

Cette lettre d'information est destinée aux membres des équipes de recherche utilisant la plate-forme bio-informatique GenoToul. Elle a pour but de vous informer sur les évolutions de l'équipe, les nouveaux outils, services, conditions d'utilisation, projets et formations mis en place.

### 1/ Les prochains cycles d'apprentissage

#### **A/ Se former aux analyses bioinformatiques des données NGS sous Galaxy du 17 au 20 mai 2016 (reste 9 places)**

Ce cycle d'apprentissage est composé de 3 modules indépendants proposés par Sigenae et bioinfo genotoul :

- **Galaxy First Steps** : Après une initiation à l'environnement Galaxy, cette journée vous permettra de prendre en main l'instance toulousaine de Galaxy et de connaître les bonnes pratiques d'utilisation.
- **Galaxy : Reads alignment and small size variants calling** : il s'agit d'1,5 jours de formations vous permettant de comprendre les principaux formats de fichiers manipulés, d'apprendre à aligner les reads sur un génome de référence et de détecter SNP et petits indels via la suite GATK.
- **Galaxy : RNAseq alignment and transcripts assemblies** : Il s'agit d'1 jour de formation à l'analyse de données RNASeq. Après une introduction sur les données de transcriptomique issues des séquençages nouvelle génération, cette formation vous permettra d'effectuer un alignement épissé sur un génome de référence, de découvrir des nouveaux gènes et des nouveaux transcripts et enfin de quantifier leur expression.

| Module   | Date          | Durée en jours | Pré-requis         |
|--|---------------|----------------|--------------------|
| Galaxy First steps                                   | 17/05/2016    | 1              | Aucun              |
| Galaxy : Alignment and variants calling              | 18-19/05/2016 | 1,5            | Galaxy First steps |
| Galaxy : RNASeq alignment and transcripts assemblies | 19-20/05/2016 | 1              | Galaxy First steps |

#### **B/ 4 jours de formation à la métagénomique 16S et 18S sous Galaxy seront organisés du 20 au 23 juin 2016 (reste 1 place)**

Vous avez ou allez obtenir des données de métagénomique 16S ou 18S Miseq ou 454 ? Vous souhaitez les analyser efficacement sans utiliser la ligne de commande ? Les équipes Sigenae, NED (GenPhySE), TWB et bioinfo genotoul vous proposent pour cela une formation de 4 jours indivisibles sous environnement Galaxy. Le premier jour sera consacré, après une initiation à l'environnement Galaxy à la prise en main de l'instance Galaxy Toulousaine. Pendant deux jours vous apprendrez à utiliser le pipeline FROGS : [http://bioinfo.genotoul.fr/fileadmin/BIO\\_INFO\\_STAT\\_2015/oral/FROGS\\_GenotoulBioinfo.pdf](http://bioinfo.genotoul.fr/fileadmin/BIO_INFO_STAT_2015/oral/FROGS_GenotoulBioinfo.pdf) constitué des étapes principales suivantes : nettoyage, clustering, affiliation taxonomique, filtre et statistiques descriptives que nous vous détaillerons. Enfin la dernière journée vous permettra d'interpréter les résultats grâce aux outils statistiques proposés.

Pré-requis nécessaire : connaissance de R ou d'un autre langage de programmation.

#### **Pour tous les cycles d'apprentissage :**

Ces formations sont organisées sur le site INRA de Toulouse Auzeville.

Les tarifs sont disponibles à l'adresse suivante : <http://bioinfo.genotoul.fr/index.php?id=115>.

Les inscriptions s'effectuent sur cette page : <http://bioinfo.genotoul.fr/index.php?id=10>.

### 2/ Mise à jour de l'instance Galaxy toulousaine

L'instance Galaxy mise à disposition par la plateforme Sigenae de Galaxy sera mise à jour au printemps 2016 afin de bénéficier de nouvelles fonctionnalités tout en offrant une continuité de services.

### 3/ E-learning en partenariat avec la formation permanente nationale INRA

Afin de répondre à une demande de formation croissante, Sigenae a mis en place en collaboration avec la formation permanente nationale INRA une offre de e-learning. Pour le moment seules deux formations sont disponibles :

- Galaxy First steps (Initiation et spécificité de l'instance Sigenae / Genotoul)
- FastQC : outil de contrôle de la qualité des séquences

Ces auto-formations sont accessibles aux agents INRA via la plateforme de e-formation de l'Inra : rubrique «

Sciences » / « Bioinformatique » à cette adresse : <https://elearning.formation-permanente.inra.fr/> (authentification avec LDAP Inra).

#### **4/ Augmentation des tarifs de location d'espace disque**

Pour faire suite à l'augmentation des frais administratifs, la plate-forme augmentera de 10 % les tarifs de location d'espace de stockage pour les projets hors INRA. Cette augmentation prendra effet 1<sup>er</sup> avril prochain.

