

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 en ligne de commande du 14 au 18 mars 2016

Ce cycle d'apprentissage est composé de 4 modules indépendants proposé par Sigenae et bioinfo genotoul :

- **Unix / Linux** : cette journée vous permettra de connaître les commandes de base servant, par exemple, à se connecter, à copier des fichiers, les éditer et en sécuriser l'accès. Vous apprendrez aussi comment gérer votre espace disque, compresser et décompresser des fichiers. Connaître ces commandes consiste un pré-requis pour les 3 autres modules. Ce module vous permettra de mieux comprendre notre infrastructure et de découvrir comment s'en servir efficacement.
- **Cluster** : cette deuxième journée vous permettra quant à elle de lancer vos premiers jobs sur le cluster de calcul. Nous aborderons aussi les réservations de ressources (CPU, mémoire ...) et nous apprendrons comment suivre les différents "jobs" lancés. Lors des travaux pratiques vous découvrirez, entre autre, ou sont situés les logiciels et les banques de données que nous mettons à votre disposition.
- **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.
- **RNAseq alignment and transcripts assemblies** : Il s'agit d'1,5 jours 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
Unix / Linux	14/03/2015	1	Aucun
Cluster	15/03/2015	1	Unix / Linux
Alignment and variants calling	16-17/03/2015	1,5	Unix / Linux
RNASeq alignment and transcripts assemblies	17-18/03/2015	1,5	Unix / Linux

B/ 4 jours de formation à la métagénomique 16S et 18S sous Galaxy seront organisés du 11 au 14 avril 2016

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 ? La CATI Bios4Biol vous propose 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) 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. Aucun pré-requis n'est nécessaire.

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/ Coupure électrique prévue le 24/11 (contrôles réglementaires)

Cette coupure va induire des coupures réseaux sur l'ensemble des systèmes de la façon suivante :

- Difficultés d'accès aux différents sites web et machines virtuelles hébergés.

- Difficultés d'accès aux différents systèmes de stockage réseaux.

Pour éviter des perturbations sur le système d'ordonnancement de calculs, le cluster de calcul et son stockage (/work) seront arrêtés pendant toute la durée de l'intervention. Nous en profiterons pour effectuer des maintenances (mises à jour et contrôles d'intégrité). En conséquences, il n'y aura **pas de cluster de calcul genotoul entre lundi 23 novembre à 18h et mardi 24 novembre à 12h**. Les jobs en cours qui ne seront pas terminés avant lundi 23 novembre à 18h seront perdus. Nous nous excusons par avance du désagrément que cela engendrera. N'hésitez pas à contacter le support si vous avez besoin de plus d'informations concernant les systèmes impactés.

3/ Nouveauté : GENOSMP a maintenant accès au /work et au /work2

Genosmp (smpq) est désormais raccordée physiquement au système de stockage GPFS (work et work2) via une interface infiniband. Vous pouvez donc lancer des calculs sur le /work ou le /work2 en lecture et écriture directement sur cette machine. Il n'y a donc plus nécessité d'utiliser le /scratch (volume disque local à la machine). Nous vous rappelons que l'utilisation de cette machine à 3 To de mémoire vive est soumise à une demande exceptionnelle de ressources limitée dans le temps (<http://bioinfo.genotoul.fr/index.php?id=82>).

4/ qlogin et qrsh : petit rappel des différences

Le qrsh et le qlogin permettent de soumettre des jobs en interactif sur le cluster de calcul. Mais il y a des différences importantes entre les deux :

- le qlogin permet un déport de l'interface graphique alors que le qrsh non.
- le qlogin n'adresse qu'une seule machine (node001) alors que le qrsh adresse l'ensemble du cluster.

Donc, **si vous souhaitez soumettre des jobs en interactif qui ne nécessitent pas l'interface graphique, le mieux est d'utiliser qrsh.**

5/ Les traitements de données sur le serveur frontal genotoul sont interdits

Pour que l'ensemble des utilisateurs de genotoul puissent travailler dans de bonnes conditions nous avons dû interdire le traitement des données sur le serveur frontal. Les serveurs genotoul sont réservés exclusivement à la connexion, au transfert de données, à la compilation, au test de la ligne de commande et à la soumission de jobs sur le cluster de calcul. Tout traitement de données lancé directement (sans « qsub », « qarray », « qrsh » ou « qlogin ») sera systématiquement interrompu sans préavis par les administrateurs système.

6/ Votre mot de passe genotoul est strictement personnel et confidentiel

Nous avons constaté une recrudescence de l'utilisation de certains comptes genotoul par d'autres utilisateurs que le détenteur du compte. La charte (<http://bioinfo.genotoul.fr/fileadmin/Documents/ChartPFBioinfoGenoToul.pdf>) de la plateforme, que vous avez acceptée lors de la création de votre compte, stipule :

- "L'autorisation d'accès est strictement personnelle, incessible et temporaire"
- "L'utilisateur est responsable des dommages qui pourraient provenir d'une mauvaise utilisation de son compte s'il est établi qu'il en a donné le mot de passe".

Toute utilisation d'un compte par un utilisateur non détenteur de ce compte pourra entraîner la fermeture **définitive** du compte et une interdiction d'utiliser l'infrastructure.

Nous vous remercions par avance de votre compréhension.

7/ Petit bilan de notre activité cette année

Durant l'année 2015, 790 utilisateurs (répartis dans 247 équipes) ont utilisé genotoul au moins une fois au cours de l'année passée parmi lesquels 450 utilisateurs du cluster ont consommé un total de 6,2 millions d'heures de calcul. Nous avons traité plus de 1200 demandes utilisateurs. Nous mettons à jour environ 200 banques de données et avons environ 500 logiciels installés (sans compter les différentes versions). Nous hébergeons 45 machines virtuelles pour des utilisateurs externes.

Nous avons accompagné cette année 35 projets de recherche sur 9 thématiques différentes et formés 135 jours.personnes. Nous avons 7 projets de développement informatiques et bioinformatiques actifs. Les publications valorisant ces travaux sont consultables sur la page « publications » de notre site web : <http://bioinfo.genotoul.fr/index.php?id=54>. A ce jours, vous nous avez transmis 11 articles publiés en 2015 ayant utilisé les ressources de la plate-forme. N'hésitez pas à nous communiquer les références des travaux ayant nécessité l'utilisation de notre infrastructure, celles-ci constituent un indicateur de notre utilité pour la communauté.

Nous vous souhaitons une bonne fin d'année et espérons que l'année prochaine permettra de nouvelles collaborations scientifiques fructueuses.

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