

## Compte-Rendu rapide de la réunion du 6 novembre 2014

### sur le projet ABDC

*Présents* : Matthieu BARRET, Tristan BOUREAU, Jérôme BOURSIER, Benoit DAMOTA, Gilles HUNAULT, Emmanuel JASPARD, David LESAIN, Vincent VIGNERON, Frédéric LARDEUX (au début).

*Excusé* : Frédéric SAUBION.

La réunion s'est déroulée en deux parties : 3 exposés puis une discussion. Aucune date de prochaine réunion n'a été fixée pour l'instant, mais nous prévoyons une seconde rencontre en mai ou juin 2015 en fonction de l'avancement des premiers travaux, avec peut-être une première réunion (partielle) sur les descripteurs de séquences d'ADN mi-décembre. Les textes des présentations et ce CR sont disponibles à l'adresse

<http://forge.info.univ-angers.fr/~gh/Abdc/>

### 1. Les trois exposés

Matthieu a présenté la métagénomique comme l'analyse des génomes des micro-organismes issus d'un environnement, avec les problèmes posés par la technologie de séquençage : fracturation des séquences d'ADN des micro-organismes de départ, mélange des fragments obtenus. Si l'assemblage des fragments est encore difficile, des méthodes et des logiciels existent et nous allons plutôt nous focaliser sur le regroupement ou « binning » des fragments.

Ensuite, Jérôme a présenté ce que sont la stéatopathie et la NAFLD (Non-Alcoholic Fatty Liver Disease) et comment l'analyse métagénomique du microbiote intestinal s'insère dans l'étude exhaustive de la pathologie. Un article de Qin qui a été publié dans Nature a déjà montré le bien-fondé d'une telle approche.

Enfin, Gilles a essayé de présenter les grandes lignes d'un futur projet collaboratif avec la mise en place d'outils pratiques nommés « descripteurs de séquences » et « descripteurs de classes » après avoir redéfini brièvement les notions d'alphabet, séquence, classe, en insistant sur l'originalité de l'approche séquences+descripteurs par rapport à une approche tableau+distances.

## 2. La discussion

Un point abordé par David fut la notion de distance et son utilité. Il ne s'agit pas d'abandonner les notions de distance et d'alignement, mais de les reléguer en fin de chaîne, pour ne pas « exploser » en termes de calculs.

Un autre point était la notion de réalisation pratique. Benoit et Matthieu devraient rapidement se mettre d'accord sur des jeux de données éprouvées de séquences d'ADN pour étayer une « preuve de concept ».

Concernant l'applicabilité aux images, évoqué par Etienne, il faudrait y penser plus en détail (mais sans doute un peu plus tard) car imaginer des alphabets de formes 2D, 3D ou de motifs couplés aux données phénotypiques et génomiques paraît très faisable.

Pour les données à venir de Jérôme, la question principale est celle du financement du séquençage.

Gilles a indiqué qu'il faudrait sans doute constituer une base de données de descripteurs avec des métadonnées pour savoir comment les utiliser (dans quel contexte, pour quelle approche...).

Emmanuel a soulevé la question des rapports entre l'ensemble des analyses de Jérôme sur la NAFLD et des séquences protéiques liées à la pathologie, notamment l'infection. Encore une piste à approfondir.

Concernant la notion de « projet » et de financement, il est beaucoup trop tôt pour y penser. Attendons d'avoir déjà réalisé quelque chose.

GH, Angers, le 7 novembre 2014.