

## STAGE M2 : structure et évolution des gènes *de novo*

**NOM, prénom de la personne proposant le stage :** Anne LOPES

Adresse Professionnelle : équipe de Bioinformatique Moléculaire - Institut de Biologie Intégrative de la Cellule (I2BC), Bâtiment 400, Université Paris-Sud, 91400 Orsay

Numéro de téléphone : 01 69 15 35 60

Adresse électronique : [anne.lopes@u-psud.fr](mailto:anne.lopes@u-psud.fr)

**Entreprise/Unité d'appartenance :** UMR 9198 - Institut de Biologie Intégrative de la Cellule (I2BC)

Domaine d'expertise de l'entreprise / laboratoire : bioinformatique, génomique, biologie des systèmes, bioinformatique structurale

Nom de l'équipe d'accueil ou du service : équipe de Bioinformatique Moléculaire

**Nom, prénom, statut et spécialité de la personne qui encadrera le stagiaire :**

Anne Lopes, maître de conférences en bioinformatique à l'université Paris-Sud

Numéro(s) de téléphone où l'on peut la joindre : 01 69 15 35 60

Adresse électronique : [anne.lopes@u-psud.fr](mailto:anne.lopes@u-psud.fr)

Autre(s) encadrant(s) : NON

**Titre du stage :** structure et évolution des gènes *de novo*

**Mots clés, orientation scientifique du sujet du stage et méthodes qui seront utilisées :**

Orientation : Bioinformatique structurale, structure des protéines, évolution

Méthodes : « protein design », « protein threading », développement, statistiques

### Description du sujet

**Contexte :** il ne fait plus aucun doute aujourd'hui, que l'émergence de gènes *de novo* (i.e. nouveaux gènes émergeant depuis le génome non-codant) joue un rôle important dans l'évolution des génomes [1]. De nombreux cas ont été et continuent d'être rapportés chez des organismes très différents tels que *S. cerevisiae*, *C. elegans*, *D. melanogaster*, *M. musculus* ou *H. sapiens* par exemple [2]. Par ailleurs, de nombreuses études montrent que ces gènes *de novo* jouent un rôle important dans l'adaptation des espèces. Ces derniers présentent en effet des profils d'expression bien spécifiques, ont été montrés comme soumis à pression de sélection négative et ont été montré comme essentiels dans de nombreux cas. Enfin, des approches de protéomiques ont récemment mis en évidence l'existence de peptides *de novo* dans des cellules humaines de cerveau [3] et dans des cellules neuronales de souris [4]. Toutes ces études montrent que la création de gènes *de novo* est un processus important et sous-estimé dans la création de diversité génétique. Cependant, les mécanismes impliqués dans l'émergence de ces gènes restent encore inconnus et sont pourtant essentiels pour comprendre les forces évolutives régissant l'apparition de ces nouvelles protéines et l'évolution des génomes en général.

**Objectifs :** ce projet se situe principalement à l'interface entre l'analyse de structures et de séquences des protéines et fera appel à des connaissances/outils relevant de l'informatique, des statistiques, de la génomique et de la bioinformatique structurale. L'objectif est de caractériser les propriétés évolutives et structurales de ces

protéines/peptides *de novo* afin de comprendre les mécanismes moléculaires permettant leur apparition dans la cellule, leur fonctionnement et finalement leur contribution dans l'apparition de nouveaux repliements protéiques.

**Travail demandé :** ce stage consistera à développer une approche automatique pour identifier ces gènes *de novo* dans le génome non-codant et à les caractériser par des approches de bioinformatique structurale. Les résultats obtenus à l'issue de ce stage permettront d'améliorer notre compréhension des mécanismes survenant (i) dans l'émergence de nouvelles protéines depuis le génome non-codant (mécanisme encore largement sous-étudié) et (ii) dans l'émergence de nouvelles structures protéiques participant ainsi à la diversité des repliements 3D protéiques actuels.

**Compétences techniques recherchées :** bonnes compétences en programmation en statistiques et bioinformatique structurale.

**Références bibliographiques :**

[1] Tautz D, Domazet-Lošo T. 2011, The evolutionary origin of orphan genes. *Nat Rev Genet.* 12:692-702

[2] Carvunis AR, *et al* 2012, Proto-genes and *de novo* gene birth. *Nature.* 487:370-4

[3] Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A. 2013, Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol.* 9:59-64.

[4] Prabakaran S, Hemberg M, Chauhan R, Winter D, Tweedie-Cullen RY, Dittrich C, Hong E, Gunawardena J, Steen H, Kreiman G, Steen JA. 2014, Quantitative profiling of peptides from RNAs classified as noncoding. *Nature Commun.* 5:5429

**Ce sujet constitue un premier pas vers un travail de thèse (oui / non) : NON**

**Durée estimée du stage :** 5 à 6 mois

**Montant (brut mensuel) de la rémunération proposée :** 554.40 euros pour 22 jours ouvrés (154H)