

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

Nous vous proposons 3 cycles d'apprentissage dans les mois qui viennent (2 sous Galaxy et un en ligne de commande) :

A) 2,5 jours de formation à l'alignement et à la détection de variants à partir de NGS sous Galaxy du 9 au 11 mai 2017

Ce cycle d'apprentissage vous est proposé par l'équipe Sigeneae.

La première journée sera consacrée à la présentation de l'environnement Galaxy, puis à la prise en main de l'instance toulousaine. Suivront ensuite 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.

Aucun pré-requis n'est nécessaire.

B) 4 jours de formation à la métagénomique 16S et 18S sous Galaxy seront organisés du 3 au 6 juillet 2017 (reste 5 places)

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 Sigeneae, NED (GenPhySE), TWB et Bioinfo Genotoul vous proposent une formation de 4 jours indivisibles sous environnement Galaxy. La première demi-journée sera consacrée, 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 dont la description est disponible à l'url :

http://bioinfo.genotoul.fr/fileadmin/BIO_INFO_STAT_2015/oral/FROGS_GenotoulBioinfo.pdf. Il est constitué des étapes principales suivantes : nettoyage, clustering, affiliation taxonomique, filtre et statistiques descriptives que nous vous détaillerons. Enfin la dernière journée et demie 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 la partie statistique.

C) 3,5 jours de formation à l'analyse bioinformatique et biostatistique des données RNAseq sur génome de référence du 16 au 19 mai 2017

Ce cycle d'apprentissage vous est proposé par les plateformes Bioinfo Genotoul, Biostat Genotoul, l'équipe Sigeneae et l'unité MIAT.

Après un bref rappel sur l'environnement Unix et l'utilisation du cluster nous vous présenterons les données RNAseq. Nous discuterons notamment des plans d'expérience et des biais des différentes technologies de séquençage de type NGS. Vous apprendrez à analyser la qualité des données RNAseq et à les nettoyer. Vous effectuerez ensuite l'alignement sur le génome de référence et vous serez en mesure de découvrir de nouveaux gènes ou de nouveaux transcrits. Vous effectuerez ensuite les comptages afin de générer le tableau qui vous servira à mettre en évidence les gènes différentiellement exprimés entre deux conditions. Le dernier jour et demi sera consacré à l'analyse statistique. C'est à dire à l'exploration des données de comptage, la normalisation et l'analyse de l'expression différentielle.

Pré-requis nécessaire : connaissance basique du langage R et de la ligne de commande Unix/Linux.

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/training-2/pricing/>.

Les inscriptions s'effectuent sur cette page : <http://bioinfo.genotoul.fr/index.php/training-2/training/>.

2/ Les logiciels installés

La liste des logiciels installés sur genotoul a été mise à jour sur le site web. Elle est accessible à la page : <http://bioinfo.genotoul.fr/index.php/resources-2/software/>.

Une mise à jour des liens « current » (situé dans votre \$PATH) sera effectuée lors de l'arrêt des serveurs prévu le 18 mars prochain. Nous vous avons envoyé un mail avec le détail des versions des logiciels impactés par cette manipulation le 28 février dernier.

