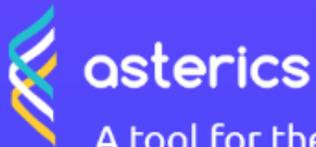


une image = un projet

JACATI BIOS4Biol 2024



pipelines, développements,
benchmarking



asterics

A tool for the exploration
and integration of omics data

ASTERICS is an online tool designed to help you to perform
your statistical and integrative analyses in an interactive
and easy-to-use way.



[+ Create new workspace](#) [Load workspace](#)

Load TCGA demo data into your workspace

I confirm that I have read and that I agree to Asterics' Privacy Policy.

[+ Create new workspace](#)

 Your unique identifier will allow you to access your work and data during 30 days after the last activity on
ASTERICS.



A tool for the exploration and integration of omics data

[Documentation](#)

Software | [Open access](#) | Published: 18 October 2023

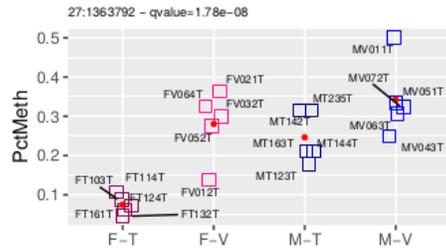
Asterics: a simple tool for the Exploration and Integration of omiCS data

[Élise Maigné](#), [Céline Noirot](#), [Julien Henry](#), [Yaa Adu Kesewaah](#), [Ludovic Badin](#), [Sébastien Déjean](#), [Camille Guilmineau](#), [Arielle Krebs](#), [Fanny Mathevet](#), [Audrey Segalini](#), [Laurent Thomassin](#), [David Colongo](#), [Christine Gaspin](#), [Laurence Liaubet](#) & [Nathalie Vialaneix](#) 

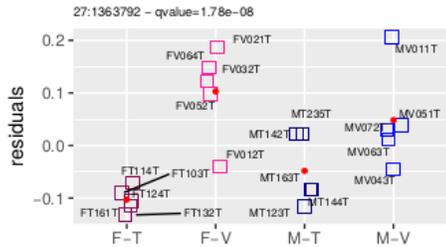
BMC Bioinformatics **24**, Article number: 391 (2023) | [Cite this article](#)

Analyse RRBS (Reduce Representation Bisulfite Sequencing) à 2 facteurs

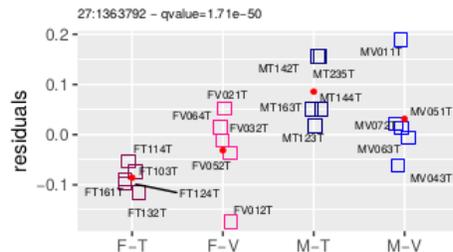
Données brutes



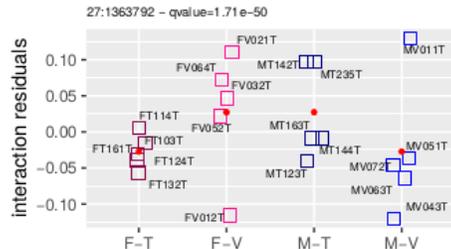
Effet milieu :
Résiduelles PctMeth~sexe



Effet sexe :
Résiduelles PctMeth~milieu



Interaction :
Résiduelles PctMeth~sexe+milieu



Utilisation du paquet R DSS

ajustement d'un modèle linéaire ($\sim 0 + \text{sexe} + \text{milieu} + \text{sexe}:\text{milieu}$)
à l'aide de la fonction `DMLfit.multiFactor`,

test des effets et interaction
avec la fonction `DMLtest.multiFactor`,

puis utilisations des fonctions `callDML` et `callDMR`

Perte des niveaux de méthylation dans les différentes conditions



Auteurs : Anne Frambourg, Véronique Duranthon, Valentin Costes, Luc Jouneau

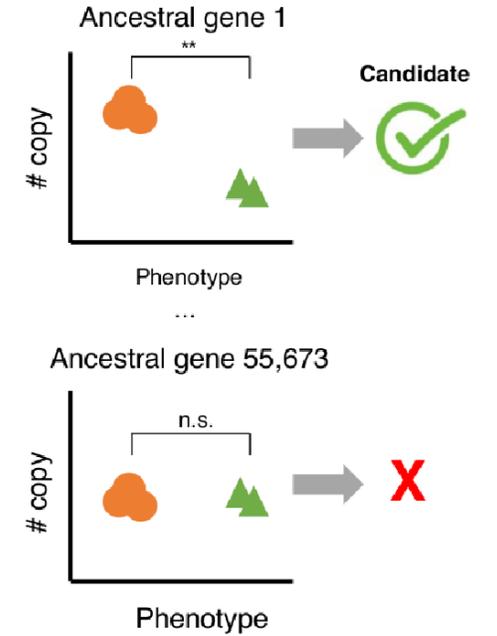
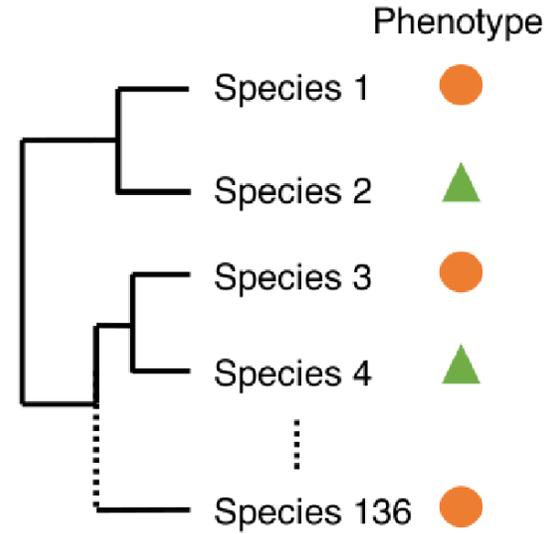
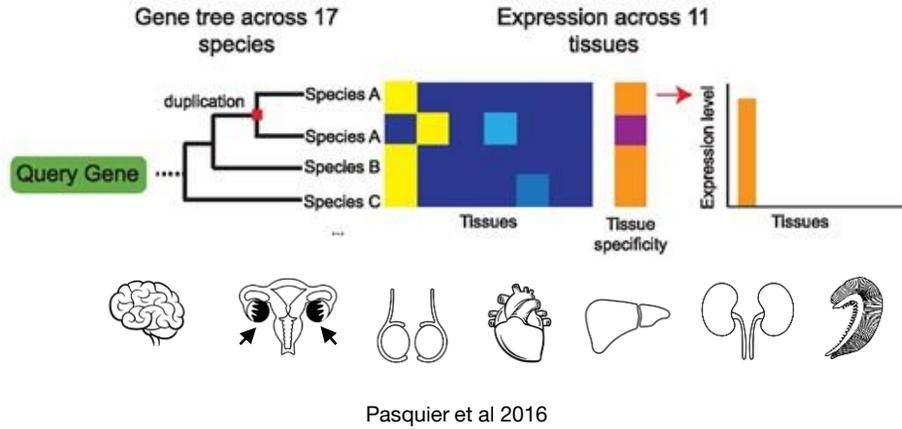


Expression

Evolution des génomes poissons et shiny app



Duplications

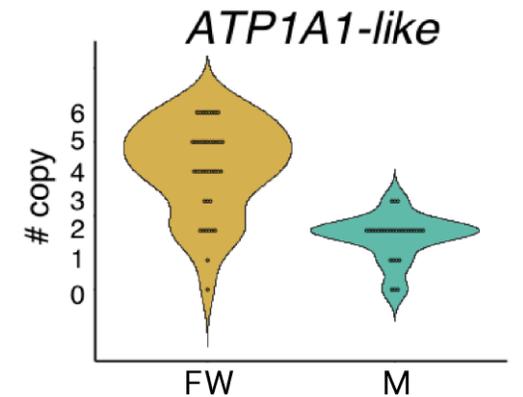
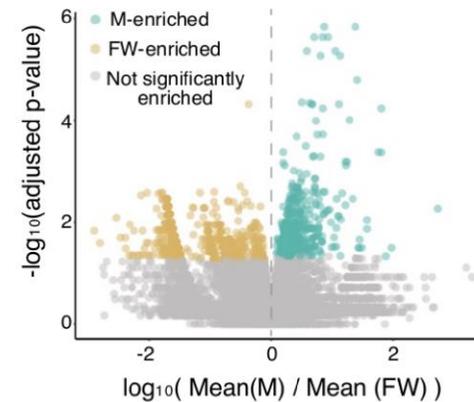


Nucleic Acids Research, 2024, 52, W65–W69
<https://doi.org/10.1093/nar/gkae264>
 Advance access publication date: 8 April 2024
 Web Server issue



FEVER: an interactive web-based resource for evolutionary transcriptomics across fishes

