

Le Genoscope, département de l'IBFJ/CEA, est un centre de recherche spécialisé dans la génomique environnementale et l'exploration des fonctions métaboliques. Parmi ses laboratoires, le LAGE (Laboratoire d'Analyses Génomiques des Eucaryotes ; UMR8030 Génomique Métabolique CNRS/CEA/Univ Paris-Saclay) étudie la structure et l'évolution des génomes eucaryotes, ainsi que la structure et la fonction des communautés environnementales eucaryotes par des approches bio-informatiques.

Plusieurs sujets de stage sont proposés au sein du laboratoire:

Sujet 1: Modélisation et analyses *climato-génomiques* de la dynamique de la biogéographie du plancton dans l'océan global: des communautés aux gènes, impacts des courants marins et du climat.

Le plancton est un acteur majeur des écosystèmes marins et du climat terrestre, entre autres à l'origine de la production primaire de matière organique via la photosynthèse soutenant toute la chaîne trophique des océans et la pompe à carbone biologique.

Ses communautés sont complexes et composées d'organismes couvrant toutes les branches principales de l'arbre du vivant: virus, bactéries, archaebactéries, eucaryotes unicellulaires et pluricellulaires. Il est essentiellement microscopique et passivement transporté par les courants.

La compréhension des mécanismes régulant sa distribution spatiale, sa *biogéographie*, est un challenge scientifique dont l'importance s'est accentuée par le contexte de changement climatique. Des travaux récents, dont certains du laboratoire d'accueil^{1,2}, proposent que la biogéographie résulterait des liens étroits entre les processus physico-chimiques environnementaux (température, apports de nutriments), le transport par les courants et les propriétés de l'évolution des génomes. Depuis 10 ans, les collectes de l'expédition *Tara Oceans* fournissent des jeux de données génomiques du plancton et physico-chimiques issues de centaines de sites de tous les océans permettant de tester des théories toujours débattues sur les écosystèmes marins^{5,6}.

L'objectif de ce stage est en premier lieu d'acquérir et comparer des résultats de régionalisations des océans à partir de données génomiques et transcriptomiques récentes^{3,4} en les combinant à des données physico-chimiques environnementales à l'aide d'outils statistiques (machine learning, PHATE, t-SNE, SOM). Les données *omiques* donnant des renseignements de différentes natures (composition taxonomique, évolution, fonctions géniques), elles donneront l'opportunité au candidat(e) de comparer les partitionnements géographiques obtenus pour modéliser et/ou en tirer de nouvelles hypothèses sur l'organisation des communautés et des écosystèmes marins à différentes échelles spatio-temporelle et biologiques.

Le/a candidat(e) motivé(e) pour poursuivre le sujet en doctorat serait amené(e) à se présenter aux concours d'obtention d'une bourse de thèse. Il/elle devra mettre en œuvre des outils de traitements numériques de données massives omics ainsi que des approches statistiques de machine learning, un savoir-faire en programmation est requis.

Ce sujet sera développé en interaction avec un doctorant et en collaboration avec le laboratoire des sciences du climat et de l'environnement du CEA (LSCE), un laboratoire d'océanographie physique de Naples et les différents laboratoires du consortium *Tara*

Oceans. Le stage se déroulera au Laboratoire d'Analyses Génomiques des Eucaryotes du CEA - Genoscope à Evry.

Contact: Olivier Jaillon (ojailon@genoscope.cns.fr)

1. Richter, D. J. et al. Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *bioRxiv* 867739 (2019). doi:10.1101/867739
2. Frémont et al. Restructuring of genomic provinces of surface ocean plankton under climate change. *bioRxiv* 34723 (2020). doi:<https://doi.org/10.1101/2020.10.20.347237>
3. Delmont, T. O. et al. Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics. *bioRxiv* 2020.10.15.341214 (2020). doi:10.1101/2020.10.15.341214
4. Carradec, Q. et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* (2018). doi:10.1038/s41467-017-02342-1
5. Karsenti, E. et al. A holistic approach to marine eco-systems biology. *PLoS Biol* (2011). doi: 10.1371/journal.pbio.1001177.
6. Sunagawa S. et al. Tara Oceans: towards global ocean ecosystems biology. *Nat Rev Microbiol.* (2020) doi: 10.1038/s41579-020-0364-5.

