

: Titre stage M2 Bioinformatique

Développement d'un outil bioinformatique d'intégration de données biologiques
.utilisant les technologies du web sémantique

Mots clés : ontologies, rdf, sparql, intégration de données

: Contexte scientifique

L'équipe « Résistance et Adaptation » de l'UMR IGEPP développe des recherches sur les interactions plantes – bioagresseurs. Le projet Epiclub s'intéresse en particulier à une maladie majeure des Brassicacées cultivées (colza, choux, brocoli, moutarde) appelée la hernie des crucifères. Cette maladie est provoquée par un protiste et conduit à d'énormes pertes de rendement. L'équipe "Résistance et Adaptation" vise à identifier les mécanismes moléculaires de résistance des Brassicacées à cette maladie. Suivant une approche multidisciplinaire, des études génétiques, génomiques, épigénomiques et physiologiques ont été menées et ont
.permis de produire de nombreuses données hétérogènes

Les travaux menés au sein de l'équipe Dyliss (Dynamics, Logics and Inference for biological Systems and Sequences) de l'IRISA/INRIA vise à modéliser les acteurs clés de processus d'adaptation biologique. Pour cela, l'équipe utilise diverses méthodes d'intégration de données, d'analyse de séquences biologiques et de
.modélisation dynamique de systèmes biologiques

Le but du projet «Epiclub» est d'intégrer les données biologiques hétérogènes de résistance à la hernie grâce à un outil (Askomics) en cours de développement. Cet outil permettra tout d'abord d'intégrer des données à partir de fichiers structurés dans une base de données RDF puis d'interroger cette base de données en construisant visuellement des requêtes sans nécessiter la maîtrise de SPARQL. Askomics est développé pour les biologistes et leur permettra d'établir une vision intégrative du contrôle de la résistance à la hernie des crucifères et développer des stratégies de
.lutte contre ce bioagresseur

: Description du stage et ses missions

Le stagiaire devra collaborer au développement et à l'utilisation de l'outil «Askomics»
: incluant une ou plusieurs des missions suivantes

Evolution des fonctionnalités d'Askomics, packaging et distribution de -1
l'application. Cette mission implique des tâches de développement en langage
.python, dans un environnement web

Préparation et chargement de données réelles (données locales et/ou -2
.données publiques) : manipulation et formatage de données bioinformatiques

Test et validation de l'outil en utilisant les données chargées. Ces tests seront -3
.basés sur des questions biologiques liées au projet Epiclub

Le stagiaire sera affecté à l'UMR INRA IGEPP, et basé dans l'équipe INRIA Dyliss sous la responsabilité d'Aurélie Evrard, post-doctorante dont l'objectif est d'exploiter les données du projet Epiclub et d'Anthony Bretaudeau qui a une forte expérience dans le traitement et l'analyse de données génomiques et le développement

d'applications. L'équipe est associée à la plateforme de bioinformatique GenOuest qui offre une infrastructure pour le calcul intensif en bioinformatique incluant un cluster de calcul (approx. 300 cpu), et des solutions de stockage (100 Tb), ainsi qu'une importante variété d'outils et de banques de données dédiés au traitement et à l'analyse des données biologiques

: Compétences requises

Le développement d'AskOmics nécessitera des compétences en programmation web (langages python et javascript) ainsi qu'une connaissance des technologies du web sémantique (RDF, SPARQL). La préparation des données nécessitera aussi des compétences en développement (python, bash, bases de données relationnelles) et une connaissance du contexte biologique (QTL, transcriptomique, ...). Des bonnes aptitudes en communication et rédaction ainsi qu'une capacité à travailler dans un esprit de collaboration sont fondamentales

: Lieu du stage

INRIA/IRISA Equipe Dyliss

: Durée du stage rémunéré

mois 6

: Encadrants

Aurélié Evrard (Post-doctorante INRA, aurelie.evrard@rennes.inra.fr, 02 99 84 73 21) et Anthony Bretaudeau (Ingénieur INRA, anthony.bretaudeau@rennes.inra.fr, 02 99 84 71 87)