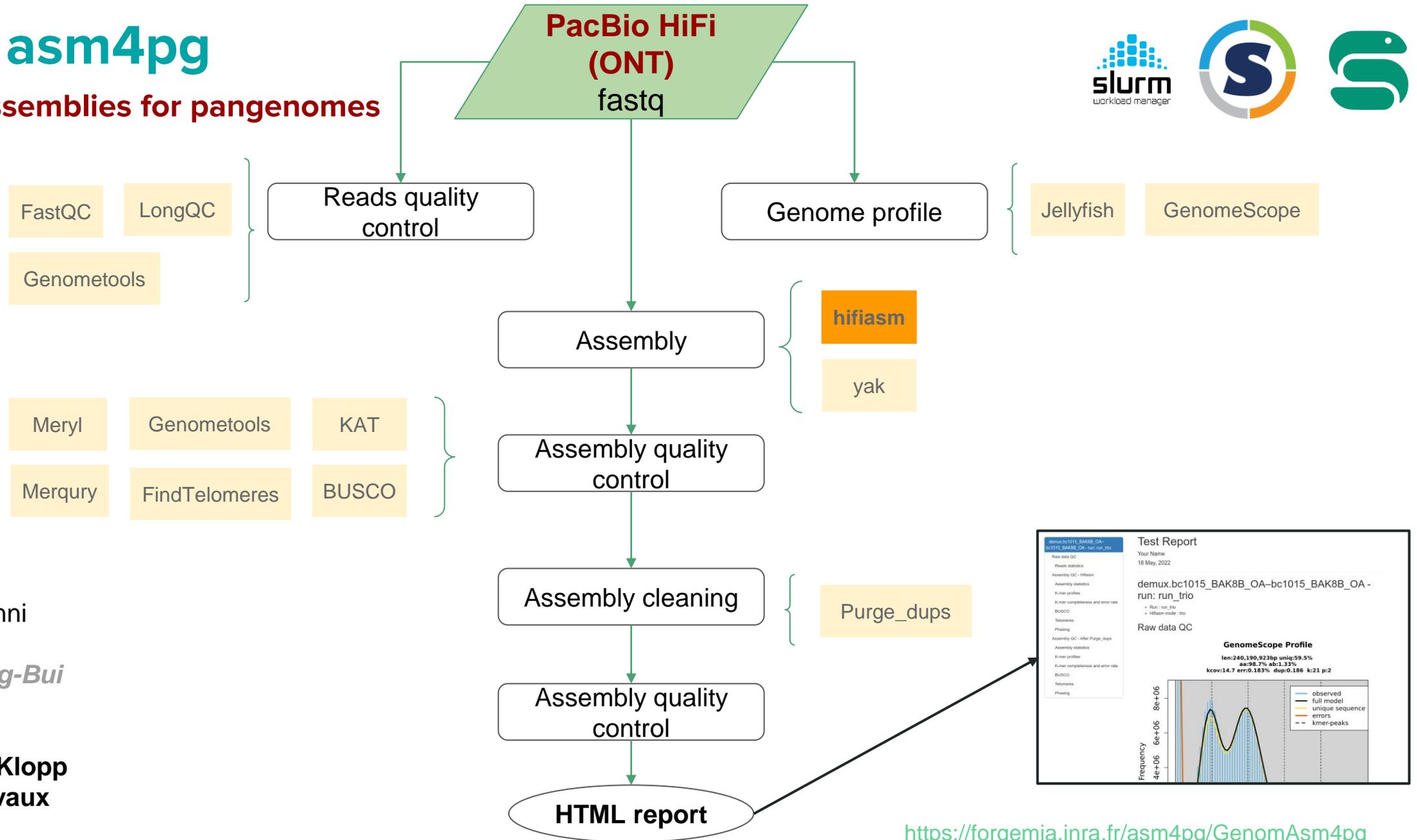
Jérôme Montfort , Francisca Hervas-Sotomayor , Aurélie Le Cam and Florent Murat *



Jérôme Montfort, Florent Murat, Cédric Cabau, Christophe Klopp, team sk8

asm4pg

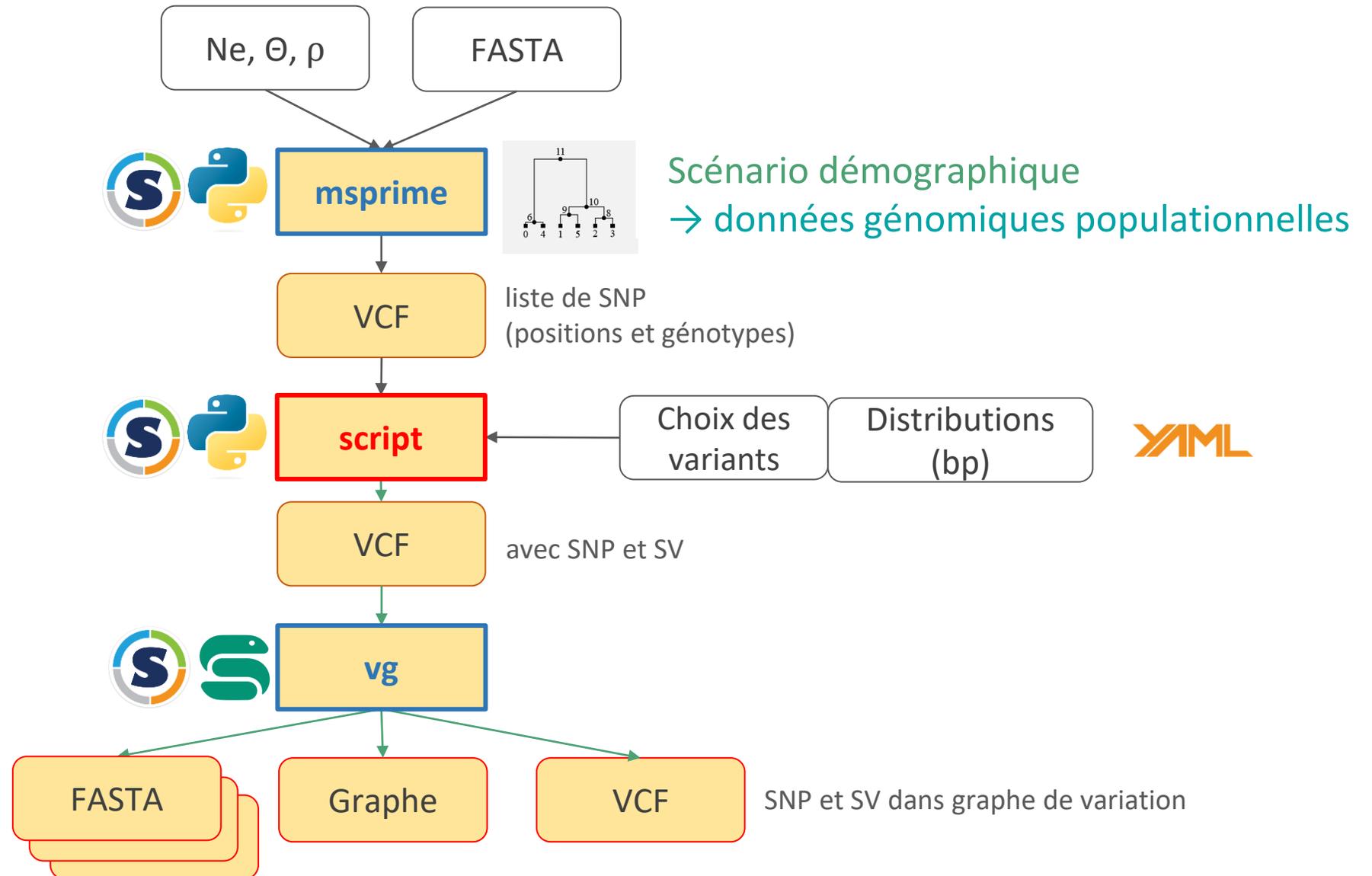
Genome assemblies for pangenomes



Auteurs :
Sukanya Denni
Lucien Piat
Quynh Trang-Bui
Smith Ken
Wu Chen
Christophe Klopp
Ludovic Duvaux

<https://forgemia.inra.fr/asm4pg/GenomAsm4pg>

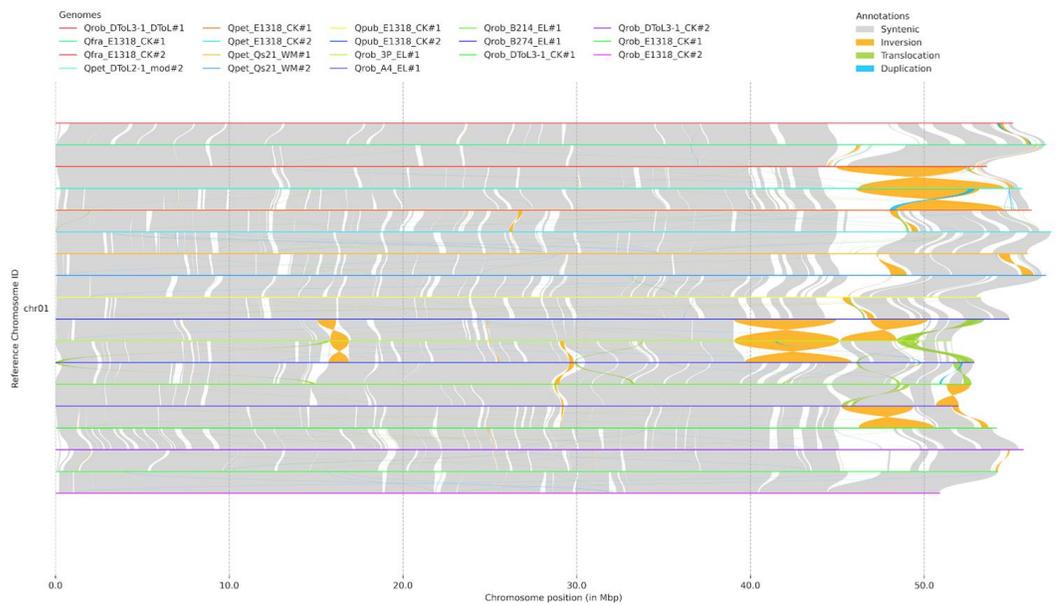
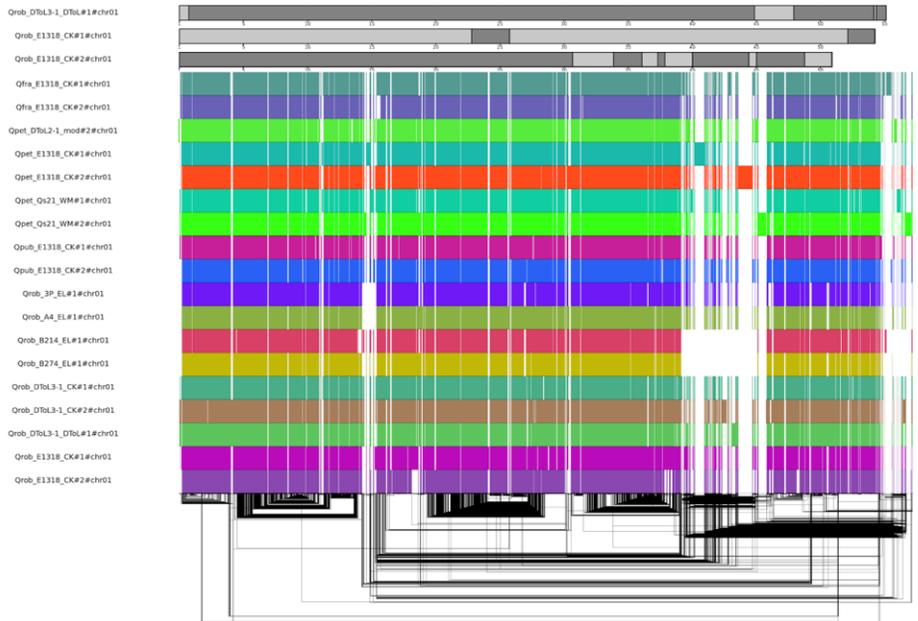
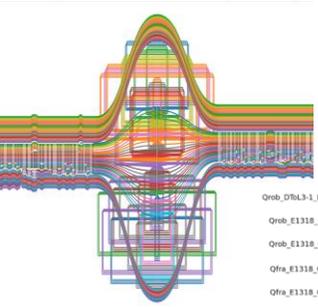
MSpangepop: simulating genomes with structural variants under a population coalescent