Sujet 2 : Étude et implémentation d'un outil de reconstruction de modèles métaboliques pour des eucaryotes unicellulaires.

1. Contexte

Les changements dans les communautés planctoniques marines, dus aux variations environnementales, impactent la chaîne trophique marine ainsi que les cycles biogéochimiques terrestres principaux (carbone, nitrate, oxygène...). Afin de mesurer et prédire ces changements et leurs conséquences, il est essentiel de décrire les relations fonctionnelles qui dirigent les assemblages de communautés d'organismes et de quantifier leurs contributions relatives dans les principaux traits et services fonctionnels des écosystèmes (fixation de l'azote, production de composés soufrés...).

Des méthodes bio-informatiques ont été développées pour reconstruire des génomes à partir de séquences environnementales (MAGⁱ: metagenome-based assembled génomes, et MGTⁱⁱ: metagenome based transcriptome). Ainsi, grâce aux données de *Tara Ocean*ⁱⁱⁱ, il est possible d'identifier les principaux acteurs planctoniques.

En complément de ces approches, nous projetons de reconstruire des modèles métaboliques (genome-scale metabolic network)^{iv}. Les réseaux métaboliques sont des outils *in silico* recensant l'ensemble des réactions métaboliques identifiées chez un organisme. Une fois un réseau métabolique généré, l'analyse des flux équilibrés (Flux Balance Analysis)^v permet de simuler les comportements d'un organisme (croissance) en fonction des conditions environnementales et de relier ces comportements aux voies métaboliques empruntées. Il est également possible de procéder à une simulation à l'échelle d'une communauté permettant de mieux comprendre les interdépendances métaboliques entre les organismes la constituant^{vi}.

Bien que le nombre de génomes d'organismes eucaryotes est en augmentation constante, peu de modèles métaboliques sont cependant disponibles. Cela est en partie dû aux étapes laborieuses des méthodes de reconstruction dites 'bottom-up'^{vii, viii}, nécessaires

pour obtenir un modèle de qualité. Pour surmonter ces difficultés, Machado *et al.*^{ix} ont alors développé l'outil CarveMe. Partant d'un modèle 'générique' et des séquences de génome annotées, CarveMe permet de reconstruire des modèles métaboliques de manière complètement automatisé, et dont les performances sont proches des modèles construits en 'bottom-up'. Cependant, cette méthode ne s'applique à ce jour qu'aux procaryotes.

2. Descriptif du stage

En amont du stage, un réseau métabolique 'générique' d'organismes eucaryotes autotrophes marins (EAM) a été préalablement généré. Les génomes (MAG/MGT) d'organismes EAM sont disponibles au sein du laboratoire. Le stage consistera à :

- i) Étudier l'outil CarveMe, et implémenter le modèle générique des EAM.
- ii) Reconstruire les modèles des MAGs/MGTs EAM disponibles au laboratoire.
- iii) En fonction du temps disponible et de la volonté du stagiaire, des analyses biologiques pourront être effectuées (modélisation à l'échelle de communautés...)

3. Conditions du stage

Le stage aura lieu en 2021 au Genoscope/CEA (2 rue Gaston Crémieux, 91000 Evry), dans l'équipe du LAGE (Laboratoire d'Analyse Génomique des Eucaryotes), et sera co-encadré par Eric Pelletier (Chercheur, eric.pelletier@genoscope.fr) et Marie Burel (Doctorante, marie.burel@genoscope.fr), pour une durée de 6 mois.

Le candidat recherché doit être en train de préparer un diplôme de niveau bac +5 (bioinformatique, biologie des systèmes...). Compétences solides en programmation (Python ou autre), à l'aise avec l'environnement Linux. Motivation, ténacité, autonomie.

