



COMPUTATIONAL METHODS FOR DE NOVO ASSEMBLY OF NEXT-GENERATION GENOME SEQUENCING DATA

le 2 juillet 2012 16h00

ENS Rennes Amphi
[Plan d'accès](#)

**Soutenance de thèse de Rayan Chikhi (ENS Cachan - IRISA).
Spécialité Informatique**



Dans cette thèse, nous présentons des méthodes de calcul (modèles théoriques et algorithmiques) pour effectuer la reconstruction de séquences d'ADN. Il s'agit de l'assemblage de novo de génome à partir de lectures (courte séquences ADN) produites par des séquenceurs à haut débit. Ce problème est difficile, aussi bien en théorie qu'en pratique.

Du point de vue théorique, les génomes sont structurellement complexes. Chaque instance d'assemblage de novo doit faire face à des ambiguïtés de reconstruction. Les lectures peuvent conduire à un nombre exponentiel de reconstructions possibles, une seule étant correcte. Comme il est impossible de déterminer laquelle, une approximation fragmentée du génome est retournée.

Du point de vue pratique, les séquenceurs produisent un énorme volume de lectures, avec une redondance élevée. Une puissance de calcul importante est nécessaire pour traiter ces lectures. Le séquençage ADN évolue désormais vers des génomes et méta-génomes de plus en plus grands. Ceci renforce la nécessité de méthodes efficaces pour l'assemblage de novo.

Cette thèse présente de nouvelles contributions en informatique autour de l'assemblage de génomes. Ces contributions visent à incorporer plus d'information pour améliorer la qualité des résultats, et à traiter efficacement les données de séquençage afin de réduire la complexité du calcul. Plus précisément, nous proposons un nouvel algorithme pour quantifier la couverture maximale d'un génome atteignable par le séquençage, et nous appliquons cet algorithme à plusieurs génomes modèles. Nous formulons un ensemble de problèmes informatiques pour incorporer l'information des lectures pairées dans l'assemblage, et nous étudions leur complexité.

Cette thèse introduit la notion d'assemblage localisé, qui consiste à construire et parcourir un graphe d'assemblage partiel. Pour économiser l'utilisation de la mémoire, nous utilisons des structures de données optimisées spécifiquement pour la tâche d'assemblage. Ces notions sont implémentées dans un nouvel assembleur de novo, Monument.

Enfin, le dernier chapitre de cette thèse est consacré à des concepts d'assemblage dépassant l'assemblage de novo classique.

THÉMATIQUE(S)

Recherche - Valorisation

CONTACT

[Rayan Chikhi](#)

Mise à jour le 4 septembre 2015

ARCHIVES

[Séminaires 2020-2021](#)
[Séminaires 2019-2020](#)
[Séminaires 2018-2019](#)
[Séminaires 2017-2018](#)
[Séminaires 2016-2017](#)
[Séminaires 2015-2016](#)
[Séminaires 2014-2015](#)
[Séminaires 2013-2014](#)
[Séminaires 2012-2013](#)
[Séminaires 2011-2012](#)
[Séminaires 2010-2011](#)
[Séminaires 2009-2010](#)
[Séminaires 2008-2009](#)
[Séminaires 2007-2008](#)
[Séminaires 2006-2007](#)
[Séminaires 2005-2006](#)
[Séminaires 2004-2005](#)
[Séminaires 2003-2004](#)
[Séminaires 2002-2003](#)

JURY :

Éric RIVALS, DR, LIRMM / rapporteur
Sante GNERRE, Group leader, Broad Institute of MIT and Harvard / rapporteur
Marie-France SAGOT, DR, Inria Grenoble et Université de Lyon 1 / examinateur
Bertil SCHMI DT, Professeur, Johannes Gutenberg-Universität Mainz / examinateur
Olivier JAILLON, Chercheur, Genoscope-CNS (CEA) / examinateur

