemple, 2 chromosomes circulaires et 3 chromosomes linéaires), trouver une couverture des sommets qui maximise le poids et est conforme à la signature donnée. Ce problème est \mathcal{NP} -complet.

L'objectif du stage sera de mettre au point une ou plusieurs heuristiques pour ce problème, d'étudier leur complexité et leur qualité en terme d'approximation, ainsi que leurs performances et pertinence sur des données simulées et données biologiques réelles. On pourra également s'intéresser à une variante du problème, qui autorise la duplication de certaines portions du génome, et en étudier la complexité.

Prérequis Compétences en optimisation combinatoire, programmation

Contacts annie.chateau@lirmm.fr, rgirou@lirmm.fr