



Offre de stage master 2

Intégration d'informations hétérogènes avec données de génomique comparative

MODALITES DE CANDIDATURE

Les candidatures (CV + lettre de motivation) doivent être adressées jusqu'au 30/09/2018 par courriel à raphael.flores@inra.fr avec l'objet suivant : [2018-stage-integration-syteny]

CONTEXTE

Le/la stagiaire recruté(e) travaillera au sein de l'unité URGI (<https://urgi.versailles.inra.fr/>) du centre INRA de Versailles. Cette unité de recherche en bio-informatique du département de Biologie et Amélioration des Plantes travaille sur l'intégration de données (équipe SI, système d'information) et l'analyse des génomes (équipe Anagen). Composée de 25 à 30 personnes, elle héberge une plateforme bio-informatique d'envergure nationale (infrastructure nationale IFB) et internationale (infrastructure européenne ELIXIR, système d'information de la Wheat Initiative...). Elle centralise, intègre et analyse les données de génomique, de génétique et de phénotypique des plantes d'intérêt agronomique, assure le développement et l'exploitation d'un système d'information et de pipelines d'analyses de données hauts débits.

L'équipe SI a développé depuis 2011 l'outil SytenyViewer (<https://urgi.versailles.inra.fr/syteny/syteny/viewer.do>) en collaboration avec l'équipe « Paléogénomique et évolution » de l'INRA GDEC et valorisé dans plusieurs publications. Cet outil permet de visualiser des résultats de génomique comparative et évolutive entre espèces de plantes (céréales et dicotylédones). Il a évolué au fil du temps afin de présenter d'autres types de données (QTL, couverture en éléments transposables, publications). Un travail exploratoire est en cours pour évaluer l'utilisation d'une base de données orientée graphe pour le stockage des données et leur interrogation.

MISSIONS

L'outil SytenyViewer sera remis à jour pour permettre d'intégrer, sur la base de l'identifiant des gènes et de leur position sur les génomes, des informations génétiques (données QTL, GWAS, polymorphismes...), génomiques (domaines fonctionnels, données d'expression...) et bibliographiques, qui sont disponibles dans les bases publiques (InterPro, NCBI, Ensembl, EBI...) ou celles des partenaires (CATdb : <http://tools.ips2.u-psud.fr/CATdb>). Pour atteindre cet objectif d'intégration des données, un travail préalable sur les identifiants des gènes sera nécessaire pour pouvoir ensuite enrichir l'outil avec des références croisées dynamiques permettant les liens vers d'autres ressources publiques.

[1]

ENVIRONNEMENT DANS LEQUEL LE/LA STAGIAIRE EVOLUERA

Le/la stagiaire recruté(e) travaillera dans l'équipe SI, pratiquant des méthodes agiles (Scrum, intégration continue, TDD), dans le contexte du projet SyntenViewer (Institut Carnot Plant2Pro, <http://www.instituts-carnot.eu/fr/institut-carnot/plant2pro>) au sein duquel l'URGI à en charge l'évolution de l'outil SyntenViewer. Il/elle sera donc également amené à réfléchir et présenter son travail dans un cadre large incluant des biologistes, spécialistes en génomique comparative.

Il/elle aura accès à un environnement informatique diversifié incluant: Java, Shell, web services, PostgreSQL, Neo4J, Talend, Elasticsearch, Solr.

CAPACITES PERSONNELLES

Curiosité pour les nouvelles technologies. Goût pour le travail en équipe. Connaissance de l'environnement GNU/Linux.

REFERENCES BIBLIOGRAPHIQUES

[0] Quraishi UM, Pont C, Ain Q-u, Flores R, Burlot L, Alaux M, Quesneville H and Salse J (2017) Combined Genomic and Genetic Data Integration of Major Agronomical Traits in Bread Wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 8:1843. doi: 10.3389/fpls.2017.01843

[1] El Baidouri, M., Murat, F., Veyssière, M., Molinier, M., Flores, R.-G., Burlot, L., Alaux, M., Quesneville, H., Pont, C., Salse, J. (2017). Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytologist*, 213 (3), 1477-1486. DOI : 10.1111/nph.14113

[2] Murat, F., Zhang, R., Guizard, S., Gavranovic, H., Flores, R.-G., Steinbach, D., Quesneville, H., Tannier, E., Salse, J. (2015). Karyotype and gene order evolution from reconstructed extinct ancestors highlight contrasts in genome plasticity of modern rosid crops. *Genome Biology and Evolution*, 7 (3), 735-749. DOI : 10.1093/gbe/evv014

[3] Murat, F., Zhang, R., Guizard, S., Flores, R., Armero, A., Pont, C., Steinbach, D., Quesneville, H., Cooke, R., Salse, J. (2014). Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16K protogenes. *Genome Biology and Evolution*, 6 (1), 12-33. DOI : 10.1093/gbe/evt200

[4] Pont, C., Murat, F., Guizard, S., Flores, R.-G., Foucrier, S., Bidet, Y., Quraishi, U. M., Alaux, M., Doležel, J., Fahima, T., Budak, H., Keller, B., Salvi, S., Maccaferri, M., Steinbach, D., Feuillet, C., Quesneville, H., Salse, J. (2013). Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo- and neoduplicated subgenomes. *Plant Journal*, 1-29. DOI : 10.1111/tbj.12366