

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 et demi de formation alignement de séquences génomiques et recherche de variations sous Galaxy seront organisés du 16 au 18 novembre 2015 (Reste 7 places)

Il s'agit de 2,5 jours 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 17 et le 18 au matin (09h00 - 12h00) vous permettront d'effectuer un alignement de séquences et une recherche de polymorphisme sous environnement Galaxy avec la suite GATK.

B/ 3 jours de formation à la métagénomique 16S et 18S sous Galaxy seront organisés du 2 au 4 décembre 2015 (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 ? Nous vous proposons pour cela une formation de 3 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. La suite de la formation (2 jours) sera consacrée à la prise en main du pipeline FROGS (http://bioinfo.genotoul.fr/fileadmin/user_upload/FROGS_poster_Jobim_2015.pdf) constitué des étapes principales suivantes : nettoyage, clustering, affiliation taxonomique, filtre et statistiques descriptives.

Ces 2 cycles ne nécessitent aucune connaissance préalable de la ligne de commande.

C/ Formations en ligne de commande du 1er au 5 février 2015

Vous souhaitez mieux appréhender le cluster de la plateforme pour pouvoir analyser de gros jeux de données ? Vous avez besoin de connaître les commandes de base Unix, et de lancer des jobs sur le cluster ? Vous souhaitez découvrir l'alignement de séquence, la détection de SNP et / ou l'analyse RNAseq ? Nous proposons une semaine de formation composé de 4 modules séparés qui vous permettra de mettre en œuvre ce type d'analyse en ligne de commande : Unix / Linux (le 1er février), cluster (le 2 février), alignement de séquences génomiques et recherches de variations (le 3 et 4 février au matin), analyse RNAseq (sans la partie biostat, celle-ci pourra être proposé quelques mois plus tard en fonction de la demande) (le 4 février après-midi et le 5 février).

Pour tous nos cycles d'apprentissage :

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

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>.

D/ La formation permanente INRA propose une formation « Analyses bioinformatiques (sous Galaxy) et biostatistiques des données RNAseq pour initiés à R (sous Rstudio) »

Il s'agit d'une action de formation à l'initiative de la cellule bioinformatique pour laquelle il reste 3 places. Le personnel INRA est prioritaire. Elle se déroulera à Toulouse du 22 septembre 2015 à 14 heure jusqu'au 25 septembre 12 heure. Les inscriptions se font via ce formulaire : <http://cvip.sphinxonline.net/v4/s/rgyju>. Pour tous renseignements complémentaires, merci de prendre contact par e-mail : Sylviane.Fortun@nantes.inra.fr.

2/ Nombre de cœurs de calcul maximum et de mémoire instantanés autorisés sur le cluster de calcul

Le tableau ci-dessous présente le nombre maximum de cœurs de calcul instantanés autorisés actuellement sur les queues selon le groupe d'appartenance des utilisateurs. Il y a des quotas globaux par groupe (contributeurs, INRA/Région et autres) et par utilisateur. Ces valeurs peuvent subir des ajustements en fonction de l'évolution de l'utilisation du cluster.

Nb max cœurs autorisés	Groupe « contributeurs »		Groupe « INRA ou Région »		Groupe « Autres »		Nb slots disponibles
	groupe entier	1 utilisateur	groupe entier	1 utilisateur	groupe entier	1 utilisateur	
workq	4120	1024	3264	512	1088	256	4352
unlimitq	680	256	128	48	32	8	680

Pour réserver plusieurs cœurs : `qsub -pe parallel_smp n myscript.sh`.

Ce tableau se décline en multipliant chaque chiffre par 8 pour obtenir la quantité maximale de RAM utilisée de façon

instantanée (Voir ci-dessous).

RAM max autorisés (G)	Groupe « contributeurs »		Groupe « INRA ou Région »		Groupe « Autres »		Nb Giga RAM disponibles
	groupe entier	1 utilisateur	groupe entier	1 utilisateur	groupe entier	1 utilisateur	
workq	32960	8192	26112	4096	8704	2048	34816
unlimitq	5440	2048	1024	384	256	64	5440

Pour réserver plus de RAM : `qsub -l mem=xG -l h_vmem=yG myscript.sh` (par défaut `mem=1G` et `h_vmem=8G`).