Auteurs :
Sukanya Denni
Lucien Piat
Ludovic Duvaux

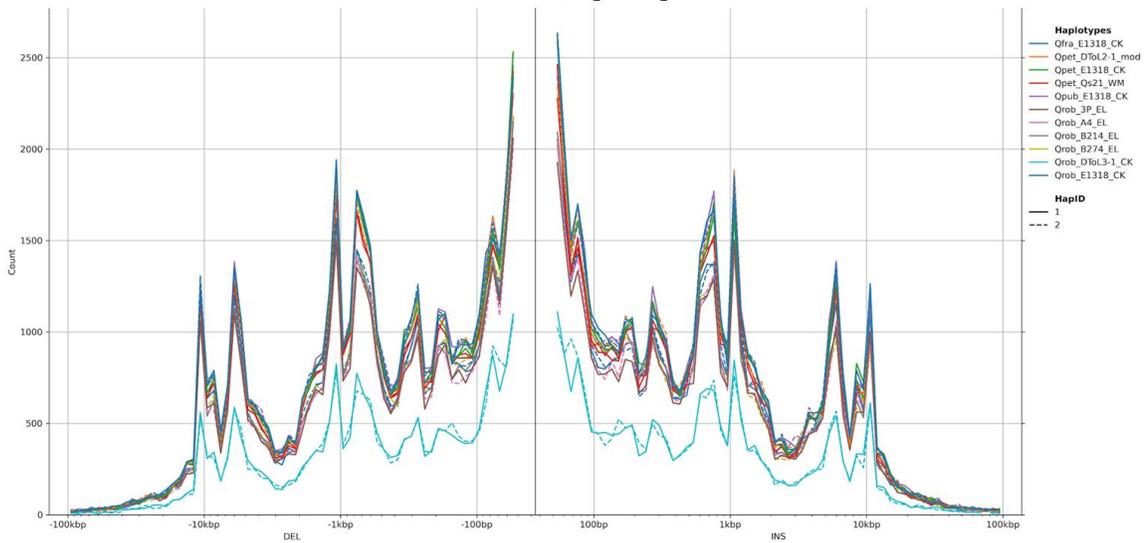
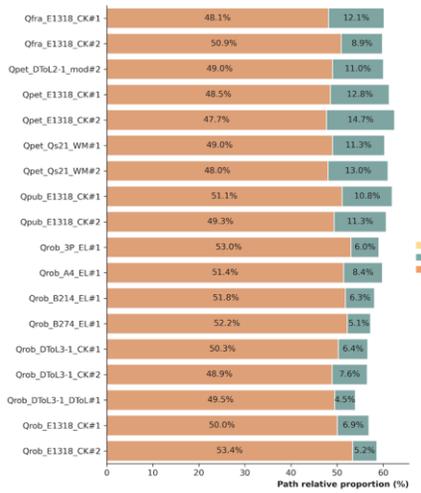
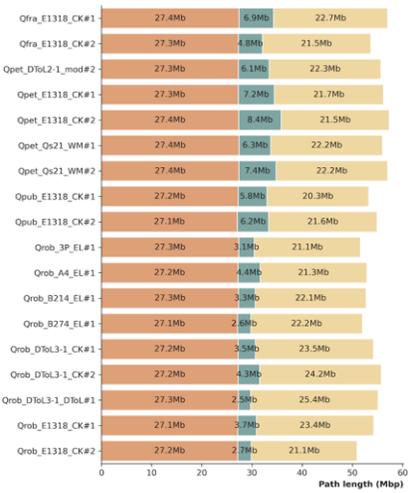


Pan1c : construction d'un pangénome à l'échelle du chromosome



Path composition by groups - 18Qr-v2a - chr01

Pan1c - 18Qr-v2a - Qrob_DTol3-1_DTOL - VG



Auteurs :

Alexis Mergez
 Christophe Klopp
 Ludovic Duvaux
 Fabrice Legeai
 Christine Gaspin
 Frederic Choulet

autour de la
formation/ réseau/ communication

Fête de la science 2023



ateliers pratiques:

- *assemblage de génome*
- *biostatistiques*



Thématique : rôle du microbiote dans la performance des chevaux d'endurance

3 ateliers destinés au collégien.ne.s :

- Assemblage de séquences
- Initiation à la programmation
- Introduction à la biostatistique

Supports partageables / réutilisables

2 classes recues

Auteurs : Gabryelle Agoutin, Lucas Auer, Cédric Cabau, Philippe Bardou, Vincent Darbot, Cervin Guyomar, Luc Jouneau, Géraldine Pascal, Didier laborie, Elise Maigné, Sabrina Legoueix

Escape game “Closed Access”

Un escape game sur la Science ouverte adapté pour INRAE

Expérience ludique et instructive pour le printemps de la données 2024

9 sessions de jeu d'environ 1h30 sur les sites de Jouy-en-Josas et d'Antony => 36 participants

Excellent accueil => continuer à le faire vivre et à le transmettre

4 sessions sur le site de Toulouse Jeudi 03 Octobre (merci B4B)

L'escape game « Closed Access », en accès libre sur Zenodo (<https://zenodo.org/records/3876138>), a été créé par la bibliothèque Michel SERRE de l'École Centrale de Lyon

Auteurs : Nathalie PEYNOT, Mathieu CHARLES, Anne FRAMBOURG, Cédric MIDOUX, Véronique OSSWALD



L'IA en science du vivant - Impact sur nos métiers - mini-symposium JOBIM 2024



*300 personnes sont venues écouter notre mini-symposium
- avec introduction des orateurs façon pièce de théâtre.*



Auteurs : Gabryelle Agoutin, Lucas Auer, Philippe Bardou , Cédric Cabau, Cervin Guyomar, Sarah Maman, Géraldine Pascal

Vincent Guigue

- Le fonctionnement des LLM et les spécificités de chatGPT
- Les limites générales (techniques) des approches de machine learning
- Les usages et applications des LLM et les développements techniques autour des autres IA génératives
- Les risques engendrés



Guillaume Cabanac

- Des cas de mésusages de la génération de textes et de données avec ChatGPT dans des articles scientifiques
- Publication de données fictives
- Usurpation d'identité d'auteurs
- Phénomène des paper mills.
- Problematic Paper Screener : un outil pour détecter les publications suspectes
- Des éléments sur les bonnes pratiques

