

Stage de Master 2 recherche : Approche par contraintes pour l'échafaudage de génomes avec répétitions

Annie Chateau - Rémi Coletta

2013–2014

Le *scaffolding* (ou échafaudage) de génome est une des étapes de la production d'un génome complet lors du séquençage de l'ADN. A ce stade, on dispose de séquences de longueurs variées, appelées *contigs*, et d'informations valuées sur l'orientation relative et la proximité de ces contigs. Le problème posé est de linéariser cet ensemble de relations entre les contigs de façon à obtenir le génome attendu, en tenant compte du poids des informations reliant les contigs. Il est possible de modéliser ce système sous forme de contraintes, notamment sous forme de *modélisation SAT*. Dans ce stage, on s'intéressera notamment aux questions suivantes :

- implémenter de façon efficace l'adaptation du modèle décisionnel existant en modèle adapté au problème d'optimisation,
- tester cette approche par rapport aux approches existantes, notamment en terme de passage à l'échelle et de pertinence biologique,
- étudier la structure des solutions pour inférer une détection de répétitions comme perturbations de la structure attendue, et mettre au point des stratégies de résolution de ces artéfacts,
- compléter la modélisation existante pour l'adapter au cas des contigs avec multiplicité (i.e. on s'intéresse au cas où un même contig peut apparaître plusieurs fois dans le génome), en testant au besoin plusieurs approches.

Prérequis Programmation par contraintes

Contacts annie.chateau@lirmm.fr, remi.coletta@lirmm.fr