

# Lettre d'information

Cette lettre d'information est destinée aux membres des équipes de recherche utilisant la plateforme 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.



## Arrêt de production

Du dimanche 26 juin au soir au lundi 27 juin un arrêt de production est prévu sur les serveurs de login et l'ensemble du cluster de calcul pour des mises à jour du système et des contrôles d'intégrité. Nous nous sommes rendus compte d'un dysfonctionnement au niveau de la gestion des quotas (avéré par notre fournisseur) qui a des répercussions potentielles sur l'ensemble de nos utilisateurs.

Nous en profiterons pour mettre à jour la version de SLURM en raison d'une faille de sécurité et pour renouveler des certificats de sécurité de serveurs. Les services concernés sont les suivants : Serveurs de login (genologin1 & 2), Cluster de calcul (y compris genoview, genosmp02 & 03), Espaces disques /work, Banques de données génomiques, Machines virtuelles (dont galaxy) qui utilisent le /work



## Empreinte Carbone

Lorsque vous vous loguez sur genologin l'équivalent carbone de votre consommation du cluster de calcul est affichée. Le quota mentionné est de 2 tonnes équivalent CO<sub>2</sub>e par an, il s'agit de l'objectif à atteindre par personne pour toutes nos activités d'ici à 2050 pour respecter les engagements de l'Accord de Paris : maintenir l'augmentation de la température mondiale à un niveau inférieur à 2 degrés par rapport au niveau pré-industriel.

## Bilan enquête 2021

### Retours sur l'enquête de satisfaction

Vous avez été nombreux (158 soit près de 16%) à répondre à notre questionnaire de satisfaction annuel et nous vous en remercions chaleureusement. Suite à vos retours nous avons rajouté des items à notre FAQ, par exemple, concernant l'utilisation de MobaXterm pour se connecter au cluster. Certaines de vos autres suggestions sont à l'étude et pourront peut-être être envisagées après la mise en production du nouveau cluster que nous sommes en train d'acquérir. A ce propos vous nous avez fait part de votre besoin en GPU, nous en avons tenu compte dans ce nouveau projet d'acquisition.

En ce qui concerne le test des logiciels mis à disposition, vous nous avez mentionné que vous souhaiteriez avec un exemple d'utilisation des outils installés. Ces exemples simples sont très souvent disponibles dans le répertoire d'installation du logiciel (/usr/local/bioinfo/src/NomDuLogiciel/). Il s'appelle exemple\_on\_cluster. Mais, si vous souhaitez l'utiliser, il vous faut obligatoirement le copier dans votre work sinon vous aurez un problème de permission.

Certains d'entre vous souhaiteraient pouvoir partager des répertoires avec des personnes appartenant à un autre groupe unix. C'est possible via les ACL (commande `nfs4_setfacl`). Des exemples d'utilisation sont disponibles dans la FAQ Linux dans la question « How to change permission on file or folder ».

Si vous avez besoin de davantage d'espace work, il est possible d'en louer, les tarifs et le formulaire de « demandes exceptionnelles » sont disponibles sur notre site web.

Vous nous avez mentionné que la fermeture des comptes à la fin des contrats des non-permanents pouvait poser problème pour récupérer les données. Nous sommes désolés mais pour des raisons légales, nous ne pouvons pas mettre l'encadrant en copie des mails envoyés pour prévenir de la fermeture du compte.

N'hésitez pas à nous poser vos questions via le formulaire : <http://bioinfo.genotoul.fr/index.php/ask-for/support/> ou le mail du support : [support.bioinfo.genotoul@inrae.fr](mailto:support.bioinfo.genotoul@inrae.fr). Vous pouvez, bien sûr, nous écrire en français.



## Qu'est-ce qu'un cluster et pourquoi l'utiliser ?

Un cluster est un ensemble de serveurs de calcul appelés nœuds qui sont interconnectés. Les nœuds peuvent être différents les uns des autres, par exemple il y a un nœud de login, des nœuds de calcul standards, des nœuds à forte mémoire, des nœuds GPU.... Il est possible de se connecter aux nœuds de login via ssh. Notre cluster est décrit sur notre site web sur cette page (pour avoir plus de détail, cliquer sur chaque composant de l'image) :

<http://bioinfo.genotoul.fr/index.php/ressources-2/ressources/>

Notre cluster est une ressource partagée entre plus d'un millier d'utilisateurs.

Il est utile d'utiliser un cluster de calcul plutôt qu'un ordinateur personnel dans les cas suivants :

- besoin de plus de ressources que disponible sur son PC (RAM, CPU, espace disque)
- le même job doit être lancé plusieurs fois sur différentes données (possibilité de le lancer en parallèle sur plusieurs nœuds)
- le programme que vous souhaitez lancer gagnerait à être « multi-threadé » via MPI ou OpenMP.

Cependant, l'utilisation d'un cluster ne rend pas toujours le calcul plus rapide. Il peut arriver que votre job sur le nœud du cluster partage les ressources avec d'autres jobs : CPU, mémoire, réseau d'accès aux espaces de stockage... Nous observons par exemple des ralentissements possibles sur les I/O (lecture/ écriture). Il faut, par conséquent, faire attention à cette contrainte lorsque vous dimensionnez vos jobs, par exemple éviter de lire des milliers de fois le même fichier en parallèle.

Lorsque vous lancez vos jobs, ils seront tout d'abord placés en file d'attente et leur priorité de passage sera calculée par l'ordonnanceur et leur lancement effectif dépendra également des ressources disponibles.

## Cycles d'apprentissage

Des places disponibles sur nos formations 2022

En partenariat avec Sigenae et SaAB (MIAT), nous vous proposons 6 cycles d'apprentissage pour lesquels il reste encore quelques places.