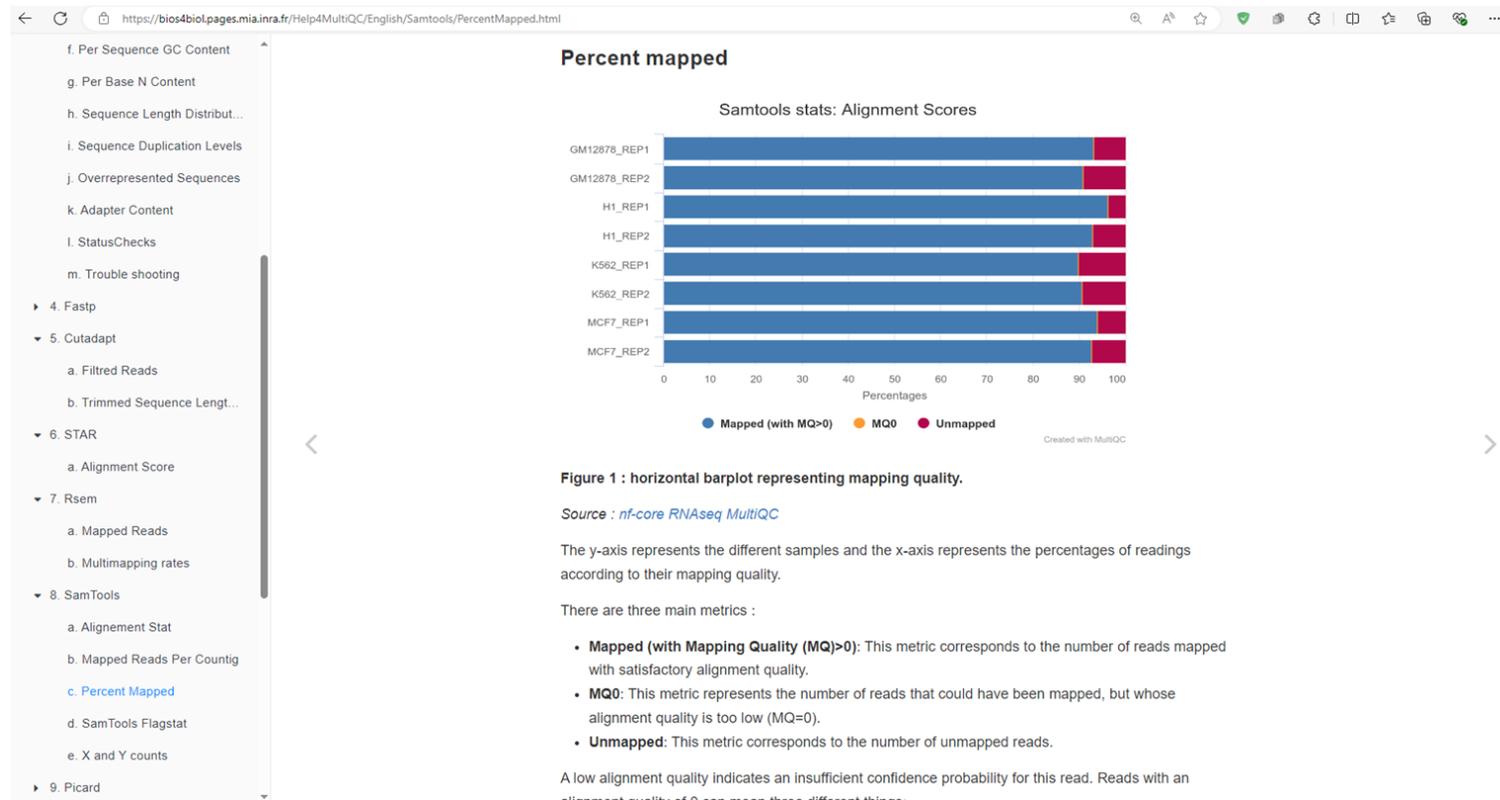


Help4MultiQC

Une documentation participative pour aider les biologistes à lire les rapports multiQC. Pour le moment nous avons documenté les sorties des WF Nextflow nf-core RNAseq et Sarek.

GitBook : <https://bios4biol.pages.mia.inra.fr/Help4MultiQC/>

Repo GitLab : <https://forgemia.inra.fr/bios4biol/Help4MultiQC/-/blob/main/README.md>



Auteurs : **Gaston Rognon, Sarah Maman, Claire Hoede, Yannick Lippi, Cervin Guyomar**

Article Kids Frontiers “Qu’est-ce que la bioinformatique”

3 réunions en marge des visios mensuelles :

Deux paragraphes proposés :

- Avènement de la bioinformatique (AF)
- Un exemple d’analyse bioinformatique : la transcriptomique (LJ)



Dans l’année écoulée, un article “**Bioinformatics: Using “Big” Data to Solve Health Mysteries**” est paru dans Kids Frontiers. Cet article est plus centré sur les Health Data, ce que l’on peut en faire et sur les problèmes de confidentialité des données personnelles que sur la bioinformatique elle-même.

Auteurs : **Gabryelle Agoutin, Anne Frambourg, Luc Jouneau, Elise Maigné**

Vie du site du CATI BIOS4BioI



Pages membres :

Tableau dans nextcloud

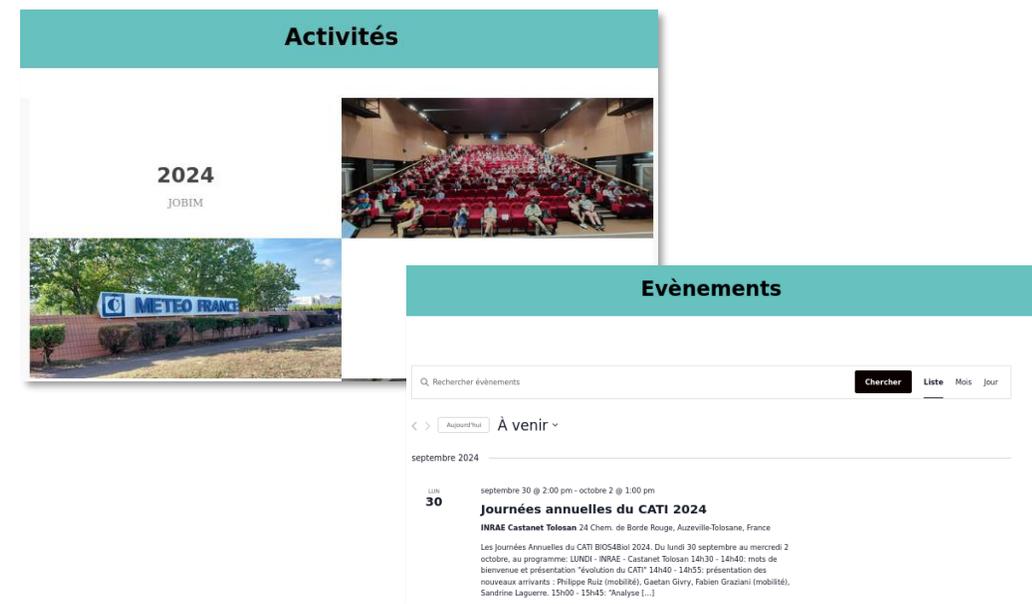
<https://nextcloud.inrae.fr/s/rTKp4xKd6T5MaWE>

+ message



Envoyez-nous votre photo !

N'hésitez pas à nous envoyer toutes informations concernant le CATI !



Site à destination des départements, DUs, collègues...

Auteurs : Jérémy Tournayre, Anne Frambourg, Géraldine Pascal, Luc Jouneau

Journée Shiny::INRAE

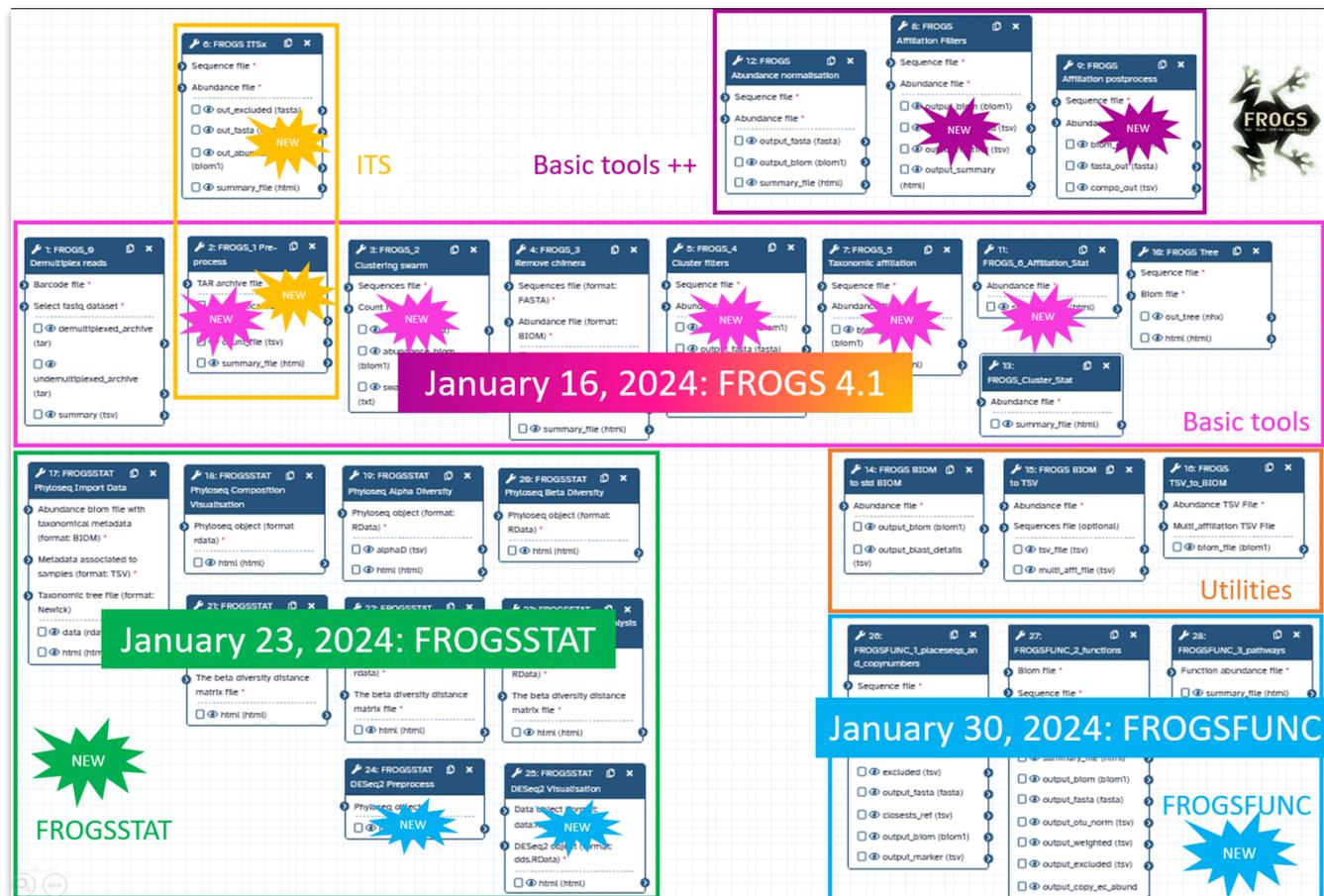
Financement SAPI pour organiser une journée Shiny (x2) dans l'institut



https://sk8.pages.mia.inra.fr/journee_shiny_inrae/edition_2025

Auteurs : Porté par **Elise Maigné** et la team SK8

Les mardis de la grenouille



```
February 6, 2024: FROGS CLI

(FROGS-v4.1.0_env) [bleuet@genobioinfo2 Formation_FROGS]$ ll
total 18
drwxr-xr-x 2 bleuet formation 4096 Feb  5 15:57 Logs
-rw-r--r-- 1 bleuet formation 1150 Feb  5 15:57 readme.txt
drwxr-xr-x 2 bleuet formation 16384 Feb  5 15:59 Results
drwxr-xr-x 2 bleuet formation 4096 Feb  2 22:34 Sequences
(FROGS-v4.1.0_env) [bleuet@genobioinfo2 Formation_FROGS]$ ll Logs/
total 16
-rw-r--r-- 1 bleuet formation 8370 Feb  5 15:59 01-preprocess.log
-rw-r--r-- 1 bleuet formation  0 Feb  5 15:56 frogs_preprocess.err
-rw-r--r-- 1 bleuet formation  0 Feb  5 15:56 frogs_preprocess.out
(FROGS-v4.1.0_env) [bleuet@genobioinfo2 Formation_FROGS]$ more Logs/01-preprocess.log
```

Mises à jour FROGS sur les nouveautés dédiées aux anciens formés.