#### **5/ Questionnaire de satisfaction 2015**

Tout d'abord nous tenons à vous remercier d'avoir répondu à notre enquête de satisfaction cette année. Lors de cette enquête plusieurs d'entre vous nous ont communiqué certaines difficultés d'utilisation des logiciels sur notre infrastructure. Afin de rassembler les informations indispensables à un unique endroit, nous avons ajouté une rubrique software à notre FAQ (<http://bioinfo.genotoul.fr/index.php?id=11>). Suite à vos demandes relatives à la non accessibilité des tutoriaux de soumission de jobs sur le cluster, un lien direct vers les diapos de formation Unix/Linux et Cluster a été mis en place sur la FAQ.

De plus pour toutes demandes exceptionnelles (augmentation d'espace disque work, accès à hypermemq...) nous vous prions de remplir le formulaire de demandes exceptionnelles : <http://bioinfo.genotoul.fr/index.php?id=82>. Si vous souhaitez demander une augmentation du quota dans Galaxy merci d'utiliser : [sigenae-support@listes.inra.fr](mailto:sigenae-support@listes.inra.fr). Pour toute autre demande de support sur notre infrastructure, merci d'utiliser le mail : [support.genopole@toulouse.inra.fr](mailto:support.genopole@toulouse.inra.fr) ou de remplir le formulaire : <http://bioinfo.genotoul.fr/index.php?id=74>.

Enfin, une page de notre site web vous permet d'avoir accès à nos dernières newsletters : <http://bioinfo.genotoul.fr/index.php?id=157>.

#### **6/ Nouveaux arrivants dans l'équipe**

Depuis quelques temps, trois personnes nous ont rejoints en CDD : Gaele Lefort (IE Biostat, Financement GenoToul), Regis Ongaro (IE Bioinfo Financement France Génomique) et Alexandre Heurteaux (IR Bioinfo/Biostat, Financement GenEndurance). Les membres de l'équipe sont présentés sur cette page régulièrement mise à jour : <http://bioinfo.genotoul.fr/index.php?id=4>.

#### **7/ Interfaces de visualisation et d'exploration de données RNAseq de novo et sRNASeq**

RNABrowse (Mariette et al, 2014) est un environnement de stockage et de visualisation orienté utilisateur permettant d'explorer les données de manière descendante. Il propose de nombreux graphiques globaux ainsi que les informations détaillées pour chaque contig. Il a été cité dans une revue récente sur le RNASeq *de novo* : [Moreton et al, 2015](#). Nous hébergeons à ce jour plus d'une vingtaine d'instances, plusieurs instances ont donné lieu à des publications : [Maganou et al, 2014](#); [Hinaux et al, 2013](#); [Riviere et al, 2015](#), [Lesur et al, 2015](#) ; [Bouzid et al, 2014](#) ; [Levasseur et al, 2014](#) ; [Canales et al, 2013](#) ; [Ueno et al, 2013](#).

sRNABrowse est un outil permettant l'annotation, la visualisation du profil d'expression et l'exploration de données sRNAseq variées. Il permet de traiter plusieurs famille d'ARNnc et de faire l'analyse de l'expression différentielle entre deux conditions. Six instances sont maintenues et une a été publiée : [Juanchich et al, 2016](#).

#### **8/ Jflow : un nouveau gestionnaire de workflow**

Il existe deux principaux types de gestionnaire de workflow selon le type de public auquel il est destiné : 1) des gestionnaires de workflows avec une interface graphique permettant la création et l'exécution de workflow mais qui sont difficilement intégrables dans un site web déjà existant. 2) Des gestionnaires de workflow sans interface graphique fournissant un cadre aux développeurs de workflows permettant une plus grande flexibilité et la puissance d'un langage de programmation de haut niveau.

Jflow, quant à lui, combine à la fois une interface graphique conviviale et une API python intuitive facilitant le travail du développeur. De plus grâce à son organisation en plugins JQuery, il est facilement intégrable dans n'importe quel site web. Il repose sur Makeflow (Albrecht et al., 2012) and weaver (Bui et al., 2012) et supporte la plupart des systèmes de gestion des ressources de calcul (Condor, SGE, Work Queue ou une machine multi-cœurs). JFlow est disponible sous la licence GNU General Public License (GPL) ici : <http://bioinfo.genotoul.fr/jflow>. Il est livré avec une documentation complète, un « quick start » et un portail de test.

Voici la référence du papier : Jérôme Mariette, Frédéric Escudié, Philippe Bardou, Ibouniyamine Nabihoudine, Céline Noiro, Marie-Stéphane Trotard, Christine Gaspin, Christophe Klopp. Jflow: a workflow management system for web applications. *Bioinformatics*. 2015. doi 10.1093/bioinformatics/btv589.

Pour toute demande d'information ou de travaux, veuillez envoyer un mail à [support.genopole@toulouse.inra.fr](mailto:support.genopole@toulouse.inra.fr) en précisant vos noms et coordonnées.