Sur notre site internet, vous trouverez les [tarifs](#) et le [formulaire d'inscription](#).

Ils se feront en présentiel sauf nouvelles contraintes sanitaires.

Vous pourrez trouver d'autres cycles d'apprentissage en bioinformatique sur le site de l'IFB (Institut français de bioinformatique).

	Date	Pré-requis	Places disponibles
<b>Linux</b>	10 octobre	Aucun	1 place
<b>Cluster</b>	11 octobre	Linux	4 places
<b>Using sed and AWK to modify large text files</b>	12 octobre	Linux	10 places
<b>First step in AWK programming</b>	13 octobre	Linux	10 places
<b>Short-read alignment and SNP calling</b>	14 – 15 novembre	Linux et Cluster	8 places



### Le coin des débutants

Si vous êtes complètement débutant avec la ligne de commande Linux, nous vous recommandons les supports pédagogiques suivants :

- Un cours très bien fait : <https://swcarpentry.github.io/shell-novice/>. Pour tout préparer en amont : <https://swcarpentry.github.io/shell-novice/setup.html>
- Un jeu sur le web : <https://luffah.xyz/bidules/Terminus/>
- Un jeu pour s'entraîner utilisable sur genologin : <https://github.com/phyver/GameShell/blob/master/doc/gameshell.md>
- Pour aller plus loin : voici un outil vous permettant de vérifier vos scripts shell : <https://github.com/koalaman/shellcheck#user-content-in-your-editor>

De plus, nos supports de formations sont disponibles sur notre site web sur la page de description de chaque formation.



# Nouvelle version de Galaxy

La fin de la migration approche

Depuis le début de l'année 2022, l'ancien serveur Galaxy v16.05 a été remplacé par un nouveau serveur Galaxy v21.05. L'ancien serveur n'est plus disponible pour y effectuer des traitements, mais uniquement pour exporter des workflows. Une machine supplémentaire a été mise en place pour l'export des historiques, permettant de les rendre compatibles avec la nouvelle version de Galaxy.

Les procédures d'export/import et les accès aux différents serveurs sont disponibles à partir de la page : [https://vm-galaxy-prod.toulouse.inra.fr/Galaxy\\_menu/galaxy-sigenae.html](https://vm-galaxy-prod.toulouse.inra.fr/Galaxy_menu/galaxy-sigenae.html)

Nous pouvons accompagner les utilisateurs dans la migration de leurs données. Contactez le support de l'équipe Sigenae si besoin. Nous comptons actuellement 198 comptes utilisés sur le nouveau serveur. Nous rappelons que les espaces de stockage de l'ancien serveur et du nouveau sont différents.

**Nous vous rapellons que les données ou les workflows qui ne seront pas migrés d'un serveur vers l'autre seront perdus lorsque l'ancien serveur sera définitivement arrêté (fin juin).** Comme dans toute migration de services, certains outils ont vu leur fonctionnement altéré, et des correctifs ont été apportés pour la grande majorité.

32 utilisateurs de Galaxy (soit environ 10%) ont répondu à notre dernière enquête de satisfaction annuelle et nous vous en remercions. Nous retenons que 2/3 des personnes ayant répondu ont suivi une formation sur Galaxy. Globalement, les utilisateurs sont satisfaits. De plus, vous nous avez fait part de neuf publications issues de l'usage de Galaxy.

Sur le plan des outils, FROGS est l'outil majoritairement utilisé. Les autres usages de Galaxy portent sur l'analyse de données RNASeq, l'alignement et l'analyse de SNP.

Pour toute demande telle que l'intégration d'un nouvel outil, contactez l'équipe support de Galaxy à l'adresse suivante : [support.sigenae@inrae.fr](mailto:support.sigenae@inrae.fr). Cela va déclencher un ticket qui sera tracé ce qui permettra à l'équipe Galaxy de répondre à vos besoins le plus efficacement possible. L'équipe étant en effectif réduit, nous traitons les tickets par ordre de priorité. Ne pas hésiter à indiquer sur le ticket si le problème rencontré est bloquant. Trois autres instances peuvent éventuellement servir de relais : <https://usegalaxy.fr/>, <https://usegalaxy.eu/>, et <https://usegalaxy.org/> .



## Trucs et astuces sur le cluster

- Pour suivre un job en cours sur le cluster vous pouvez utiliser les commandes `squeue` ou `scontrol` comme indiqué dans la FAQ (Job submission : « How can I monitor a running job »). Mais si vous souhaitez savoir en direct ce qu'il se passe sur le nœud vous pouvez vous connecter sur le nœud sur lequel tourne le job avec la commande `srunc -p --pty --jobid=XXX bash`. Vous pouvez ensuite utiliser la commande `top`, ou faire directement `srunc -p --pty --jobid=XXX top -u login`. Nous avons rajouté une FAQ sur le sujet dans la catégorie Job submission (« How to use srunc to check running jobs »).

- Pour utiliser un logiciel, il faut charger son environnement permettant d'avoir accès à ses dépendances. Nous avons mis en place des scripts permettant de trouver plus facilement les modules utiles. Par exemple, pour trouver le nom d'un module : `search_module soft_name` (ou partie du nom du logiciel).

De la même façon pour chercher les packages R installés, nous mettons un script à votre disposition. Il s'utilise ainsi : `search_R_package package_name`.

Et enfin, le même type de script existe pour les packages python : `search_Python_package package_name`.

Les trois scripts sont mentionnés dans notre FAQ à la section software sur ce lien : [http://bioinfo.genotoul.fr/index.php/faq/software\\_faq/](http://bioinfo.genotoul.fr/index.php/faq/software_faq/)