+ nouvelle formation d'une journée : FROGS CLI

Auteurs : **Gabryelle Agoutin, Lucas Auer, Géraldine Pascal**

BLOG sur R/RStudio , biostats https://lcauquil.pages.mia.inra.fr/mon_blog/

Blog à part

Tips, tutoriels pour R et RStudio

17 Default

Sep 12, 2024
0 min

Comment choisir une version de R spécifique au démarrage de RStudio en 1 clic

RStudio nous donne la possibilité d'utiliser n'importe quelle version de R installée.



Categories

- All (20)
- RStudio (1)
- assumption (1)
- code styling (1)
- dplyr (1)
- emoji (1)
- facet (1)
- flextable (1)
- fonction (1)
- ggplot (4)
- ggrepel (1)
- grep (1)
- growthcurver (1)
- gtExtras (1)
- lwalk (2)
- link (1)
- list (1)
- map (5)
- mix model (1)
- model (2)
- names (1)
- package (1)
- pak (1)
- pals (1)
- paste (1)
- performance (1)
- pipe (1)
- pivot_longer (1)
- purrr (5)
- quarto (2)
- rlist (1)

Jul 24, 2024
1 min

Package pak

Le package `pak` <https://pak.r-lib.org/index.html> permet l'installation de packages R depuis différents dépôts distants ou locaux de manière rapide, sécurisée et facile.

Jul 24, 2024
2 min

Data first, nom d'un pipe

Ou comment utiliser les pipes sans se faire piéger

Pour rappel un pipe permet de chaîner plusieurs processus de sorte que la sortie d'un processus alimente directement l'entrée du processus suivant Définition wikipedia

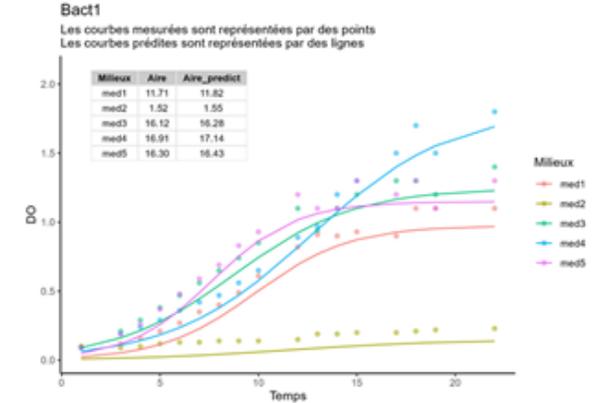


Tutoriels:

- ggplot
- tableaux résultats
- modèles mixtes
- package purrr
- RCy3

News:

- package pak



Tips:

- Insérer des emojis dans quarto
- Couper un jeu de donnée en 2
- Choisir sa version de R en 1 clic
- ...

Gene	Molec_A			Molec_B			Lot	Molecule	Lot_Molecule
	Lot_A	Lot_B	Lot_C	Lot_A	Lot_B	Lot_C			
Gene_1	3.84 ± 0.45 b	2.46 ± 0.42 a	2.31 ± 0.37 a	3.9 ± 0.45 b	2.94 ± 0.52 a	2.48 ± 0.31 a	***	NS	NS
Gene_2	1.94 ± 0.12 b	1.53 ± 0.13 a	1.49 ± 0.12 a	1.95 ± 0.12 b	1.67 ± 0.15 a	1.55 ± 0.1 a	***	NS	NS
Gene_3	0.03 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.03 ± 0.01 b	0.02 ± 0.01 a	0.02 ± 0 a	***	NS	NS
Gene_4	0.29 ± 0.06	0.2 ± 0.04	0.21 ± 0.06	0.2 ± 0.04	0.17 ± 0.05	0.17 ± 0.06	NS	*	NS
Gene_5	0.05 ± 0.01	0.03 ± 0.01	0.03 ± 0.01	0.06 ± 0.01 b	0.04 ± 0.01 ab	0.03 ± 0.01 a	**	NS	NS
Gene_6	0.02 ± 0.01	0.02 ± 0.01	0.03 ± 0.02	0.03 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	NS	NS	NS
Gene_7	0.06 ± 0.02	0.04 ± 0.02	0.04 ± 0.02	0.1 ± 0.03 b	0.08 ± 0.03 b	0.03 ± 0.01 a	***	*	*
Gene_8	0.11 ± 0.02	0.13 ± 0.02	0.13 ± 0.02	0.12 ± 0.02	0.14 ± 0.01	0.12 ± 0.01	NS	NS	NS
Gene_9	0.1 ± 0.02 b	0.04 ± 0.01 a	0.03 ± 0.01 a	0.1 ± 0.02 b	0.05 ± 0.01 a	0.03 ± 0.01 a	***	NS	NS
Gene_10	0.21 ± 0.01	0.23 ± 0.03	0.24 ± 0.04	0.22 ± 0.02	0.22 ± 0.01	0.23 ± 0.02	NS	NS	NS
Gene_11	0.38 ± 0.06	0.37 ± 0.07	0.47 ± 0.06	0.4 ± 0.09	0.4 ± 0.05	0.38 ± 0.03	NS	NS	NS
Gene_12	0.07 ± 0.01	0.08 ± 0.02	0.06 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.06 ± 0.01	NS	NS	NS

***: p < 0.001
** : p < 0.01
* : p < 0.05
NS: p ≥ 0.05

Auteur : Laurent Cauquil

Formation git/gitlab déclinée sur tous les centres INRAE

Formation initialement montée sur le centre PACA,
Point de départ = IDE de l'utilisateur



Session 1 : initiation à Git

Session 2 : GitLab et dépôt distant

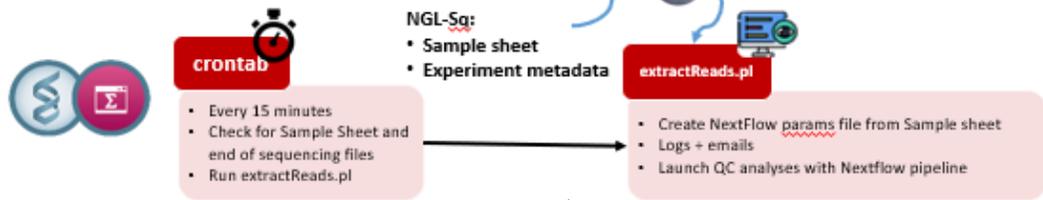
Session 3 : Les branches

Réseau national de ~30 formateurs git/gitlab

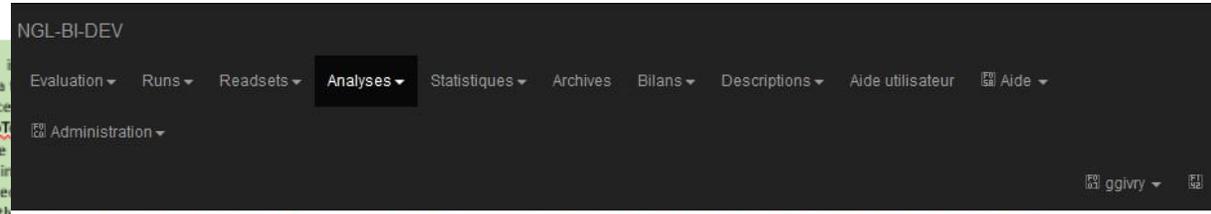
infra/systeme/BD

Contrôles qualité données NGS sous nextflow et NGL-Bi

NGS quality analysis workflow per technology :



Each technology manages the data (High-Performance systems of Genom) integrity, prepare pipeline. Depending analyses performed pipelines. Current



20240913-MISEQ_6-LM5H3-FC9AB114-7370-11EF-B725-DA64627E7159 Nouveau

Général Téléchargement Avancé Infos workflow Historique d'évaluation

rsync

You could use the following rsync method and just specify /yourRepository/

```
rsync -avhL sbsuser@genobioinfo.toulouse.inrae.fr:/save/user/sbsuser/public_html/download/214d92b9-39d8-42e9-8722-a6735fedd7c3_1724745330.111229/20240913-MISEQ_6-LM5H3-FC9AB114-7370-11EF-B725-DA64627E7159/ /yourRepository/
```