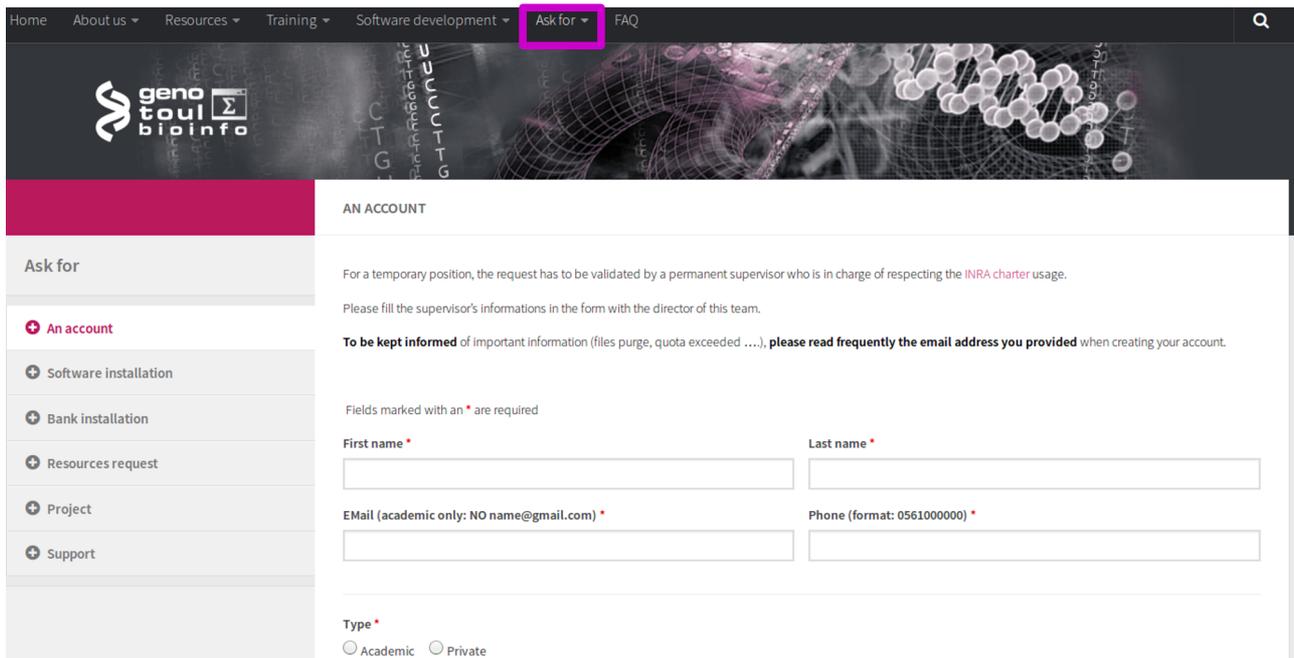
Enfin, pour trouver un logiciel installé nous vous conseillons d'utiliser la commande suivante :

```
ls /usr/local/bioinfo/src|grep -i soft_name
```

De même pour connaître les différents modules configurés et les utiliser, les commandes : `module avail`, `module display NomDuModule` et `module load NomDuModule` s'avèrent particulièrement utiles.

3/ La nouvelle rubrique « Ask for »

A partir de la rubrique « Ask for » de notre nouveau site web vous avez maintenant accès à plusieurs formulaires de demandes. Vous pouvez demander un compte sur le cluster et sur l'instance Galaxy (« An account »), l'installation ou la mise à jour d'un logiciel (« Software installation »), l'installation ou la mise à jour d'une banque (« Bank installation »). Vous pouvez également nous solliciter pour obtenir davantage de ressources : RAM, CPU, espace disque (« Ressources request »), ce service peut être payant selon les cas. Si vous souhaitez bénéficier de notre expertise pour vous aider dans vos projets traitants des NGS le formulaire ad hoc est celui intitulé « Projets ». Si vous avez besoin d'un support technique un formulaire est également accessible sous la rubrique Ask for (« Support »).



The screenshot shows the 'AN ACCOUNT' form on the genotoul bioinfo website. The 'Ask for' menu is highlighted in pink. The form includes the following fields and options:

- First name *** (text input)
- Last name *** (text input)
- Email (academic only: NO name@gmail.com) *** (text input)
- Phone (format: 0561000000) *** (text input)
- Type *** (radio buttons for Academic and Private)

Additional text on the page includes: "For a temporary position, the request has to be validated by a permanent supervisor who is in charge of respecting the INRA charter usage." and "Please fill the supervisor's informations in the form with the director of this team." A note states: "To be kept informed of important information (files purge, quota exceeded ...), please read frequently the email address you provided when creating your account."

4/ Petit rappel sur les comptes Genotoul

Nous vous rappelons que votre user et votre mot de passe sont strictement **personnels** et **confidentiels** pour des raisons de sécurité.

De plus les comptes non-utilisés depuis plus de 1 an ainsi que les comptes expirés des personnels temporaires (date de fin renseignée lors de la demande du compte) sont systématiquement clos et les données purgées.

5/ Fonctionnement de l'espace /work

L'espace disque utilisateur (/work) est un espace temporaire de calcul et peut être purgé si besoin (i.e ; lorsque le taux total de remplissage atteint 80 %). Lorsque ce nettoyage sera nécessaire, les utilisateurs concernés seront prévenus. **Pour être sûr de conserver les résultats de vos analyses il est nécessaire de les copier dans l'espace disque /save** (prévu à cet effet).

