

**Encadrants et laboratoire :**

Laboratoire : BIOSSE, anciennement Mer Molécules Santé (MMS)

Encadrantes : Nathalie CASSE, MCF HDR (biologie)

Aurélie PETICCA, Doctorante (bioinformatique)

Emails : [nathalie.casse@univ-lemans.fr](mailto:nathalie.casse@univ-lemans.fr)

[aurelie.peticca@univ-lemans.fr](mailto:aurelie.peticca@univ-lemans.fr)

Adresse : Laboratoire BIOSSE - Le Mans Université,

Avenue Olivier Messiaen 72085 Le Mans

Site internet : <https://bit.ly/3oY5212>

**Mots clefs :**

Génome, Transcriptome, RNA-seq, lncRNA, Eucaryotes, Microalgues, *Tisochrysis lutea*, *Phaeodactylum tricornutum*

**Description du projet :**

*Tisochrysis lutea* est une espèce de *Isochrysis* du genre des *Haptophytes*. Elle est connue pour sa très grande utilisation en tant qu'algue de fourrage dans la culture des bivalves et larves de crevettes. Son premier génome a été publiée en 2018. Quant à *Phaeodactylum tricornutum*, il s'agit d'une espèce de diatomée de la famille des *Haeodactylaceae*, dont le génome est connu depuis 2008. Ses applications sont diverses et vont de son utilisation comme précurseur de biocarburant à l'exploitation de sa capacité de biosynthèse et de son taux de croissance élevés comme hôte d'expression de protéines recombinantes.

L'objectif du laboratoire est de rechercher les longs ARN non codants (lncRNA) chez deux espèces de microalgues. Le/La stagiaire accompagnera une doctorante en bioinformatique dont le sujet comprend l'étude des lncRNA et leur relation avec les éléments transposables sur deux espèces de microalgues dont *Tisochrysis lutea*. Le travail du/de la stagiaire sera donc de créer un pipeline d'analyses automatisé capable d'annoter les lncRNA à partir d'un génome et de données RNA-seq d'une même microalgue. Pour cela il/elle aura à sa disposition un grand jeu de données ainsi que des contacts ayant déjà travaillé sur le sujet.

Des analyses génomiques à grande échelle ont mis en évidence l'existence d'une multitude de lncRNA. Bien que la fonction de la majorité de ces lncRNA reste à l'heure actuelle inexplorée, il semble clair que plusieurs de ces transcrits jouent un rôle important dans la régulation de l'expression génique et sont impliqués dans diverses pathologies, notamment le cancer. De fait, des études fonctionnelles ont démontré la participation des lncRNA à divers processus physiologiques.

### **Méthodologie :**

Le but de ce stage est de rechercher les lncRNA présents dans les données de séquençage des transcriptomes de deux microalgues : *Tisochrysis lutea* et *Phaeodactylum tricornutum* ; grâce à un pipeline automatisé qui sera produit par le/la stagiaire. Les différents logiciels seront proposés et/ou recherchés par ce/cette dernier(e), et le langage utilisée devra être bash et/ou python. Le/La stagiaire devra donc être capable de comparer et regrouper les résultats obtenus via différentes méthodes dans son pipeline afin d'obtenir une annotation des lncRNA la plus robuste possible.

Une fois l'annotation terminée, il sera possible de comparer cette annotation à celle des éléments transposables pour vérifier s'il existe des annotations semblables entre les deux. Mais il ne s'agit pas du but principal de ce stage, seulement d'une ouverture.

Une publication sur les résultats obtenus est prévu, le/la stagiaire sera cité(e) dans les auteurs à la hauteur de sa contribution.

### **Missions :**

- Reconstruction d'un transcriptome à partir de données RNA-seq
- Annotation des lncRNA depuis les transcriptomes et les génomes fournis
- Comparaison avec des annotations déjà existantes

### **Profil du candidat :**

- Expertise en environnement UNIX
- Connaissances solides en langages bash, Python et/ou R
- Expériences dans l'étude de données génomiques et transcriptomiques
- Bonne capacité d'organisation, de travail en autonomie et de communication avec les scientifiques en biologie
- Connaissances basiques sur les éléments transposables peuvent être un plus
- Expériences dans l'étude d'organismes non modèles appréciées

### **Publications relatives au projet :**

Berthelie J, Casse N, Daccord N, Jamilloux V, Saint-Jean B, Carrier G. A transposable element annotation pipeline and expression analysis reveal potentially active elements in the microalga *Tisochrysis lutea*. *BMC Genomics*. 2018 May 22;19(1):378. doi: <https://doi.org/10.1186/s12864-018-4763-1>. PMID: 29783941; PMCID: PMC5963040.

Bowler, C., Allen, A., Badger, J. et al. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244 (2008). <https://doi.org/10.1038/nature07410>

Carrier G, Garnier M, Le Cunff L, Bougaran G, Probert I, et al. (2014) Comparative Transcriptome of Wild Type and Selected Strains of the Microalgae *Tisochrysis lutea* Provides Insights into the Genetic Basis, Lipid Metabolism and the Life Cycle. *PLOS ONE* 9(1): e86889. <https://doi.org/10.1371/journal.pone.0086889>

Carrier G., Baroukh C., Rouxel C., Duboscq-Bidot L., Schreiber N., Bougaran G., Draft genomes and phenotypic characterization of *Tisochrysis lutea* strains. Toward the production of domesticated strains with high added value, *Algal Research*, Volume 29, 2018, Pages 1-11, ISSN 2211-9264, <https://doi.org/10.1016/j.algal.2017.10.017>

Rastogi, A., Maheswari, U., Dorrell, R.G. et al. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci Rep* 8, 4834 (2018). <https://doi.org/10.1038/s41598-018-23106-x>