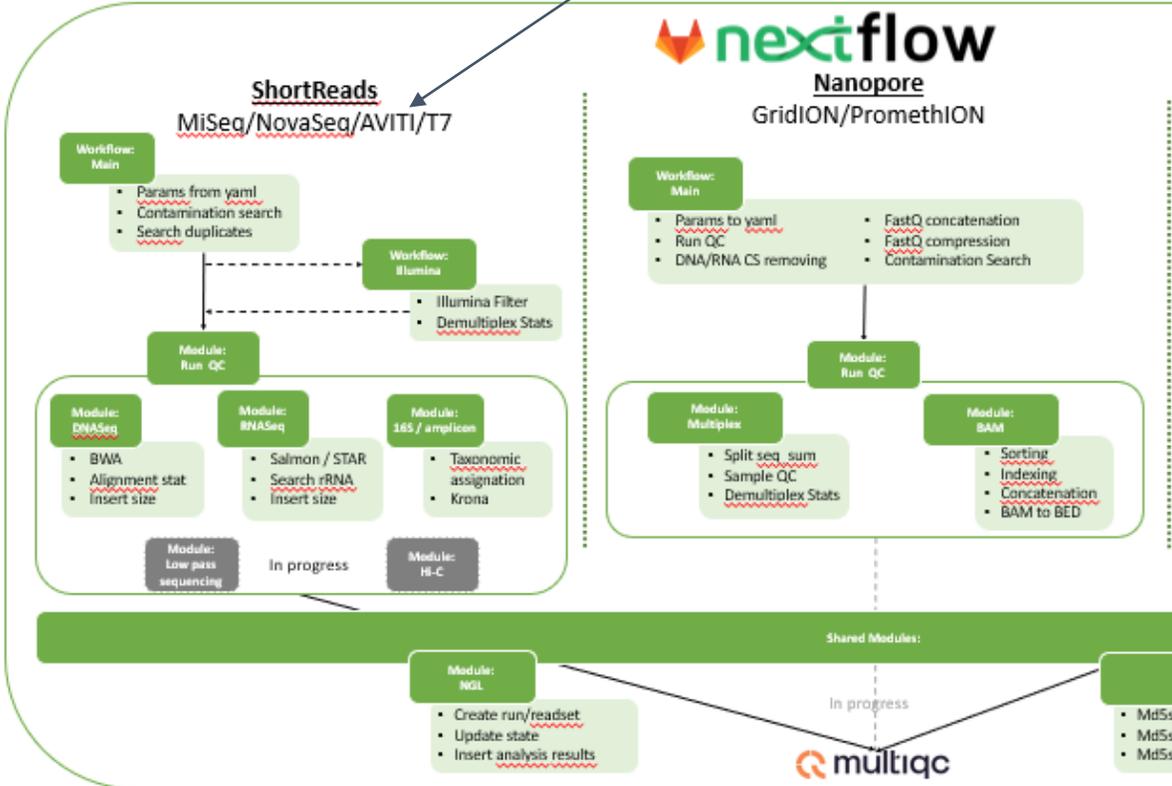
wget

You could use the following wget method

```
wget --no-check-certificate -ic https://web-genobioinfo.toulouse.inrae.fr/~sbsuser/download/214d92b9-39d8-42e9-8722-a6735fedd7c3_1724745330.111229/20240913-MISEQ_6-LM5H3-FC9AB114-7370-11EF-B725-DA64627E7159/wgetFiles.txt
```

Readsets (283)

N° Piste	Code	Fichiers	Etat	Projet	Nom échantillon	% déposé	% Séquences valides / piste	Nb Séquences valides	Nb Bases	% >= Q30	Score Qualité moyen	Valid QC ?
1	GENO886459:G12-84-24-LM5H3-P1-A095-1726411156	E74-S04-LA-R3_TTGCTA-LM5H3_L001_R1.fastq.gz E74-S04-LA-R3_TTGCTA-LM5H3_L001_R2.fastq.gz	Disponible	84-24	E74-S04-LA-R3	0,35	0,22	54 790	27 395 000	91,91	35,94	Oui
1	GENO886460:B11-84-24-LM5H3-P1-A178-1726411156	E75-S07-ZC-R1_ACCTCA-LM5H3_L001_R1.fastq.gz E75-S07-ZC-R1_ACCTCA-LM5H3_L001_R2.fastq.gz	EVAL. QC en attente	84-24	E75-S07-ZC-R1	0,35	0,28	68 463	34 231 500	90,28	35,57	---
1	GENO886461:H6-84-24-LM5H3-P1-A178-1726411156	E76-S06-TR_AAGCTA-LM5H3_L001_R1.fastq.gz	EVAL. QC en attente	84-24	E76-S06-TR	0,35	0,28	67 837	33 918 500	90,30	35,58	---



Technologies :

- Pipelines : bash, perl, R, python, Nextflow
- NGL : play framework, java, REST API, dtools, angularJS, MongoDB, MariaDB
- e-Sitoul : Java, perl, REST API, PostgreSQL

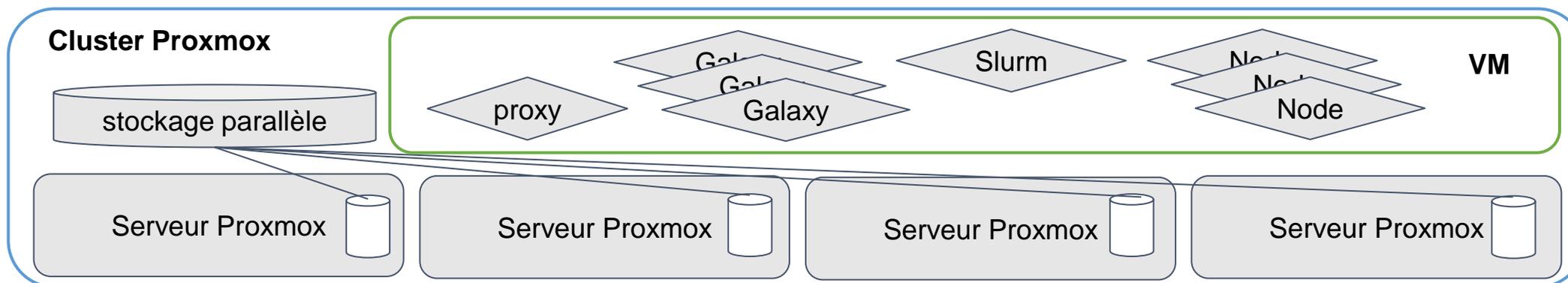
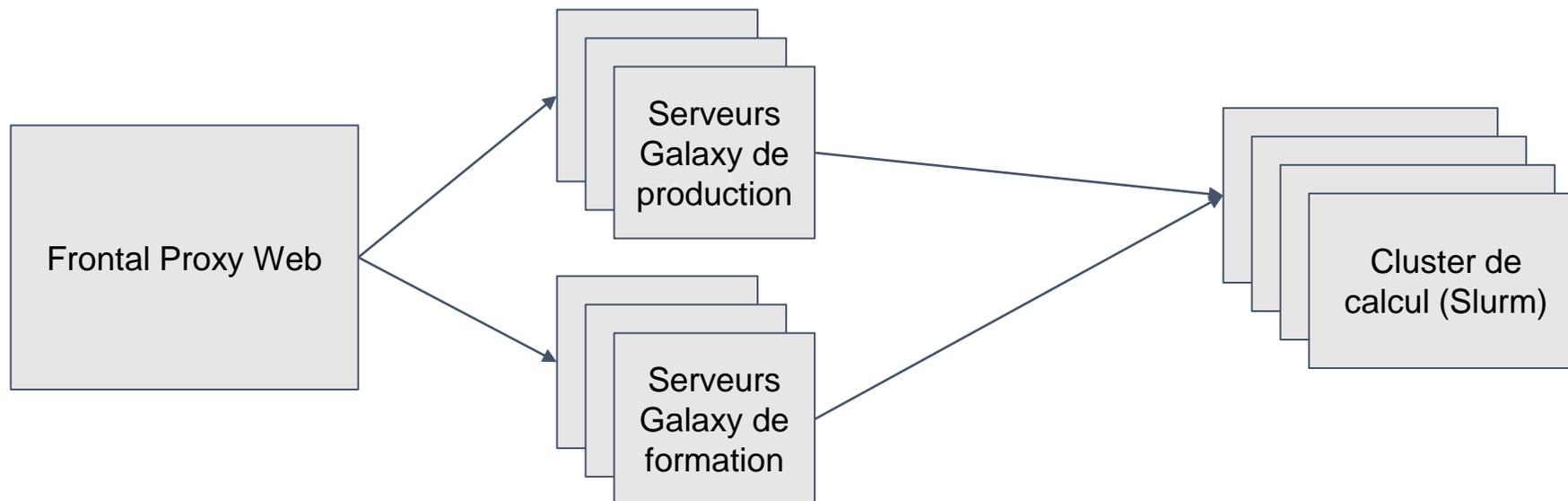
Thanks :
GeT-PlaGe info and bioinfo teams for the maintenance and develop for hosting our information system and our NGS data, the CEA/GeT-PlaGe

Auteurs : Didier L., Marie-Stéphane Trotard, Céline Noirot, tout GeT-PlaGe

Testeur à venir : Laurent Cauquil, autres?