6/ Quelques consignes de bonnes pratiques

Pour que l'ensemble des utilisateurs de genotoul puissent travailler dans de bonnes conditions le traitement des données sur le serveur frontal est interdit. Les serveurs genotoul sont réservés exclusivement à la connexion, au transfert de données, à la compilation, au test très rapide 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. Ces quatre commandes sont explicitées dans notre FAQ, rubrique « Job Submission » (http://bioinfo.genotoul.fr/index.php/faq/job_submission_faq/) à la question : « Which commands can I use to submit my job » ?

De plus, lorsque vous êtes en qlogin et en qrsh, si vous avez fini vos traitements, nous vous conseillons de fermer votre session car cela occupe un nœud et compte dans votre quota CPU.

7/ Les news de l'instance Galaxy Sigenae / BioInfo Genotoul

L'instance Galaxy Sigenae / BioInfo Genotoul a été utilisée par 230 utilisateurs en 2016. Nous vous remercions de la confiance que vous nous accordez.

Cette instance Galaxy a été mise à jour en octobre 2016 afin de vous faire bénéficier d'une interface récente et de banques mieux gérées. Elle est référencée depuis peu par l'équipe américaine du Galaxy Project tant en terme d'instance (<https://galaxyproject.org/community/deployment/sigenae-bioinfo-genotoul/>) que de formations (<http://bit.ly/gxytrnGenoToul> et <http://bit.ly/gxytrnmap>).

L'instance s'est récemment enrichie de nouveaux outils pour :

- le traitement de vos données sRNAseq (la suite d'outils MirDeep2 a été ré-installée et améliorée).
- pour le variant calling (AnnoVar, SnpEff, Sequenza, Contra, Stacks, GATK 3.5).
- pour les données RNAseq (RSEM, la suite Cufflinks et les SARTools).
- autres outils (PICRUSt, LEfSe, SSPACE).

Pour en savoir plus sur le côté administration de Galaxy, la 8ième conférence mondiale aura lieu cette année en France (GCC 2017 Montpellier du 26 au 30 juin, <https://gcc2017.sciencesconf.org/>).

Le nombre de publications remerciant l'instance Galaxy Sigenae / BioInfo GenoToul est un indicateur de notre utilité pour nos recherches de financements par exemple. C'est pourquoi nous vous proposons le modèle de phrase suivant lorsque vos publications traitent d'un sujet ayant bénéficié des ressources fournies par cette instance : "We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees and Sigenae group for providing help and/or computing and/or storage ressources thanks to Galaxy instance <http://sigenae-workbench.toulouse.inra.fr>".

Dans le cadre d'une collaboration, vous pouvez directement citer la / les personnes qui ont participé au projet de la manière suivante : Name, Sigenae group, GenPhySE, INRA Auzeville CS 52627 31326 Castanet Tolosan cedex.

8/ Coupure de service entre vendredi 17 mars à 17h et lundi 20 mars

Une opération de maintenance des équipements électriques du datacenter INRA toulousain est planifiée le samedi 18 mars prochain. Elle nécessite un arrêt total de tous les services. En conséquence, la plateforme BioInfo GenoToul devra interrompre toutes ses infrastructures de calcul, stockage, hébergement de sites web et machines virtuelles pendant tout le week-end.

Les systèmes hébergés **seront arrêtés le vendredi 17/03 à partir de 17h et redémarrés progressivement à partir du lundi matin 20/03.**

Nous sommes désolés pour la gêne occasionnée par cette coupure indépendante de notre volonté.

