

# Sujets de stage pour Master 2 en Bioinformatique au Genoscope

Contact : [stage\\_rdbioseq@genoscope.cns.fr](mailto:stage_rdbioseq@genoscope.cns.fr)

## Annotation de génomes de tiques

Dans le cadre d'un projet France Génomique ([france-genomique.org](http://france-genomique.org)), le Genoscope réalise le séquençage de plusieurs génomes de tiques, dont *Ixodes ricinus*, vecteur de multiples pathogènes d'animaux vertébrés, comme la bactérie *Borrelia burgdorferi* responsable de la maladie de Lyme chez l'humain. L'analyse de ces génomes va permettre d'apporter un éclairage à l'échelle moléculaire de la diversité génétique de ces espèces, ainsi que sur la relation avec leurs pathogènes. L'approche choisie fait appel à plusieurs technologies allant du séquençage sur Illumina HiSeq4000 à l'exploitation de bibliothèques Hi-C. La grande taille de ces génomes (environ 2Gb) et leurs hétérozygoties importantes représentent un challenge supplémentaire dans l'obtention de leurs séquences. Dans le but de procéder à des analyses de niveaux d'expressions des gènes d'*Ixodes ricinus*, les transcriptomes de plusieurs tissus de cette espèce ont été séquencés.

L'étudiant aura pour objectif de réaliser l'annotation de ces génomes. Pour cela, il devra mettre en œuvre les méthodes et outils en usage au laboratoire pour la prédiction des gènes Eucaryotes, et au besoin les adapter aux propriétés des génomes étudiés. Ainsi, l'étudiant devra faire un état des lieux des ressources disponibles pour la prédiction de gènes appliquées aux génomes de tiques, les exploiter ainsi que les transcriptomes générés dans le cadre de ce projet, et proposer un set de gènes localisés sur chacun de ces génomes à la suite d'une validation experte.

Le stage se déroulera au sein d'une équipe d'environ une quinzaine de personnes impliquées dans différentes thématiques bioinformatiques, allant du séquençage à l'analyse des génomes et de transcriptomes (<http://www.genoscope.cns.fr/rdbioseq>). L'étudiant devra faire preuve d'esprit d'équipe, de curiosité et d'adaptation dans le travail qui lui sera confié. Il devra être familier avec l'environnement type unix, avoir des connaissances de bases dans les langages de script (comme perl, python, bash, awk ou R) et en biologie des génomes.

## Utilisation de longues lectures pour l'amélioration du catalogue de gènes du projet Tara Océans

L'ensemble des données du projet Tara Océans ont été générées à partir de la technologie commercialisée par Illumina. Cette technologie séquence des fragments de petites tailles, il faut donc passer par une étape d'assemblage pour reconstituer les gènes présents dans un

échantillon. L'arrivée du séquençage 'longue lecture' Oxford Nanopore Technologies (ONT) permet de s'affranchir de cette étape d'assemblage et donc, permettra en théorie de reconstituer des gènes plus complets. Dans ce contexte, nous proposons un stage orienté 'recherche en bioinformatique', visant à développer une méthode permettant de tirer parti de ces nouvelles données afin d'améliorer le catalogue des gènes existant.

Le stage se déroulera au sein d'une équipe d'environ une quinzaine de personnes impliquées dans différentes thématiques bioinformatiques, allant du séquençage à l'analyse des génomes et de transcriptomes. L'étudiant devra faire preuve d'esprit d'équipe, de curiosité et d'adaptation dans le travail qui lui sera confié. Il devra être familier avec l'environnement type unix, avoir des connaissances de bases dans les langages de script (comme perl, python, bash, awk ou R) et en biologie des génomes.

## Séquençage nanopore et bioinformatique

Le Genoscope participe à de nombreux projets visant à séquencer le génome d'organismes modèles. La plateforme de séquençage du Genoscope génère une grande quantité de données à l'aide de la technologie commercialisée par Oxford Nanopore Technologies (ONT, nanoporetech.com). Cette technologie permet de lire un fragment d'ADN qui transite dans un pore. Le passage du fragment d'ADN génère un courant électrique qui sera converti en bases à l'aide de logiciels, cette étape est appelée le basecalling. De nombreux outils de basecalling existent et permettent d'adresser différentes questions en fonction des caractéristiques du génome (contenu en GC, homopolymère, répétition ...) et de la question biologique (bases modifiées, ADN ou ARN). Le Genoscope souhaite intégrer et utiliser ces outils afin d'obtenir des données de haute qualité et adaptées aux projets de séquençage. La mission du stagiaire consistera à effectuer la veille technologique de ce domaine, d'évaluer les outils existants et d'intégrer les outils choisis au pipeline existant.

Le stage se déroulera au sein d'une équipe d'environ une quinzaine de personnes impliquées dans différentes thématiques bioinformatiques, allant du séquençage à l'analyse des génomes et de transcriptomes. L'étudiant devra faire preuve d'esprit d'équipe, de curiosité et d'adaptation dans le travail qui lui sera confié. Il devra être familier avec l'environnement type unix, avoir des connaissances de bases dans les langages de script (comme perl, python, bash, awk ou R) et en biologie des génomes.

**Mots clés:** Basecalling, Epigénétique, Traitement du signal, Oxford Nanopore

**Référence:** Performance of neural network basecalling tools for Oxford Nanopore sequencing (Genome Biology 2019)

## Informations complémentaires

Lieu du stage : Genoscope – 2 rue Gaston Crémieux – 91000 – Evry

Equipe d'accueil : R&D BioSeq (<http://www.genoscope.cns.fr/rdbioseq>)

Contact : [stage\\_rdbioseq@genoscope.cns.fr](mailto:stage_rdbioseq@genoscope.cns.fr)

Durée du stage : 6 mois

Rémunération : À partir de 700€ brut (selon type d'étude), tickets restaurant et éventuellement une aide au logement ou transport

Pour candidater, merci de nous envoyer vos CV accompagné d'une lettre de motivation par mail à [stage\\_rdbioseq@genoscope.cns.fr](mailto:stage_rdbioseq@genoscope.cns.fr) . Les annonces ont également été publiées sur le site de la sfbi (<https://www.sfbi.fr/emplois-en-cours>).