- i. Tom O. Delmont et al., « Nitrogen-Fixing Populations of Planctomycetes and Proteobacteria Are Abundant in Surface Ocean Metagenomes », *Nature Microbiology* (juillet 2018) doi:10.1038/s41564-018-0176-9.
- ii. Vorobev et al., « Transcriptome Reconstruction and Functional Analysis of Eukaryotic Marine Plankton Communities via High-Throughput Metagenomics and Metatranscriptomics », *Genome Research* (2020), doi:10.1101/gr.253070.119.
- iii. Karsenti et al., « A Holistic Approach to Marine Eco-Systems Biology », *PLOS Biology* (2011) doi:10.1371/journal.pbio.1001177.
- iv. Maarleveld et al., « Basic concepts and principles of stoichiometric modeling of metabolic networks », *Biotechnology Journal* (2013) doi:10.1002/biot.201200291.
- v. D Orth, Thiele, et Palsson, « What Is Flux Balance Analysis? », *Nature Biotechnology* (2010) doi:10.1038/nbt.1614.
- vi. Budinich et al., « A Multi-Objective Constraint-Based Approach for Modeling Genome-Scale Microbial Ecosystems », *PLOS ONE* (2017) doi:10.1371/journal.pone.0171744.
- vii. Karp, Paley, et Romero, « The Pathway Tools Software », *Bioinformatics (Oxford, England)* (2002) doi:10.1093/bioinformatics/18.suppl_1.s225.
- viii. Aite et al., « Traceability, Reproducibility and Wiki-Exploration for "à-La-Carte" Reconstructions of Genome-Scale Metabolic Models » *PLOS Computational Biology* (2018) doi:10.1371/journal.pcbi.1006146.
- ix. Machado et al., « Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities », *Nucleic Acids Research* (2018) doi:10.1093/nar/gky537.

Sujet 3 : Modélisation métabolique: Caractériser l'hôte inconnu de la cyanobactérie symbiotique UCYN-A.

1. Contexte

Les changements dans les communautés planctoniques marines, dus aux variations environnementales, impactent la chaîne trophique marine ainsi que les cycles biogéochimiques principaux (carbone, nitrate, oxygène...). Afin de mesurer et prédire ces changements et leurs conséquences, il est essentiel de décrire les relations fonctionnelles qui dirigent les assemblages de ces communautés (fixation de l'azote, production de composés soufrés...). Grâce aux données de Tara Oceanⁱ, il est possible d'identifier les principaux acteurs planctoniques. En effet, des méthodes bio-informatiques ont été développées afin de reconstituer des génomes à partir de séquences environnementales tels que les MAGⁱⁱ (metagenome-based assembled genomes) ou les MGTⁱⁱⁱ (metagenomics-based transcriptomes). Contrairement aux MAG, il n'y a pas de 'binning' (basé sur les caractéristiques des séquences) effectué pour les MGT ; des gènes appartenant à des taxonomies différentes peuvent se retrouver en leur sein, capturant ainsi des organismes en symbiose.

C'est le cas du MGT-v1_29 dont une partie des gènes correspondent à la cyanobactérie UCYN-A (symbiose obligatoire), et l'autre, à son éventuel hôte, encore inconnu. Un candidat potentiel serait voisin de l'haptophyte *Braarudosphaera bigelowii*^{iv}. *A contrario*, UCYN-A est bien caractérisée. Plusieurs fonctions et voies métaboliques essentielles y sont absentes, telles que la photosynthèse oxygénique, la fixation du carbone (cycle de Calvin), le cycle de Krebs, et le cycle de l'urée. Elle possède cependant la nitrogénase et est donc dite diazotrophe, et synthétise la pseudocobalamine (vitamine B12). Son rôle n'est pas très clair, mais le fait qu'UCYN-A ait la voie métabolique complète dans un génome aussi réduit indique son importance. De plus, dans la partie de MGT-v1_29 associée à l'hôte, nous avons le gène metH qui permet de modifier la B12 bactérienne en B12 assimilable pour les eucaryotes.

Afin de mieux caractériser le métabolisme de l'hôte, et comprendre les interdépendances métaboliques entre ces deux organismes et leur environnement, nous projetons d'utiliser des modèles métaboliques (genome-scale metabolic network)^v. Les réseaux métaboliques sont des outils *in silico* recensant l'ensemble des réactions métaboliques identifiées chez un organisme. Une fois un modèle généré, l'analyse des flux équilibrés (Flux Balance Analysis, FBA)^{vi} permet de simuler les comportements d'un organisme (croissance) et de les relier aux voies métaboliques empruntées. Des outils tels que Mocabpy^{vii} permettent de procéder à une simulation à l'échelle de communautés, mais ne sont pas adaptés au cas spécifique de la symbiose.