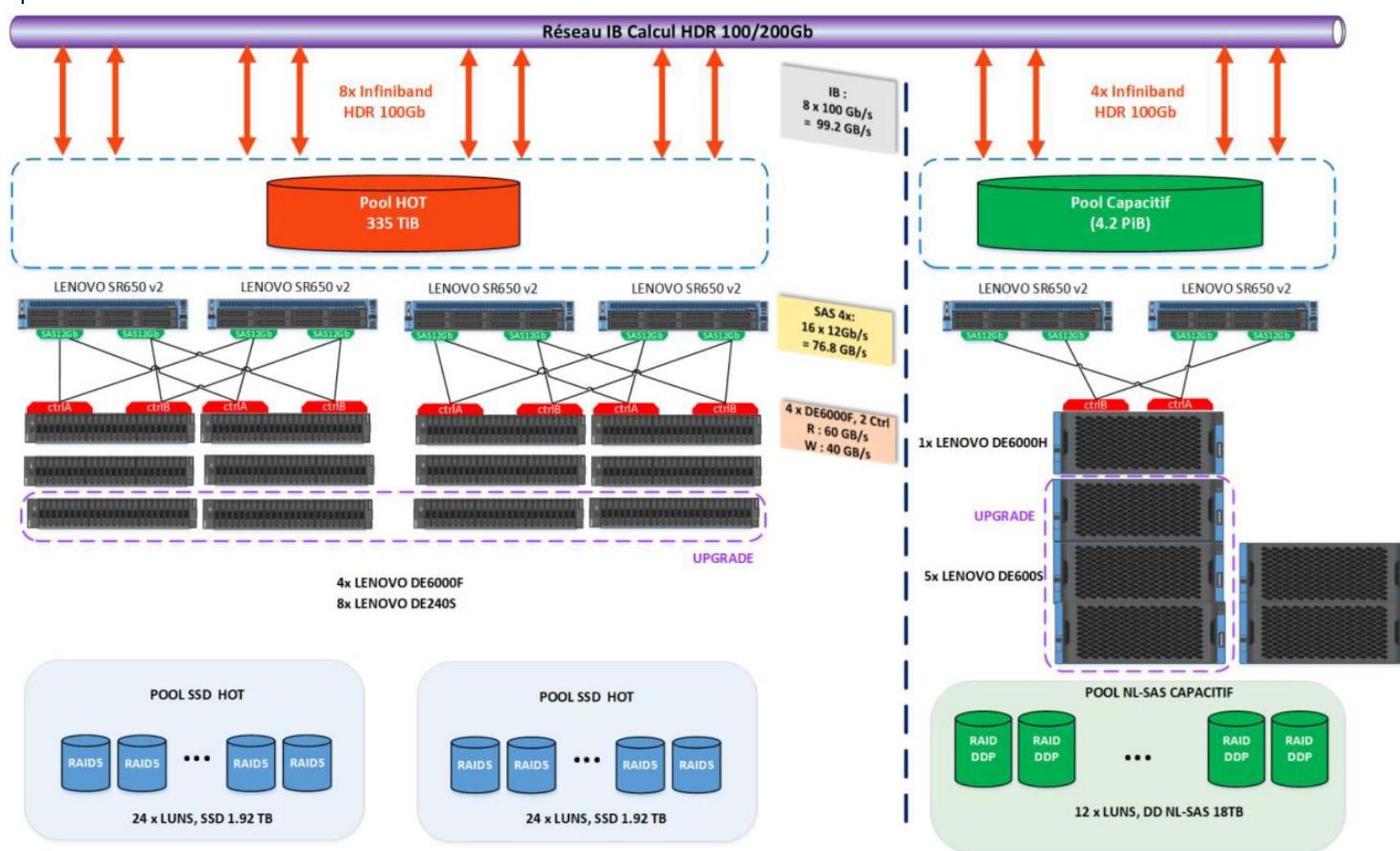
Galaxy

Gestionnaire Web de traitements en bioinformatique ... mais pas que



Auteur : **Patrice DEHAIS**

Extension /work (06/2024)



Auteurs :

Marie-S
Trotard,

Patrice
Dehais,

Didier
Laborie

- Un espace rapide et performant « hot » de 335 TiB composé de :
 - 4 x nœuds d'I/O
 - 4 x baies DE6000F 2U24 et 8 x tiroirs d'extension DE240S 2U24
 - 288 x SSD 1.92TB
- Un espace « capacitif » de 4.2 PiB composé de :
 - 2 x nœuds d'I/O
 - 1 x baie DE6000H – 4U60 et 5 x tiroirs d'extension DE600S – 4U60
 - 360 x NL-SAS 18TB

En route vers la pérennisation SK8 !



Infrastructure pour gérer et déployer des application R shiny.

100 applis hébergées. Financement pour se créer une infrastructure pour gérer le passage à l'échelle !



SPE

SA



SK8 Team @ Hackathon SK8 2024

Auteurs : ~20 personnes, 10 CATI représentés.

Porté par Jean-François Rey (CATI IMOTEP) et **Elise Maigné** (CATI Bios4Biol)

Analyse omics

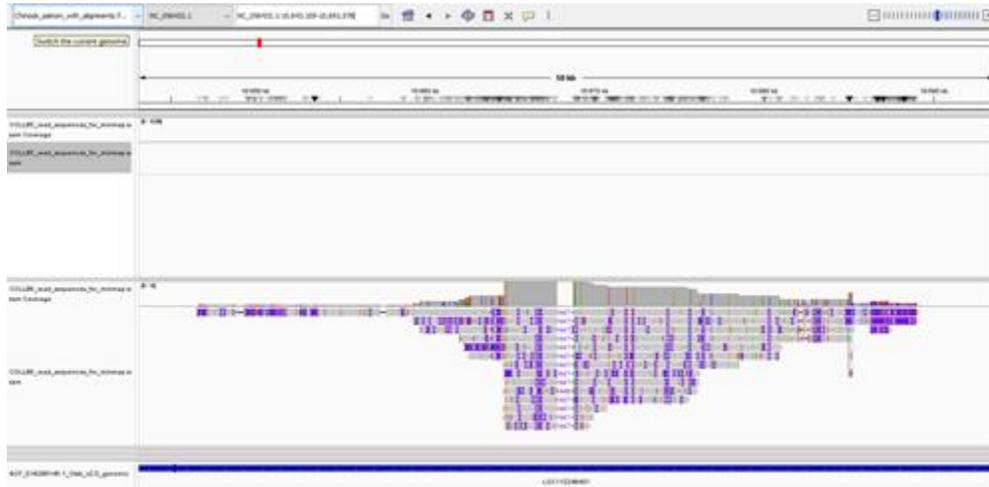
Analyse du génome de lignées cellulaires transgéniques de saumon Chinook

Séquençage long reads (Nanopore) de 2 lignées cellulaires transgéniques (profondeurs : 8.6 X et 13 X)

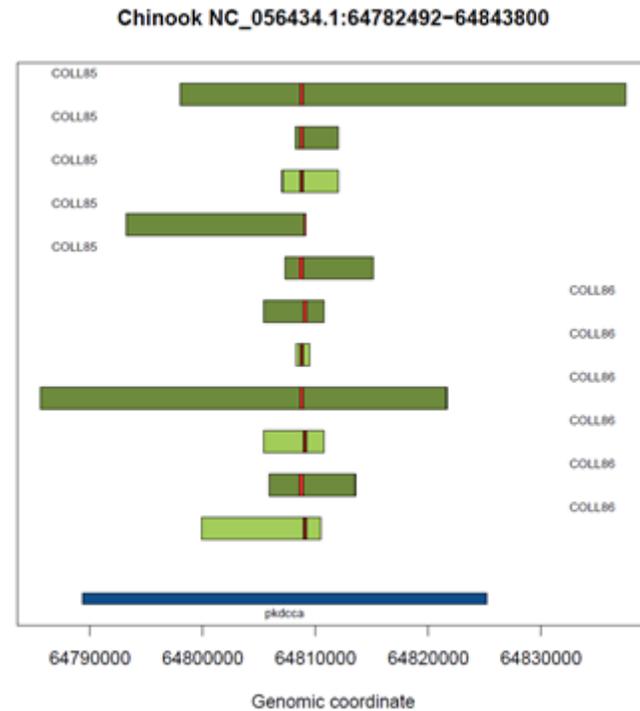
Où et en combien de copies les transgènes se sont insérés dans le génome hôte ?

Alignement minimap des long reads sur le génome du Chinook et sur les séquences des transgènes

Production de cartes d'insertion des transgènes



Visualisation sous IGV des alignements sur le génome du Chinook

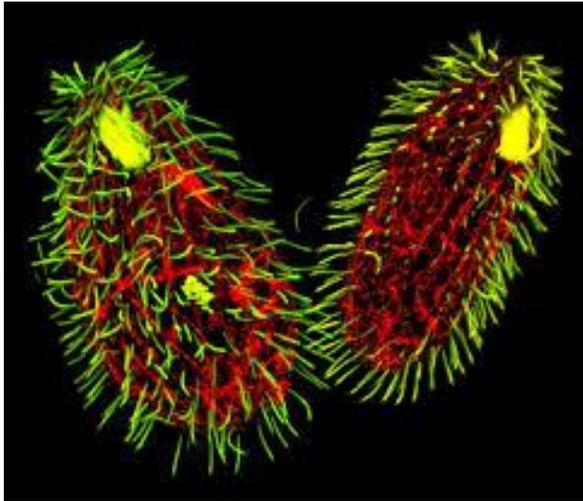


1 site d'insertion au sein d'un gène Chinook retrouvé dans les 2 lignées



La séquence du transgène est morcelée

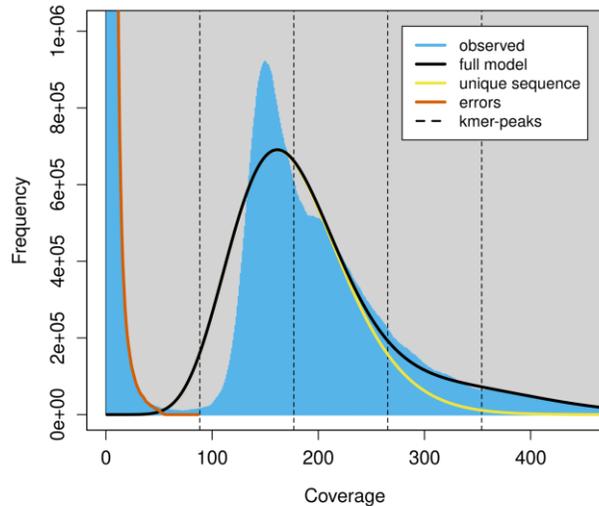
Aide à l'assemblage du génome de *Tetrahymena thermophila*, Quel MIC MAC!



