

Lettre d'information

Rédactrice : Claire Hoede

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.



Migration vers SLURM

La plate-forme Bioinformatique GenoToul vous rappelle que le cluster de calcul genotoul/SGE est actuellement en cours de fermeture progressive excepté pour les applications et les machines virtuelles en cours de migration ainsi que pour l'accès à la queue smpq.

Pour vous aider les documents importants sont :

- La [documentation utilisateur SLURM](#).
- Le [support de formation](#) au cluster.
- La [FAQ](#).

Les anciens nœuds de calcul (auparavant sous SGE) sont progressivement intégrés au cluster SLURM et seront disponibles sur les partitions workq, unlimitq sans manipulation particulière.

Si vous souhaitez accéder spécifiquement (pour des besoins particuliers) aux anciens / nouveaux nœuds de calcul, il vous faudra rajouter l'option `--constraint=ivy / broadwell`.

Pour un job « completed » uniquement, nous vous conseillons d'utiliser la commande `seff job_id` pour connaître les ressources telles que la RAM et le temps CPU que votre job a utilisés.

@BioinfoGenotoul
<http://bioinfo.genotoul.fr/>

Comptes utilisateur

Avec plus de **1300 utilisateurs** répartis sur le territoire national, la plate-forme **Genotoul Bioinfo atteint les limites permettant de fournir un service de qualité** (puissance de calcul nécessaire, espace de stockage disponible, support aux utilisateurs...). Cette situation exige la mise en place de mesures exceptionnelles pour garantir la qualité du service rendu et la pérennité de l'infrastructure aux utilisateurs actifs disposant déjà d'un accès à celle-ci.

Cette mesure exceptionnelle **mise en œuvre temporairement** vise à **ne plus ouvrir de nouveaux comptes** utilisateurs jusqu'à revenir au nombre de 1000, valeur seuil évaluée comme permettant un fonctionnement encore intensif mais gérable. Dans un premier temps le retour à ce seuil se fera par le refus ferme de nouvelles demandes d'ouverture de compte (hors collaborateurs directs du personnel de la plateforme) et le départ d'utilisateurs (départ à la retraite, fin de contrat pour les non permanents, pas d'activité sur la plate-forme pendant un an pour les permanents,...).

L'ouverture de nouveaux comptes sera réactivée dès que le nombre d'utilisateurs sera en-dessous de la valeur seuil. La mesure est donc transitoire et sera ré-évaluée en regard de l'obtention de nouveaux moyens et/ou de la mise en place d'une nouvelle organisation.

Plusieurs plateformes de l'Institut National de Bioinformatique (IFB) sont ouvertes à l'accueil de nouveaux utilisateurs. Nous encourageons les utilisateurs ayant besoin de nouveaux comptes à se rapprocher de ces collectifs (Genouest, BiRD, ABIMS, CBIB, IFB-core).

Banques de données

La plateforme maintient plus de 240 banques de données qui ont entre 1 et 7 types d'indexations. Nous souhaitons faire un état des lieux de l'utilité de l'ensemble de ces banques. Pour cela nous vous remercions par avance de bien vouloir répondre au questionnaire suivant: <https://enquetes.inra.fr/index.php/survey/index/sid/577292/newtest/Y/lang/fr>.

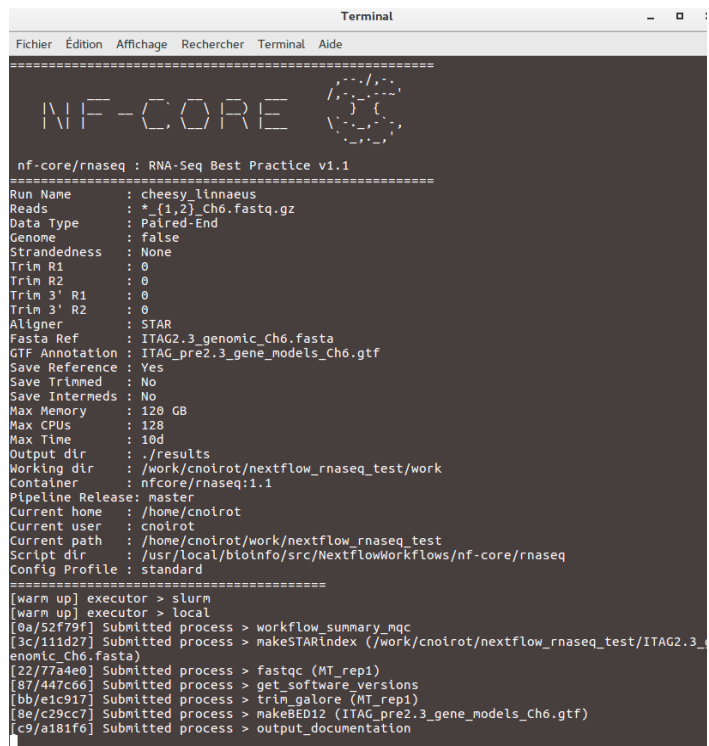
nextflow

Le gestionnaire de workflow [nextflow](#) et un workflow RNASeq sont disponibles pour tous sur [genologin](#) !

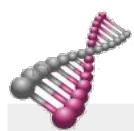
Nous avons mis à disposition le [workflow RNASeq](#) (FastQC, TrimGalore, STAR ou HISAT2, featureCounts, StringTie...) expertisé par la communauté [nf-core](#). Une très bonne documentation est disponible sur leur site.