9/ Bilan de l'enquête de satisfaction annuelle 2016

110 personnes ont répondu à notre dernière enquête de satisfaction annuelle, certains d'entre vous ont pris le temps de laisser des commentaires riches d'enseignements, et nous vous en remercions. L'un des résultats intéressants de cette enquête est que parmi les 76 personnes qui déclarent utiliser le cluster de calcul genotoul, 37 % ont publié les résultats obtenus. Au sujet de l'utilisation du cluster certains d'entre vous ont pointé le manque de tutoriaux disponibles. A ce sujet, nous nous permettons de vous rappeler les formations Linux et Cluster que nous proposons périodiquement (<http://bioinfo.genotoul.fr/index.php/training-2/training/>) ainsi que les supports de formation disponibles sur notre site web (<http://bioinfo.genotoul.fr/wp-content/uploads/FormationCluster3DiaParpage.pdf>). De plus si les quotas d'espace disque que nous vous accordons ne suffisent pas (1Tb sur le /work et 250 Gb sur /save), nous avons mis en place un service de location d'espace disque. Les tarifs sont disponibles ici : <http://bioinfo.genotoul.fr/index.php/resources-2/pricing/>.

L'équipe en charge du support est plébiscitée par vos avis et commentaires, par contre certains d'entre vous nous ont fait part de leur insatisfaction au sujet de l'affichage des versions des différentes banques de données.

Les numéros d'accèsion des génomes sont disponibles dans un fichier que vous pouvez trouver :

- pour Ensembl :
`/bank/ebi/ensembl/SPECIES_genome/current/flat/current_fasta/SPECIES/dna/README`
La version de l'annotation faite par Ensembl, quant à elle, est dans le fichier :
`/bank/ebi/ensembl/SPECIES_genome/current/flat/version.txt`
- pour le NCBI:
`/bank/ncbi/genomes/SPECIES/current/flat/README_CURRENT_RELEASE`

L'outil de mise à jour automatique des banques que nous utilisons ne nous permet pas de faire mieux, nous nous excusons de la gêne occasionnée, nous ne disposons pas actuellement des moyens humains nécessaires pour l'améliorer.

Par contre, grâce à l'outil Biomaj Watcher (<http://genoweb.toulouse.inra.fr/BmajWatcher/> : accessible depuis la page databanks de notre site web) vous avez accès à de nombreuses informations sur les banques dont les chemins d'accès sur genotoul.

The image shows two windows from the Biomaj Watcher application. The left window displays the 'Bank overview' for 'ensembl_canis_familiaris'. It includes sections for Properties, Production directories, and Release. The 'Release' section shows details for a 2016-11-23 release, including session date, duration, and file sizes. The right window shows a list of banks with columns for Release and Latest session date. A pink arrow points to the 'ensembl_canis_familiaris' entry in the list, with a text box saying 'Double cliquer sur la banque qui vous intéresse'. Another pink arrow points to the 'Production directory' field in the left window, with a text box saying 'Chemin d'accès sur genotoul'. At the bottom right of the right window, it says 'Total : 229'.

Plus spécifiquement au sujet de l'instance Galaxy : 34 utilisateurs de Galaxy ont répondu à l'enquête. 56 % précisent ne pas avoir besoin de support, mais 32 % souhaitent suivre une formation à l'utilisation de Galaxy. Nous vous rappelons à ce sujet le prochain cycle d'apprentissage qui aura lieu du 9 au 11 mai prochain (voir le chapitre 1 de cette newsletter). 82 % des utilisateurs ont trouvés que les résultats obtenus leur ont été utiles. Les principaux types d'outils utilisés sont ceux servant à la manipulation de séquences (47% des répondants), la manipulation de fichiers (38%), l'alignement de séquences (38%) et le traitement de données RNASeq (26%).

Certains d'entre vous ont regretté l'absence de certains outils et/ou de certains paramètres sur l'interface de l'instance Galaxy toulousaine. Nous vous rappelons que l'équipe Sigenae (sigenae-support@listes.inra.fr) est à votre disposition pour ajouter les outils et les paramètres dont vous avez besoin pour vos analyses bioinformatiques.

Lors de cette enquête, la conception de workflows et le choix des paramètres pour les différents outils ont également été évoqués comme pouvant être une limitation à l'utilisation de Galaxy. Les formations à Galaxy peuvent vous permettre de mieux appréhender les outils nécessaires au traitement de vos données, et vous aider à construire des workflows qui vous serez ensuite en mesure de paramétrer et de lancer sur vos propres jeux de données.

10/ Les mouvements dans l'équipe

Céline Noirot a réintégré l'équipe après un passage au CHU de Toulouse.

Pour toute demande d'information ou de travaux, veuillez remplir le formulaire adéquat sur la page : <http://bioinfo.genotoul.fr/index.php/ask-for/support/>.