2. Descriptif du Stage

En amont du stage, les réseaux métaboliques de i) l'hôte (via les outils AuReMe^{viii}, MetaCyc^{ix}) et ii) UCYN-A (CarveMe^x, BiGG^{xi}) seront générés. Le stage consistera à :

- i) Étudier l'outil Mocabpy et implémenter le mode 'symbiose' (collaboration avec son créateur, Marko Budinich).
- ii) Caractériser l'hôte, et analyser les interdépendances métaboliques entre cet organisme et UCYN-A.

3. Conditions du stage

Le stage aura lieu en 2021 au Genoscope/CEA (2 rue Gaston Crémieux, 91000 Evry-Courcouronnes), dans l'équipe du LAGE (Laboratoire d'Analyse Génomique des Eucaryotes), et sera co-encadré par Eric Pelletier (Chercheur, eric.pelletier@genoscope.fr) et Marie Burel (Doctorante, marie.burel@genoscope.cns.fr), pour une durée de 6 mois.

Le candidat recherché doit être en train de préparer un diplôme de niveau bac +5 (bioinformatique, biologie des systèmes...). De réelles compétences en programmation (Python ou autre) ainsi que de bonnes connaissances en biologie sont exigées.

- i. Karsenti *et al.*, 'A Holistic Approach to Marine Eco-Systems Biology', *PLOS Biology*, 9.10 (2011), e1001177 <<https://doi.org/10.1371/journal.pbio.1001177>>.
- ii. Delmont *et al.*, 'Nitrogen-Fixing Populations of Planctomycetes and Proteobacteria Are Abundant in Surface Ocean Metagenomes', *Nature Microbiology*, 3.7 (2018), 804–13 <<https://doi.org/10.1038/s41564-018-0176-9>>.
- iii. Vorobev *et al.*, 'Transcriptome Reconstruction and Functional Analysis of Eukaryotic Marine Plankton Communities via High-Throughput Metagenomics and Metatranscriptomics', *Genome Research*, 2020 <<https://doi.org/10.1101/gr.253070.119>>.
- iv. Hagino *et al.*, 'Discovery of an Endosymbiotic Nitrogen-Fixing Cyanobacterium UCYN-A in *Braarudosphaera Bigelowii* (Prymnesiophyceae)', *PLoS ONE*, 8.12 (2013) <<https://doi.org/10.1371/journal.pone.0081749>>.
- v. Maarleveld *et al.*, 'Basic Concepts and Principles of Stoichiometric Modeling of Metabolic Networks', *Biotechnology Journal*, 8.9 (2013), 997–1008 <<https://doi.org/10.1002/biot.201200291>>.
- vi. D Orth, Thiele, and Palsson, 'What Is Flux Balance Analysis?', *Nature Biotechnology*, 28.3 (2010), 245–48 <<https://doi.org/10.1038/nbt.1614>>.
- vii. Budinich *et al.*, 'A Multi-Objective Constraint-Based Approach for Modeling Genome-Scale Microbial Ecosystems', *PLOS ONE*, 12.2 (2017), e0171744 <<https://doi.org/10.1371/journal.pone.0171744>>.
- viii. Aite *et al.*, 'Traceability, Reproducibility and Wiki-Exploration for “à-La-Carte” Reconstructions of Genome-Scale Metabolic Models', ed. by Jens Nielsen, *PLOS Computational Biology*, 14.5 (2018), e1006146 <<https://doi.org/10.1371/journal.pcbi.1006146>>.
- ix. Caspi *et al.*, 'The MetaCyc Database of Metabolic Pathways and Enzymes - a 2019 Update' <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6943030/>> [accessed 7 July 2020].
- x. Machado *et al.*, 'Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities', *Nucleic Acids Research*, 46.15 (2018), 7542–53 <<https://doi.org/10.1093/nar/gky537>>.
- xi. King *et al.*, 'BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models', *Nucleic Acids Research*, 44.D1 (2016), D515–22 <<https://doi.org/10.1093/nar/gkv1049>>.

