

# Lettre d'information

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 cycles d'apprentissage mis en place.

## Besoin de davantage de RAM ?

Nous avons très récemment doublé la mémoire RAM sur 15 nœuds du cluster. Il s'agit des nœuds [134-148].

Voici donc comment se compose le cluster de calcul actuellement :

	Nb threads	RAM	Autre information
<b>node001 à 068</b>	40 chacun	256Go chacun	Ivy (cluster de 2014)
<b>Node101 à 148</b>	64 chacun	256Go ou 512Go	Broadwell (Cluster de 2017)
<b>genosmp02</b>	96	1,5 To	Accessible sur demande
<b>genosmp03</b>	192	3To	Accessible sur demande
<b>genoview</b>	64	128Go	Carte GPU NVIDIAK40 Accessible sur demande

De plus parmi la RAM disponible, 4G de RAM sont exclusivement réservés sur chacun des nœuds de calcul pour le bon fonctionnement du système.

Vous retrouverez ces informations grâce à la commande "scontrol show node" depuis les nœuds de login.

### Nous recrutons !

La plateforme Bioinfo Genotoul recrute un ingénieur en développement d'application au concours externe :

<https://jobs.inrae.fr/concours/concours-externe-s-ingenieurs-cadres-techniciens-h-f/ie21-mathnumm-1>



## Comment faire passer vos jobs plus vite sur le cluster ?

Les ressources du cluster sont partagées entre tous les utilisateurs. Nous constatons aussi beaucoup de sur-réservation de mémoire (mémoire réservée, mais non utilisée par le job), ce qui ralentit le passage des gens sur-réservant tout en privant les autres de ces ressources non utilisées. Pour diminuer le temps d'attente de vos jobs il est nécessaire d'ajuster au plus près les ressources réservées (CPU, RAM, temps). Pour vous aider voici quelques conseils :

- au sujet des réservations de mémoire RAM sur le cluster (option `--mem`) : pour vérifier vos besoins, utilisez la commande `seff <jobid>` sur 1 job test en état COMPLETED.

- bien ajuster le nombre de cpus/coeurs/threads demandés par votre programme avec ceux réservés avec Slurm (option `--cpus-per-task` de `srun` ou `sbatch`).

- spécifier la durée maximum de votre job en utilisant la commande « `--time=HH:MM:SS` » sans dépasser celle définie par défaut sur la queue utilisée (96h sur la `workq`).

La priorité d'un job est calculée par slurm en utilisant plusieurs facteurs : la taille du job (en CPU et RAM demandés), le temps demandé (plus le job est petit et court plus il passera vite), la partition demandée (la `workq` est prioritaire sur la `unlimitq`) et le temps passé par le job dans la file d'attente (plus le job attend, plus sa priorité augmente).

Nous fonctionnons avec des comptes utilisateurs pour lesquels nous avons mis en place des groupes (« contributeurs », « INRAE ou régions », « autres »). Chaque groupe se partage des quantités de ressources différentes (CPU et RAM), mais il n'y a pas de différences de priorité, au sens de slurm, entre les groupes.



## Cycles d'apprentissage 2021

En partenariat avec Sigeneae, NED (GenPhySE), SaAB (MIAT), nous vous proposons 8 cycles d'apprentissage pour lesquels il reste encore des places.

Nos cycles d'apprentissage qui se déroulent au printemps sont organisés en distanciel en raison du contexte sanitaire. Sur notre site internet, vous trouverez les [tarifs](#) et le [formulaire d'inscription](#).

De plus, les supports de formations (diaporamas et exercices) sont en général disponibles en bas des pages de description des cycles d'apprentissage.

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

Le plateforme de biostatistique toulousaine propose également des formations que vous trouverez [ici](#).

	Date	Pré-requis	Places disponibles
<b>First step in awk programming</b>	25 mars 2021 14 octobre 2021	Linux	1 place en mars 9 places en octobre
<b>How to run a nf-core nextflow workflow on GenoToul Bioinfo ?</b>	26 mars 2021 15 octobre	Linux Cluster	1 place en mars 10 places en octobre
<b>Linux</b>	11 octobre 2021	Aucun	9 places
<b>Cluster</b>	12 octobre 2021	Linux	9 places
<b>Using sed and awk to modify large text file</b>	13 octobre 2021	Linux	9 places
<b>Short reads alignment and small size variants calling</b>	15 – 16 novembre 2021	Linux Cluster	7 places

## Bilan questionnaire de satisfaction 2020

Vous avez été nombreux (169 soit 18%) à répondre à notre questionnaire de satisfaction annuelle et nous vous en remercions chaleureusement.

Certains utilisateurs indiquent qu'ils constatent des temps d'attente anormalement long sur le cluster. En effet, le cluster connaît certains pics d'activités périodiques qui entraînent momentanément des files d'attente plus longues (en sbatch comme en srunch). Le cluster est une ressource partagée et sa disponibilité dépend de son utilisation. Pour diminuer les files d'attente, il est nécessaire d'ajuster au mieux les réservations de CPU et de RAM aux besoins de ses jobs (voir la page précédente de cette lettre d'informations).

N'hésitez pas à nous contacter via les formulaires sur notre site web rubrique « Ask for »: les logiciels, les banques de données sont mis à jour sur demande. Si vous avez des questions même très naïves ou des suggestions, utilisez le formulaire dédié disponible sur notre site web :

<http://bioinfo.genotoul.fr/index.php/ask-for/support/>

Vos remarques nous ont permis de mettre notre FAQ à jour : vous y trouverez de nouvelles questions ou des précisions sur d'anciennes questions. Comme par exemple les questions : « Is a software already installed on the cluster? », « How to find an already install R package? », « Unexpected disconnection and broken pipe message », « Set up your personal web space (public html) », « How to set my cpu and memory pre-allocation? », « How can I get my jobs processed faster on the cluster? ».

De nombreux utilisateurs de Galaxy ont répondu à cette enquête. Nous sommes heureux de constater que cette plateforme est utile à vos projets. La majorité des utilisateurs Galaxy de l'instance BioInfo GenoToul/Sigeneae utilisent le pipeline FROGS (permettant l'analyse de données de metabarcoding). A ce sujet, FROGS évolue, vous pouvez maintenant utiliser DESeq2, les ITS sont complètement pris en charge et Picrust2 est en cours d'intégration. Vous retrouverez les précédentes newsletters Galaxy depuis la page d'[accueil Galaxy](#) en cliquant sur "Galaxy News".

Encore merci à tous de vos retours.