3/ Quelles sont les queues disponibles sur le cluster ?

Le tableau ci-dessous présente les différentes queues disponibles, leur accessibilité, leur priorité (utilisé dans le calcul de la priorité des jobs), le temps maximum d'un job et le nombre total de cœurs de chacune d'elles.

Queue	Accessibilité	Priorité	Durée maximale en heure	Nb max de cœurs
workq	Tous les utilisateurs	300	96	4120
unlimitq	Tous les utilisateurs	100	illimité	680
smpq	Sur demande	0	illimité	240
hypermemq	Sur demande	0	illimité	96
interq (qlogin)	Tous les utilisateurs	100	48	40

La durée maximum d'un job sur la workq est donc passé de 48 à 96 heures.

4/ Quoi de neuf sur l'instance toulousaine de Galaxy ?

A/ De nouveaux outils :

De nouveaux outils ont été rajoutés dans l'instance Galaxy toulousaine accessible ici : <http://sigenae-workbench.toulouse.inra.fr>.

- Pour traiter les données Miseq ou 454 de métagénomique 16S / 18S, le pipeline FROGS (http://bioinfo.genotoul.fr/fileadmin/user_upload/FROGS_poster_Jobim_2015.pdf) a été implémenté et encapsulé dans Galaxy.
- Dans le cadre du groupe de travail "statistiques" du CATI BIOS4BIOL, des outils d'analyse statistiques ont été ajoutés dans l'instance toulousaine de Galaxy (rubrique Graph/Display Data) tel que l'outil "hierarchical clustering" qui permet de générer un graphique de clustering (ou regroupement) hiérarchique à partir d'une table de comptage. Les outils sont associés à une documentation complète.
- L'outil "List my work on Genotoul" vous permet de lister l'ensemble de vos répertoires et fichiers contenus dans votre /work/ afin de connaître facilement le chemin d'accès à un fichier donné et l'outil "Show current jobs running" vous permet de suivre l'évolution de vos jobs sur le cluster.

B/ La version 3 de l'école Galaxy4Bioinformatics se tiendra à Toulouse du 2 au 5 novembre 2015 :

La version 3 de l'Ecole Thématique Galaxy4Bioinformatics « DEVELOPPEMENT ET INTEGRATION D'APPLICATIONS SOUS GALAXY » se tiendra sur Toulouse pendant 4 jours : du lundi 2 novembre 14 heure au jeudi 5 novembre.

L'école s'articulera autour de 4 axes principaux (les interventions seront assurées par les membres du Groupe de Travail Galaxy France) : présentation des concepts de workflows et de l'architecture Galaxy, atelier sur les bonnes pratiques de développements et d'intégration d'outils, perspectives : interopérabilité, API, web services, etc. et atelier «Toolshed» ou entrepôt d'outils Galaxy.

5/ La journée bioinfo/biostat régionale du 25 juin dernier fut un succès

Cette journée a réuni plus de 100 personnes autour des thématiques de bioinformatique et de biostatistiques. Le programme et les résumés sont disponibles ici : <http://bioinfo.genotoul.fr/?id=160>. Les présentations seront bientôt mises en ligne.

6/ Remerciements et Publications

Vous êtes de plus en plus nombreux à utiliser l'infrastructure de Genotoul Bioinfo et à la citer dans vos publications. Nous vous en remercions. Le nombre de publications remerciant la plateforme 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 de la plateforme : "We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing help and/or computing and/or storage resources".

Dans le cadre d'une collaboration, vous pouvez directement citer la / les personnes qui ont participé au projet de la manière suivante : Name, bioinformatics platform Toulouse Midi-Pyrenees, MIAT UR 875, INRA Auzeville CS 52627 31326 Castanet Tolosan cedex.

Nous vous invitons à nous communiquer les références de ces publications via l'adresse mail : anim.bioinfo@toulouse.inra.fr.

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