Sujet 4 : Une approche transcriptomique pour élucider la persistance des spécialistes et généralistes parmi les parasites planctoniques

1. Contexte

Le parasitisme est un mode de vie très fréquent dans la nature et une source majeure de pression évolutive tant pour les hôtes que pour les parasites. Contrairement aux relations parasitaires bien connues rencontrées dans l'environnement terrestre, nous sommes encore loin de comprendre pleinement les processus impliqués dans les interactions entre les micro-algues et ses parasites, avec une partie importante des effecteurs impliqués dans l'interaction hôte-parasite restant à découvrir dans des niches écologiques telles que la communauté planctonique.

Une des thématiques de recherche au sein du LAGE se porte sur la caractérisation, par des approches « omiques », de la relation hôte-parasite dans certaines branches de l'arbre de vie eucaryote. Cette approche est utilisée aujourd'hui sur un groupe de protistes parasites ubiquistes capables d'infecter certaines micro-algues toxiques responsables d'efflorescences connues sous le nom de marées rouges, les *Amoebophrya*, mal connues en comparaison des autres membres du plancton marin (1).

En culture, les souches appartenant au complexe d'espèces *Amoebophrya* ont un spectre d'hôtes très étroit : alors que certains clades (groupes génétiques) n'infectent qu'une seule espèce de dinoflagellés, d'autres sont capables d'infecter plusieurs espèces de dinoflagellés, voire deux genres différents.

Sur le terrain, la même espèce d'*Amoebophrya* peut infecter la même espèce d'hôte, année après année. Ainsi, les souches appartenant à un même clade ont des spectres d'hôtes assez similaires.

Les écosystèmes planctoniques côtiers sont par nature caractérisés par de fortes fluctuations environnementales. Cela devrait théoriquement conduire à la sélection naturelle de parasites généralistes au détriment des spécialistes. La spécificité de ces parasites pourrait s'expliquer, au moins en partie, par leur bagage génétique. Alors la persistance et le succès écologique des spécialistes parmi les parasites planctoniques marins reste encore un paradoxe à dévoiler.

2. Descriptif du stage

Le stage aura lieu en 2021 (pour une durée de 6 mois) au sein de l'équipe LAGE (Genoscope/CEA; 2 rue Gaston Crémieux, 91000 Evry).

Au cours du stage, le candidat explorera plusieurs voies pour dévoiler les gènes et protéines importants impliqués dans la spécialisation parasitaire, en utilisant le groupe *Amoebophrya* comme modèle d'étude. Ainsi, l'**objectif du stage M2** sera d'étudier *in silico* des nouveaux transcriptomes de *Amoebophrya* provenant des infections des différents espèces d'hôtes (dinoflagellés) par des souches généralistes d'*Amoebophrya* en s'appuyant sur les transcriptomes de références (2, 3), d'analyser leurs spécificités par génomique comparative puis d'interpréter les variations d'expressions de certains gènes dans les différents échantillons en fonction du stade de cycle de vie. A cette fin, le candidat devra apprendre à manipuler ces grandes quantités de données à l'aide d'outils bio-informatiques afin d'extraire les informations biologiques d'intérêt.

Finalement, et en fonction du temps disponible et de la volonté du stagiaire, une étude de l'impact à l'échelle mondiale de ces parasites pourra être menée grâce à la collection de séquences planctoniques obtenues au cours du projet Tara Océans (4, 5).

3. Profil du candidat :

- Connaissances en génomique, évolution et bio-informatique (une première expérience en analyse d'expression différentielle sera un plus).

- Maîtrise de l'environnement UNIX.
- Un savoir-faire en programmation (Python ou autre) avec des compétences solides en programmation en R sont requises.

Pour postuler, veuillez envoyer un CV accompagné d'une lettre de motivation à betina@genoscope.cns.fr

1. Guillou *et al.* 2008. *Environmental Microbiology*, 10: 3349-3365. doi:10.1111/j.1462-2920.2008.01731.x
2. Farhat *et al.* 2018, *Front Microbiol.* 9:2251. doi:10.3389/fmicb.2018.02251.
3. John *et al.* 2019. *Sci Adv.* 5(4):eaav1110. doi: 10.1126/sciadv.aav1110.
4. Karsenti *et al.* 2011, *PLOS Biology* doi:10.1371/journal.pbio.1001177.
5. Carradec *et al.* 2018. *Nat Commun.* 9(1):373. doi: 10.1038/s41467-017-02342-1.