

## **Approches métabotéomiques en microbiologie environnementale**

**Jean Armengaud<sup>1</sup>, Elodie Briant-Litzler<sup>2</sup>, Eric Quéméneur<sup>1</sup>**

<sup>1</sup> CEA – Institut de Biologie Environnementale et Biotechnologie, Marcoule

<sup>2</sup> Bertin Technologies, Département Systèmes pour les Biotechnologies, Montigny-le-Bretonneux

**Objectif :** Innover dans le traitement de données massives de spectrométrie de masse afin de sonder la diversité microbienne d'un échantillon biologique pour y identifier rapidement les agents pathogènes.

La spectrométrie de masse est un outil performant pour analyser systématiquement le contenu protéique d'un échantillon biologique. L'analyse d'échantillons biologiques hétérogènes constitue désormais un enjeu et nécessite d'augmenter considérablement le nombre d'indices et la fiabilité de la mesure. Des jeux de données de millions de spectres de fragmentation de spectrométrie de masse en tandem sont désormais à comparer à des banques de données de plusieurs milliers de génomes de bactéries (chacun d'entre eux codant pour des milliers de protéines différentes). Nous souhaitons tester de nouveaux modes opératoires afin de réussir le difficile compromis entre gain de temps pour l'ensemble du procédé et gain de sensibilité.

Le stage de Master2 sera consacré à tester et optimiser différentes conditions de préparation et d'analyse d'échantillons biologiques issus de prélèvements environnementaux. Pour cela, l'étudiant aura la charge de mettre en place des outils originaux de fouille de données adaptés à cette problématique. Des logiciels spécifiques seront conçus afin de créer des bases de données ad hoc évolutives par apprentissage, de développer des modèles numériques permettant de confronter le grand nombre de données expérimentales générées par la spectrométrie de masse à des données simulées, d'extraire l'information pertinente en fonction des conditions expérimentales testées, et d'évaluer le degré de confiance dans les résultats par des tests statistiques adaptés (évaluation de faux-positifs sur des données massivement parallèles).

Le sujet de thèse proposé dans la continuité de ce travail s'inscrit dans la perspective de créer à l'aide d'un industriel une ligne de produits, d'équipements et de services originaux sur ce créneau. Des bases de données propriétaires caractérisant certaines espèces bactériennes caractéristiques devront être développées. Des innovations en termes de tests statistiques pourront être explorées. Des simulations numériques permettront de valider la combinaison optimale des différentes approches testées et d'évaluer le niveau de confiance dans les résultats produits. Différentes méthodes de fouille de données pourront être également explorées et comparées. Celles-ci pourront notamment s'appuyer sur des données issues d'échantillons biologiques dangereux, simulées et réelles.

L'étudiant doit maîtriser des langages de programmation pour les traitements de données (C, C++, ou FORTRAN), de tri (Perl), de gestion des données (MySQL), et éventuellement d'interface web (Java ou PHP). L'étudiant doit faire preuve d'autonomie (informatique et gestion de projet), de sens d'innovation (propositions d'amélioration de tests statistiques par exemple), et s'intéresser à l'interface biologie/biophysique (interactions avec des experts en spectrométrie de masse et microbiologie).

### Références

[1] Demirev PA, Fenselau C (2008) Mass spectrometry in biodefense. *J. Mass Spectrom.* 43:1441-57.

[2] VerBerkmoes NC, Deneff VJ, Hettich RL, Banfield JF (2009) Systems biology: functional analysis of natural microbial consortia using community proteomics. *Nat Rev Microbiol.* 7:196-205.

## Références du laboratoire d'accueil

- [1] Armengaud J (2009) A perfect genome annotation is within reach with the Proteomics and Genomics Alliance. [Curr. Opin. Microbiol. 12:292-300.](#)
- [2] De Groot A, Dulermo R, Ortet Ph, Blanchard L, Guérin Ph, Fernandez B, Vacherie B, Dossat C, Jolivet E, Siguier P, Chandler M, Barakat M, Dedieu A, Barbe V, Heulin Th, Sommer S, Achouak W, Armengaud J (2009) Alliance of Proteomics and Genomics to unravel the specificities of Sahara Bacterium *Deinococcus deserti*. [PLoS Genetics. 5: e1000434.](#)
- [3] Zivanovic Y\*, Armengaud J\*, Lagorce A, Leplat C, Guérin Ph, Dutertre M, Anthouard V, Forterre P, Wincker P, Confalonieri F (2009) Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea. *Genome Biology*. 10: R70. (\*) These two authors contributed equally to this work.
- [4] Armengaud J (2009) Proteogenomics and system biology: the quest for the ultimate missing parts. *Expert Reviews of Proteomics* [in preparation.](#)

## Candidatures à adresser à :

Jean Armengaud, CEA/IBEB, Centre de Marcoule, 30207 Bagnols-sur-Cèze  
e-mail : [jean.armengaud@cea.fr](mailto:jean.armengaud@cea.fr)  
Elodie Brient-Litzler, Bertin Technologies, 78180 Montigny-le-Bretonneux  
e-mail : [brient-litzler@bertin.fr](mailto:brient-litzler@bertin.fr)

Le poste est à pourvoir pour l'année universitaire 2009-2010 et sera basé au CEA de Marcoule (30).