

Stage de Master 2 sur plateforme technologique

Développement des méthodes d'analyses bioinformatiques de données ribo-seq : ARIBO

Mise en application au travers de l'étude de la traduction d'une cellule infectée par un virus géant.

Plateforme : PACA Bioinfo (Laboratoire Information Génomique et Structurale)

Tuteur : Sébastien Santini

Adresse : Parc Scientifique de Luminy. 13288 Marseille, France

Téléphone : 04 13 94 67 85

Email : santini@igs.cnrs-mrs.fr

Projet de stage

Depuis la caractérisation de mimivirus en 2003 (1) plusieurs virus géants ont été découverts donnant lieu à la création de diverses familles (2–5). Parmi elles, celle des *Pandoraviridae* regroupe certains des plus gros individus avec une particule de 1.2 μm , un génome d'ADN double brin d'environ 2.5Mb contenant presque 2000 gènes dont 85% sont de fonction inconnue (2). L'étude de ces génomes a montré que beaucoup de ces gènes pourraient correspondre à des gènes créés *de novo* à partir de régions non géniques (6, 7).

Le laboratoire IGS tente de comprendre comment les *Pandoraviridae* créent de nouveaux gènes et l'impact sur leur évolution et leur physiologie (projet PandoNovo ANR-22-CE12-0041). Une des étapes consiste à comprendre comment et quand ces gènes sont traduits lors du cycle infectieux.

Pour cela le laboratoire s'appuie sur l'analyse bioinformatique à grande échelle du « traductome » du virus et de son hôte grâce à deux méthodes : le profilage ribosomique (ribo-seq) et la protéomique par spectrométrie de masse à haut débit. Pour l'approche ribo-seq, 18 échantillons ont été séquencés (6 points du cycle infectieux prélevés en triplicats). Brièvement, les ARNm ont été isolés avec leurs ribosomes, puis digérés en conservant les zones protégées par ces derniers. Ces séquences préservées ont été séquencées par la technologie Illumina.

Le profilage ribosomique des ARNm au cours du cycle infectieux permettra de cartographier l'occupation des ribosomes et de caractériser en même temps la traduction des gènes de l'hôte et du virus géant (les gènes créés *de novo* mais également les gènes dits « cœurs »). Ceci permettra aussi de mettre en évidence les éléments régulateurs de la traduction. A plus long terme, l'analyse de ce type de données pourra être étendue à d'autres familles de virus géants présentant un cycle de réplication différent (cycle dit « strictement cytoplasmique » et les résultats comparés à ceux obtenus lors de ce stage (cycle « nucléocytoplasmique »).

Au sein de la plateforme PACA-Bioinfo, le/la candidat(e) recruté(e) aura pour objectif de mettre en place les techniques de traitement et d'analyses bioinformatique de données de « ribo-seq » au travers d'une étude ciblée d'un virus géant et de son hôte. Il/Elle évoluera dans un environnement naturellement interdisciplinaire où il sera nécessaire de comprendre la biologie des organismes étudiés (ici le virus et son hôte) mais aussi les aspects informatiques pour la manipulation des données dans un contexte hautement technique (cluster de calcul). Il/Elle identifiera les programmes les plus performants et les mieux adaptés pour cette étude. Il/Elle les mettra en œuvre sur les données disponibles et si possible fournira un pipeline de traitement transposable à d'autres expériences. Dans une démarche de type projet, généralisable à d'autres environnements professionnels, il/elle mènera

l'analyse des résultats et les présentera au responsable scientifique. Il/Elle suivra et participera à l'intégration du nouveau type de données dans l'écosystème du laboratoire en adéquation avec les principes FAIR de science ouverte.

Profil attendu du candidat

Le/La candidat(e) aura de préférence un parcours intégrant de la bioinformatique et/ou de la bio-analyse ou un fort attrait pour l'une de ces disciplines. Il/Elle devra être autonome et montrer un réel intérêt pour les aspects de développement informatique à visée bio-analyses. Cette offre cible en particulier le Master de biologie structurale et génomique (BSG-GAD) et le Master de bioinformatique (DLAD).

Pièces à fournir :

- CV
- relevé de notes et classements de la Licence 3
- relevé de notes et classements du Master 1
- lettre de motivation
- lettre de recommandation

Bibliographie

1. Scola BL, Audic S, Robert C, Jungang L, Lamballerie X de, Drancourt M, Birtles R, Claverie J-M, Raoult D. 2003. A Giant Virus in Amoebae. *Science* 299:2033–2033.
2. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie J-M, Abergel C. 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341:281–286.
3. Abergel C, Legendre M, Claverie J-M. 2015. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol Rev* 39:779–796.
4. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C, Couté Y, Rivkina E, Abergel C, Claverie J-M. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci U S A* 111:4274–4279.
5. Deeg CM, Chow C-ET, Suttle CA. 2018. The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. *eLife* 7:e33014.
6. Legendre M, Fabre E, Poirot O, Jeudy S, Lartigue A, Alempic J-M, Beucher L, Philippe N, Bertaux L, Christoforou E, Labadie K, Couté Y, Abergel C, Claverie J-M. 2018. Diversity and evolution of the emerging Pandoraviridae family. *Nat Commun* 9:2285.
7. Claverie J-M, Abergel C, Legendre M. 2018. [Giant viruses that create their own genes]. *Med Sci MS* 34:1087–1091.