

# Développement d'un pipeline d'assemblage et d'annotation de génome bactérien déployable sur le cloud.

Philippe Veber<sup>1</sup> et Stéphane Delmotte<sup>1</sup>

<sup>1</sup>Laboratoire de Biométrie et Biologie Évolutive

5 septembre 2017

Les techniques de séquençage haut-débit permettent aujourd'hui d'inspecter de multiples aspects des systèmes vivants comme leur génome, la transcription au sein des cellules ou encore l'état de la chromatine. S'il est bien souvent nécessaire pour analyser ces données de développer une approche spécifique au plan expérimental suivi (on pourrait parler d'analyse "à façon"), on retrouve néanmoins certaines situations de manière suffisamment récurrente pour suggérer l'intérêt d'automatiser les analyses dans ces cas-là.

Dans cette optique, le LBBE et le PRABI-AMSB ont conjointement commencé le développement de services d'analyse complètement automatisés, et l'objet de ce projet est de contribuer à cet effort, en développant un service d'assemblage et d'annotation de génome bactérien, accessible via un navigateur web. Un enjeu essentiel sera que le service puisse être déployé très simplement par un utilisateur. L'approche proposée est de rendre ce service disponible sur le cloud de l'Institut Français de Bioinformatique [1] et sur celui du LBBE.

Le projet comporte essentiellement 3 grandes parties :

- bibliographie pour déterminer les programmes pertinents disponibles
- développement et évaluation du pipeline sur données de référence
- programmes et scripts de déploiement sur cloud

L'écriture du pipeline s'appuiera sur la bibliothèque bistro [2] développée au LBBE. Cette bibliothèque est écrite dans le langage OCaml [3], mais

peut s'utiliser avec un sous-ensemble relativement réduit du langage. L'apprentissage des bases d'OCaml fera partie intégrante du travail attendu et s'appuiera notamment sur un cours d'introduction dispensé pour les participants. En complément, les plus motivés pourront envisager de participer au MOOC animé par l'Université Paris-Diderot [4].

Le sujet proposé permettra d'aborder des sujets variés : assemblage et annotation de génomes, développement de pipelines bioinformatiques, machines virtuelles et cloud, conteneurs Docker, programmation fonctionnelle, suivi de version avec git, voire un peu de programmation web. Il est adapté pour des groupes de 3 à 5 personnes (motivées), et permettra de contribuer à la mise en place d'un service en production.

Pour toute question, envoyez un message à :

`philippe.veber@univ-lyon1.fr`

## Références

- [1] <https://www.france-bioinformatique.fr/>
- [2] <https://github.com/pveber/bistro>
- [3] <http://ocaml.org/>
- [4] <https://www.fun-mooc.fr/courses/parisdiderot/56002S02/session02/about>