

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/ 2 jours de formation Linux/Cluster seront organisés les 23 et 24 mars 2015 (reste 6 places)

Vous souhaitez utiliser notre infrastructure de calcul via la ligne de commande, mais vous avez besoin d'aide pour bien démarrer : connaître les commandes de base, mieux comprendre notre infrastructure et découvrir comment s'en servir efficacement. Cette formation a été conçue dans ce but.

Lors de la première journée, vous apprendrez comment vous connectez, comment copier des fichiers, les éditer et en sécuriser l'accès. Vous apprendrez aussi comment gérer votre espace disque, compresser et décompresser des fichiers. La formation au cluster de calcul (qui aura lieu le deuxième jour) vous permettra, quant à elle, de lancer vos premiers "jobs" sur le cluster. Lors de cette journée, 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, où sont situés les logiciels et les banques de données que nous mettons à votre disposition.

Les deux journées sont indépendantes mais savoir utiliser Linux/Unix est un pré-requis pour la formation à l'utilisation du cluster de calcul.

B/ 2 jours et demi de formation Galaxy seront organisés du 27 au 29 avril 2015 (Reste 1 place le premier jour)

Il s'agit de deux jours et demi indivisibles de formation à certains traitements de données des séquenceurs haut débit 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. Le 28 et le 29 au matin (09h00 - 12h00) vous permettront d'effectuer un alignement de séquences et une recherche de polymorphisme sous environnement Galaxy avec la suite GATK.

Ce cycle ne nécessite aucune connaissance préalable de la ligne de commande.

Pour tous nos 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>.

La plupart des formations que nous dispensons sont aussi disponibles sur la plate-forme d'e-learning sig-learning à l'adresse suivante : <http://sig-learning.toulouse.inra.fr>.

2/ Les traitements de données lancés directement sur les serveurs frontaux genotoul sont dorénavant interdits

Ces derniers temps nous avons déploré des impossibilités d'accès aux serveurs frontaux « genotoul » pour l'ensemble des utilisateurs à cause de traitements de données lancés directement sur ces serveurs. Pour que chacun puisse travailler, nous vous rappelons que 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. Désormais tout traitement de données lancé directement sur les serveurs genotoul sera systématiquement interrompu sans préavis par les administrateurs système.

Il est donc impératif d'utiliser les commandes "qsub" (batch), "qsh" (interactif) ou "qlogin" (interactif avec redirection graphique) pour tout traitement de données. N'hésitez pas à consulter la FAQ et les tutoriels de formation sur notre site web si vous avez besoin de plus d'informations sur ces commandes.

3/ Le support Galaxy : sigenae-support@listes.inra.fr

Si vous êtes utilisateur de l'instance Sigenae de Galaxy, l'adresse mail sigenae-support@listes.inra.fr vous sera certainement utile. Vous pouvez donc la rajouter dans vos contacts mail si vous pensez avoir besoin d'aide technique,

d'intégrer de nouveaux outils dans l'instance Sigene de Galaxy, si vous avez besoin qu'on ajoute une nouvelle banque de données ou encore pour nous communiquer une demande de formation.

4/ Les quotas d'utilisation du cluster de calcul

A/ Les quotas instantanés :

Il existe des quotas instantanés d'utilisation sur la mémoire et les processeurs en fonction du groupe d'appartenance de l'utilisateur et de la file de calcul. Pour connaître ces quotas, utilisez la commande :

```
qconf -srqs
```

Les quotas actuellement mis en œuvre sont notés « enabled TRUE » et la limite en nombre de slots ou de quantités de RAM sont notés après le mot clef « limit ».

De plus, nous avons limité le nombre total de jobs instantanément lancé par utilisateur (2.500 jobs maximum) dans la limite de 10.000 jobs maximum en queue dans le système d'ordonnement.

A/ Les quotas cumulés :

Il existe des quotas de calcul utilisateur cumulés sur l'année civile (par défaut à 100.000h de calcul pour les personnels académiques). Ces quotas ont été purgés au 1er janvier 2015. Pour connaître votre quota de calcul cumulé, utilisez la commande :

```
qquota_cpu user_name
```

Si votre projet nécessite davantage d'heures de calcul, mémoire ou espace disque, il vous suffit de remplir une demande de ressource exceptionnelle : <http://bioinfo.genotoul.fr/index.php?id=82>.

Ces divers quotas concernent aussi les utilisateurs de Galaxy.

5/ Une librairie d'utilitaires en bioinformatique mise à disposition

Le CATI (Centre Automatisé du Traitement de l'Information) Bios4BioI (Bioinformatique et statistique pour la Biologie) met à disposition ses utilitaires de bioinformatique à l'adresse suivante : <https://mulcyber.toulouse.inra.fr/projects/bioinfoutils/>.

Ces scripts sont installés sur genotoul dans le répertoire /usr/local/bioinfo/Scripts/bin, la liste de ces logiciels et la documentation utilisateur est disponible via notre site web dans l'onglet Services / Our bioinformatics scripts ou à l'URL suivante : <http://bioinfo.genotoul.fr/index.php?id=159>.

Pour avoir la documentation complète, il vous faudra utiliser perldoc pour un script perl ou l'option -help dans tous les cas.

7/ Enquête de satisfaction 2014

En premier lieu nous voudrions souligner le très fort taux de réponse à notre questionnaire cette année. Nous avons reçu 108 réponses (soit 16,6 % de retour). Merci beaucoup à tous de votre mobilisation, nous espérons pouvoir répondre au maximum de vos suggestions.

Parmi ces suggestions, certains d'entre vous ont mentionnés qu'ils souhaiteraient une classification des outils disponibles sur notre infrastructure. Celle-ci est trouvable sur notre site web en cliquant sur l'onglet Ressources / software (<http://bioinfo.genotoul.fr/index.php?id=5>). Par défaut tous les logiciels sont listés par ordre alphabétique. Si vous cliquez sur la catégorie qui vous intéresse à droite de la page, la liste se met à jour pour n'afficher que les logiciels de la catégorie choisie.

Nous avons par ailleurs mis à jour la description de notre cluster sur la page <http://bioinfo.genotoul.fr/index.php?id=158>, trouvable à partir de l'onglet ressources / hardware / genotoul cluster.

Enfin, une page rassemblant les dernières lettres d'informations de la plateforme a récemment été ajoutée sur notre site web. Celle-ci est accessible via l'onglet Help puis Newsletters.

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

Name	Description	Keywords
ace2sam.py	Convert an ace file to a sam file	assembly ace sam
ace_stats.py	Compute length and depth statistics and generate graphics on a ace file (zipped accepted)	454 assembly
addHetero.pl	This script search for heterologous sites in -fasta_het file and transfers these sites to the input fasta file.	heterologous, fasta
addReadGroup.pl	addReadGroup.pl Adds/Replaces the @RG header with given ID/library/sample/platform/platform unit/center tags in a sam file/stream	Readgroup GATK RG
align2fasta.py	Convert a cap3 output file (align zipped, gz or bz2 or not) to an align multifasta file	cap3 align convert multifasta
sam_check_forward_reverse.pl	Read SAMBAM input files and extract the mapping read strand information from the SAM flag.	sam bam forward reverse strand assembly contig
sam_check_pairs.pl	Read input SAMBAM files, extract the mapping reference and the pair information from the SAM flag.	sam bam pair paired-end assembly contig
bam2stat.pl	Calculate BWA BAM statistics: number of read paired/single uniquely/multimot initialer sequence mapped:	Alignment BAM
biasSelection.pl	Filter the blast outputs and display Query Id, Ref hit name, Hit start.	blast filter