Pour l'utiliser, à l'instar de tous les autres outils, vous avez une page dédiée sur notre site web et un répertoire contenant des données et lignes de commande d'exemples `/usr/local/bioinfo/src/NextflowWorkflows/example_on_cluster`

Merci de nous faire remonter tout type de problème.



```
Terminal
-----
NF-CORE
nf-core/rnaseq : RNA-Seq Best Practice v1.1
-----
Run Name      : cheesy_linnaeus
Reads         : * (1,2) Ch6.Fastq.gz
Data Type     : Paired-End
Genome        : false
Strandedness  : None
Trim R1       : 0
Trim R2       : 0
Trim 3' R1    : 0
Trim 3' R2    : 0
Aligner       : STAR
Fasta Ref     : ITAG2.3_genomic_Ch6.fasta
GTF Annotation : ITAG_pre2.3_gene_models_Ch6.gtf
Save Reference : Yes
Save Trimmed  : No
Save Interneds : No
Max Memory    : 120 GB
Max CPUs      : 128
Max Time      : 10d
Output dir    : ./results
Working dir   : /work/cnoirot/nextflow_rnaseq_test/work
Container     : nfcore/rnaseq:1.1
Pipeline Release: master
Current home  : /home/cnoirot
Current user  : cnoirot
Current path  : /home/cnoirot/work/nextflow_rnaseq_test
Script dir   : /usr/local/bioinfo/src/NextflowWorkflows/nf-core/rnaseq
Config Profile : standard
-----
[warn up] executor > slurm
[warn up] executor > local
[0a/52f79f] Submitted process > workflow_summary_mqc
[3c/111d27] Submitted process > makeSTARindex (/work/cnoirot/nextflow_rnaseq_test/ITAG2.3_
enomic_Ch6.fasta)
[22/77a4e0] Submitted process > fastqc (MT_rep1)
[87/447c66] Submitted process > get_software_versions
[bb/e1c917] Submitted process > trim_galore (MT_rep1)
[8e/c29cc7] Submitted process > makeBED12 (ITAG_pre2.3_gene_models_Ch6.gtf)
[c9/a181f6] Submitted process > output_documentation
```



FAQ Bioinfo tips

Dans la catégorie « Bioinfo Tips » vous trouverez deux nouvelles FAQ pour vous aider à générer vos fichiers de commandes pour un « array of job ».

L'une à partir de commandes bash, par exemple pour une commande fastQC :

```
for i in `ls *.fastq.gz`; do echo "module load bioinfo/FastQC_v0.11.7; fastqc $i"; done > mescommandesfastqc.sh
```

L'autre à partir d'un script one-line perl. Exemple pour générer le fichier de commande STAR sur des fichiers de lectures pairées :

```
ls /directory/*.fastq.gz | paste - - | perl -lane '($ech)=$F[0]=~/(*).fastq.gz;/ print "module load XXX; STAR --genomeDir star-index --readFilesIn $F[0] $F[1] --outFileNamePrefix $ech ..." > mesSTARcommandes.sh
```

Toutes les explications détaillées de ces lignes sont accessibles ici : http://bioinfo.genotoul.fr/index.php/faq/bioinfo_tips_faq/.



Les logiciels

Nous avons enrichi la documentation liée aux logiciels installés sur genologin. Vous disposez maintenant d'un fichier **How_to_use** accessible dans le répertoire d'installation du logiciel et via [la page logiciel](#) de notre site web. Ce fichier pointe vers un répertoire d'exemples contenant un exemple de script permettant de lancer le logiciel en batch. L'installation des logiciels nécessite par conséquent plus de temps qu'auparavant, c'est pourquoi nous vous conseillons d'anticiper davantage les demandes d'installation de logiciels via le [formulaire](#).

La page [software](#) de la FAQ a été enrichie pour vous faciliter la prise en main de la nouvelle façon d'appeler les logiciels.

A l'occasion de la migration sur slurm, l'environnement Galaxy mis à disposition par l'équipe [Sigenae](#) va évoluer au **printemps 2019**. La principale nouveauté sera la création de plusieurs serveurs Galaxy dédiés à des types de traitements. Ce nouvel environnement cohabitera avec l'ancienne pour **une période de 3 mois**, pour permettre aux utilisateurs de basculer sur la nouvelle infrastructure (données et outils).

Cycles d'apprentissage

1^{er} trimestre 2019

En partenariat avec Sigenae, NED (GenPhySE), SaAB (MIAT) and TWB, nous vous proposons 5 cycles d'apprentissage, 4 en ligne de commande et 1 sous Galaxy :

	Date	Pré-requis	Places disponibles
RNAseq denovo assembly	12 février	Linux	4 places
Linux	11 mars	Aucun	2 places
Cluster	12 mars	Linux	1 places
Modify and extract information from large text files	13 mars	Linux	9 places
Metagenomic: amplicons and stats with FROGS	18 – 21 mars	Aucun	5 places

Nos formations sont organisées sur le site INRA de Toulouse Auzeville. Sur notre site vous trouverez [les tarifs](#) et le [formulaire d'inscription](#).

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).



Vos publications

Vous trouverez sur le site web la liste des publications ayant utilisé l'infrastructure bioinfo genotoul.

Si vous avez publié en utilisant nos ressources et que votre article n'y apparaît pas merci de nous en communiquer les références par mail à anim.bioinfo-Occitanie-Toulouse@inra.fr



Nous vous souhaitons une excellente année 2019 !

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 est un indicateur de notre utilité pour nos demandes de nouvelles ressources. 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-Pyrennees 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 de la manière suivante : Name, INRA, UR 875 Unité de Mathématique et Informatique Appliquées, Genotoul Bioinfo, Auzeville, CS 52627, 31326 Castanet Tolosan cedex.