<https://www.nikonsmallworld.com/galleries/2005-photomicrography-competition/tetrahymena-thermophila-protzoa-cells>

GenomeScope Profile

len:116,485,153bp uniq:77.1%
aa:100% ab:0.001%
kcov:88.5 err:0.2% dup:15.2 k:21 p:2



Tetrahymena thermophila (inspiration Wikipedia)

C'est un genre de ciliés pouvant être commensaux ou pathogènes.

Les espèces sont très communes dans l'eau douce

Il a la particularité d'exister en 7 sexes (types sexuels) différents, ce qui donne 21 différentes combinaisons possibles d'accouplement.

Il a deux génomes :

- le micronoyau (MIC) de la lignée germinale diploïde (5 paires de chromosomes)
- le macronoyau (MAC) qui est une multiplication d'un sous partie (zones géniques) sur micronoyau (280 chromosomes)

Assemblage du génome MAC

Dans les séquences les deux sources (MIC et MAC) sont présentes.

Pour séparer les séquences du MIC de celles du MAC on peut utiliser les fréquences des kmers.

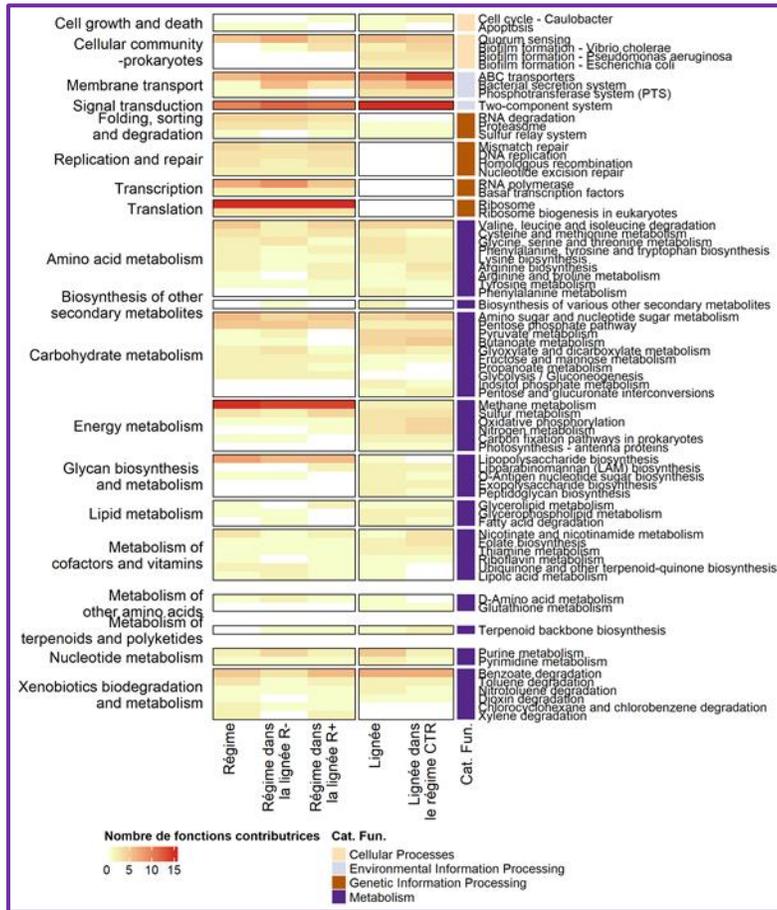
Les séquences MAC contiennent des kmers fréquents et des kmer d'erreurs (influence limitée dans l'espace). Celles du MIC contiennent les longues zones de kmers peu fréquents.

Une fois les séquences séparées on peut assembler les séquences du MAC (hifiasm)

Assemblages	taille	N50	L50	2 télomères
Public (GCF_000189635.1)	103 0143 75	520 943	59	25
Nouvelles données Moulis	102 580 396	1 007 618	31	144

Auteur : Christophe Klopp, Collaboration : Philippe Hervé (CNRS Moulis)

Caractérisation du microbiote caecal de poule pondeuse



Métatranscriptome

Métagénome

Catalogue de gènes microbiens

Expression de gènes microbiens

Abondance de gènes microbiens

Analyse différentielle
Analyse de redondance

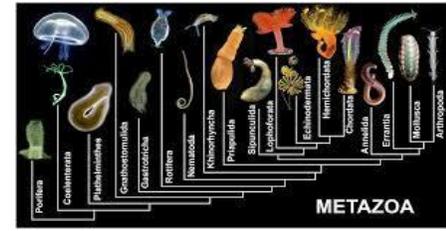
Analyse intégrative (PTA)

Quels sont les effets de la lignée ou du régime alimentaire sur les activités du microbiote caecal de poule pondeuse ?

Quelle est la relation entre l'abondance des micro-organismes et l'expression des fonctions ?



Co-évolution de la densité en CpG et des génomes animaux



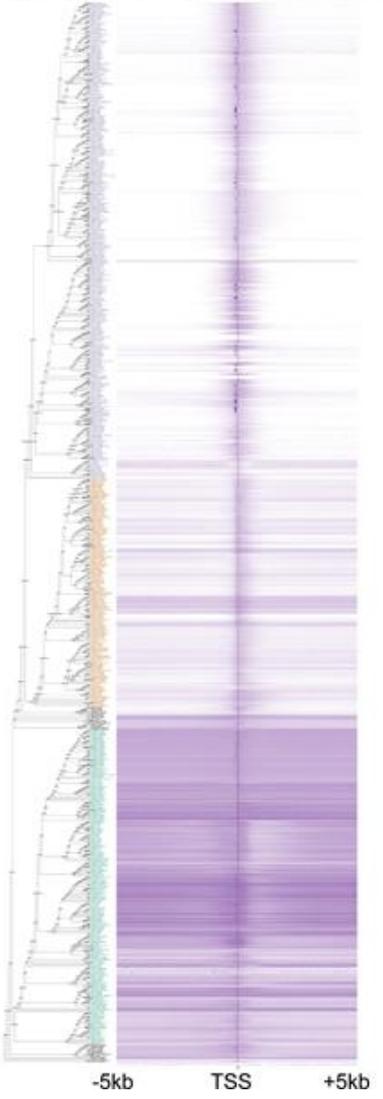
989 génomes

Metazoa evolution tree

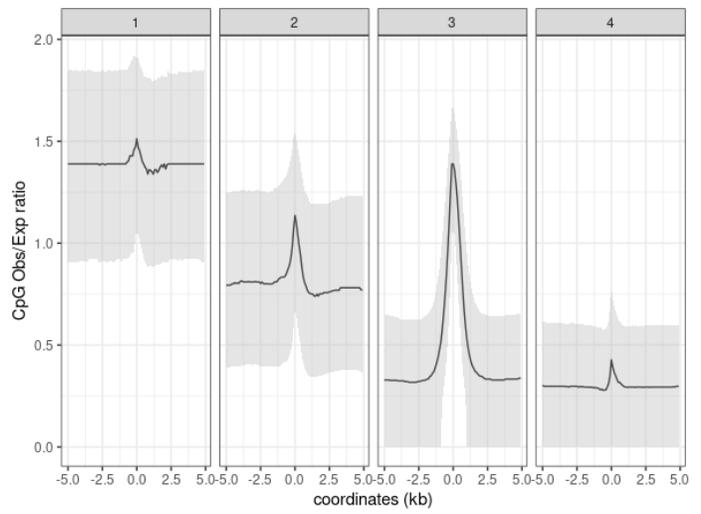
Sarcopterygii

Actinopterygii

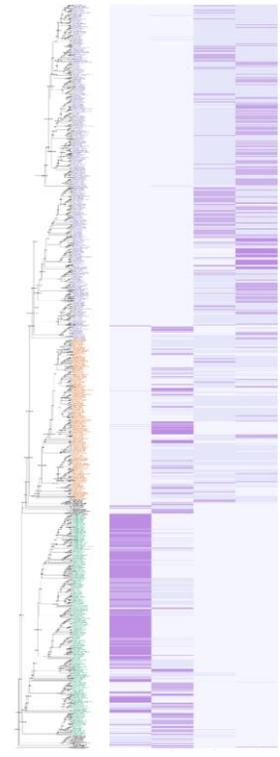
Protostomia



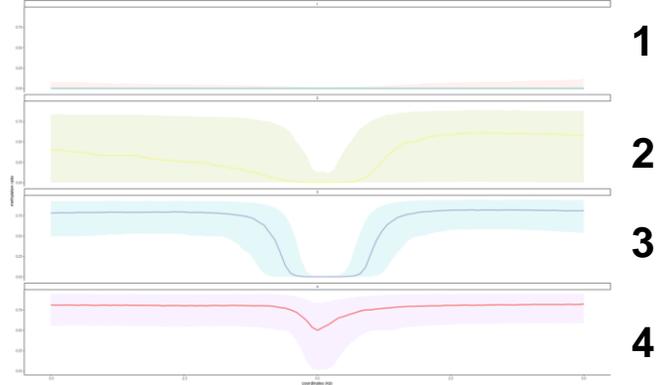
profiles CpG Obs/Exp



fréquence/génome



profiles méthylation



Co-évolution de la densité en CpG et des génomes animaux

> Sarcopterygii , Actinopterygii , Protosomia

Quatres classes de densité de CpG autour des TSS

> Répartition différente selon le clade

Corrélation négative entre la densité en CpG (classes) et le niveau de méthylation

> Conservation au sein des métazoaires

Auteurs : A BRIONNE, A LAURENT