

# Lettre d'information

Cette lettre d'information est destinée aux membres des équipes utilisant la plate-forme bio-informatique GenoToul. Elle a pour but de répondre aux questions et commentaires que vous nous avez fait remonter via le questionnaire de satisfaction annuel et de vous annoncer la prochaine mise en production du nouveau cluster.



## Des nouvelles de l'équipe

La composition de l'équipe bouge ! Tout récemment, Julien Henry nous a rejoint par concours externe. Sa mission principale sera de développer la thématique de l'intégration de données omiques d'un point de vue statistique. Il sera aidé pour cela de Nathalie Vialaneix à 25 % sur la plateforme.

Julien a été élève de l'École Centrale de Lyon et a obtenu un master en statistique. Il a ensuite été « Data Scientist » pendant 5 ans dans plusieurs start-ups en Allemagne. En 2021, il a rejoint INRAE Toulouse en CDD, dans l'unité MIAT, puis dans l'unité GenPhySE.

En ce moment, deux CDD collaborent avec nous. Hanae Chouali, sur le projet ANR Symway, qui a fait des assemblages de génomes et de l'analyse de données RNASeq, terminera son contrat fin septembre pour voler vers de nouveaux horizons. Julien Touchais, quant à lui, est développeur web sur le projet ANR MetRibo et est parmi nous jusqu'à la fin décembre 2024.

Vous pourrez retrouver toute l'équipe sur notre site web à la rubrique « About us » puis « Contact us » : <https://bioinfo.genotoul.fr/index.php/about-us/contact-us/>

## Rappel : migration sur le nouveau cluster



Nous souhaitons conclure la migration du cluster « genologin » vers le nouveau cluster « genobioinfo » fin décembre 2023.

Vous trouverez la présentation faite en mai dernier qui contient de nombreuses informations utiles sur la nouvelle infrastructure et les changements qu'elle induit sur [ce lien](#).

Suite au mail d'information du 5 juillet dernier, nous vous rappelons qu'il n'y a désormais plus de nouvelles installations ni d'augmentation de quotas (disque et heures de calcul) possible sur le cluster « genologin ». De plus, la charge de travail importante des administrateurs système les contraint à déprioriser les demandes faites au support concernant l'ancien cluster par rapport à celles concernant le nouveau.

### Point sur la migration de vos données :

Les données qui sont dans le save n'ont pas besoin d'être transférées. Elles sont partagées entre l'ancienne et la nouvelle infrastructure. Les données qui sont dans le /home et le /work ne sont PAS récupérées. Si vous souhaitez récupérer les données depuis votre espace disque personnel /work/username, alors vous pouvez utiliser la commande suivante (depuis le nouveau serveur genobioinfo ou via slurm : sbatch ou srun) :

```
rsync -avh
/oldwork/username/ /work/user/username/
```

Les données qui sont dans les espaces projets (/work/project) seront transférées par notre équipe. Vous serez contacté-e pour cela au fur et à mesure des renouvellements de contrats ou bien en fonction de vos demandes).

Pour planifier la migration de vos machines virtuelles ou si vous avez des questions ou des difficultés à utiliser la nouvelle infrastructure, n'hésitez pas à nous contacter : [support.bioinfo.genotoul@inrae.fr](mailto:support.bioinfo.genotoul@inrae.fr).



## Achat de stockage « save »

Cet été, nous avons fait évoluer l'infrastructure de stockage des données pérennes (avec réplication sur site distant).

Cela concerne tous les espaces disques autres que /work par exemple /home, /save, etc.

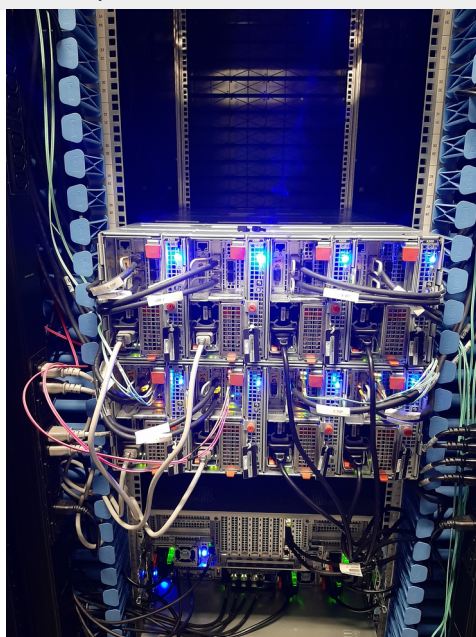
Il s'agissait de remplacer deux baies de disques en fin de maintenance et en profiter pour doubler la capacité.

Nous avons désormais 1.5Po utile (en réplication) sur chacun des 2 sites que nous exploitons (INRAe Auzeville et DROCC Datacenter Régional Occitanie).

Cela permettra d'accueillir de nouveaux projets tout en permettant une meilleure sécurité des données que vous souhaitez conserver.

Ce matériel a été financé grâce au projet EquipEx+ Mudis4LS (*Mutualised Digital Spaces for FAIR data in Life and Health Science*) de l'IFB (Institut Français de Bioinformatique).

Attention : l'espace disque /work n'est pas concerné par cette opération (il s'agit d'un autre équipement spécifique adapté au cluster de calcul mais NON répliqué). Il fera l'objet d'une extension l'année prochaine.



Arrière d'une des deux nouvelles baies de stockage Isilon

## Cycles d'apprentissage

Des places disponibles sur nos formations 2023

En partenariat avec Sigenae, nous vous proposons des cycles d'apprentissage récurrents. En particulier, cet automne une nouvelle formule de la session Linux est proposée. Elle dure 2 jours et est axée sur les commandes principales à connaître dans le cadre de l'utilisation de notre nouveau cluster. La session cluster sera également mise à jour en profondeur pour s'adapter à notre nouvelle infrastructure.

Les *tarifs* sont disponibles sur notre site internet.

La plateforme *GenoToul Biostat* met aussi en place différentes formations.

Si vous ne trouvez pas la formation que vous souhaitez, l'IFB (l'Institut Français de Bioinformatique) référence une grande diversité de cycles d'apprentissage en bioinformatique, comme des sessions sur la phylogénie, ou l'assemblage et l'annotation. Pensez également aux formations permanentes de vos différents instituts. Elles proposent de nombreuses formations en informatique.

Les prochains cycles d'apprentissage que nous proposons sont listés dans le tableau ci-dessous.

Pour vous inscrire, rendez-vous sur [cette page](#).

	Date	Pré-requis	Places disponibles
<b>Linux</b>	16 et 17 octobre	Aucun	9 places
<b>Cluster</b>	18 octobre	Linux	7 places
<b>Short reads alignment and small size variant calling</b>	13 et 14 novembre	Linux et Cluster	7 places
<b>Improve your command line skills by learning a few words on perl</b>	28 novembre	Linux	7 places

## Assemblage de génomes

Publication

La plate-forme est heureuse de vous partager une publication, qui comprend pas moins de six membres de l'équipe ! Il s'agit du premier (d'une longue série, nous l'espérons) *data paper* issu du projet SeqOccln, co-porté avec GeT-PlaGe. Dans cet article vous pourrez mesurer l'intérêt d'utiliser des lectures longues de type PacBio HiFi pour l'assemblage du génome bovin: nous avons ajouté environ 500 méga à l'assemblage de référence !

Voici la référence de cette publication :

*Eché, C., Iampietro, C., Birbes, C. et al. A Bos taurus sequencing methods benchmark for assembly, haplotyping, and variant calling. Sci Data 10, 369 (2023).*

**SeqOccln**