

une image = un projet

Journées annuelles
CATI BIOS4Biol 2023



pipelines, développements,
benchmarking

D-Genies news

D-GENIES About Run Results Gallery Documentation Install Contact Legal

Launch map analysis

Name of your job

E-mail

We will send you results by mail

[New alignment](#) [Plot alignment](#) **[Batch alignments](#)**

Batch file

Text file following [batch format](#). Must end with `.txt`

```
1 # An align job
2 type=align id_job=example_align tool=minimap2 options=repeat:many
  query=example://Fusarium_culmorum.EF1.dna.toplevel.fa.gz
  target=example://Fusarium_graminearum.RR1.dna.toplevel.fa.gz
3
4 # An all versus all align job
5 type=align id_job=example_align_ava tool=minimap2 options=repeat:many
  target=example://Fusarium_graminearum.RR1.dna.toplevel.fa.gz
6
7 # A plot job
8 type=plot id_job=example_plot backup=example://backup.tar.gz
9
```

User files

File	Size	State	Actions

[Quick help for tools](#)

Minimap2 v2.24

Better for low similar fasta files
Please use `tool=minimap2`

Options¹

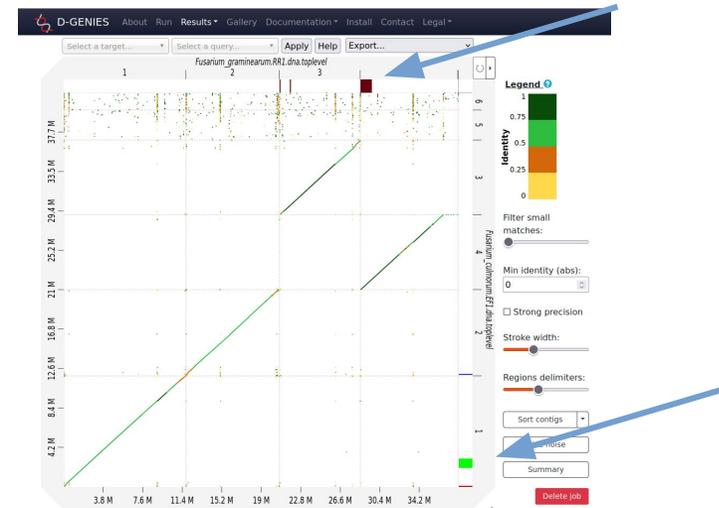
Default options are displayed in **bold**.
Please use `options=key1,key2,...` where **keys** are:

Repeatedness (mutually exclusive)

Trade-off between alignment precision and computation requirements (CPU Time + RAM) by ignoring the top fraction of most frequent minimizers

- `repeat:few` : few repeats ?
- `repeat:some` : some repeats ?
- `repeat:many` : many repeats ?

In development : adding annotations on the dot plot borders from bed or wiggle files



Philippe Bordron
Christophe Klopp
Bioinfo Genotoul

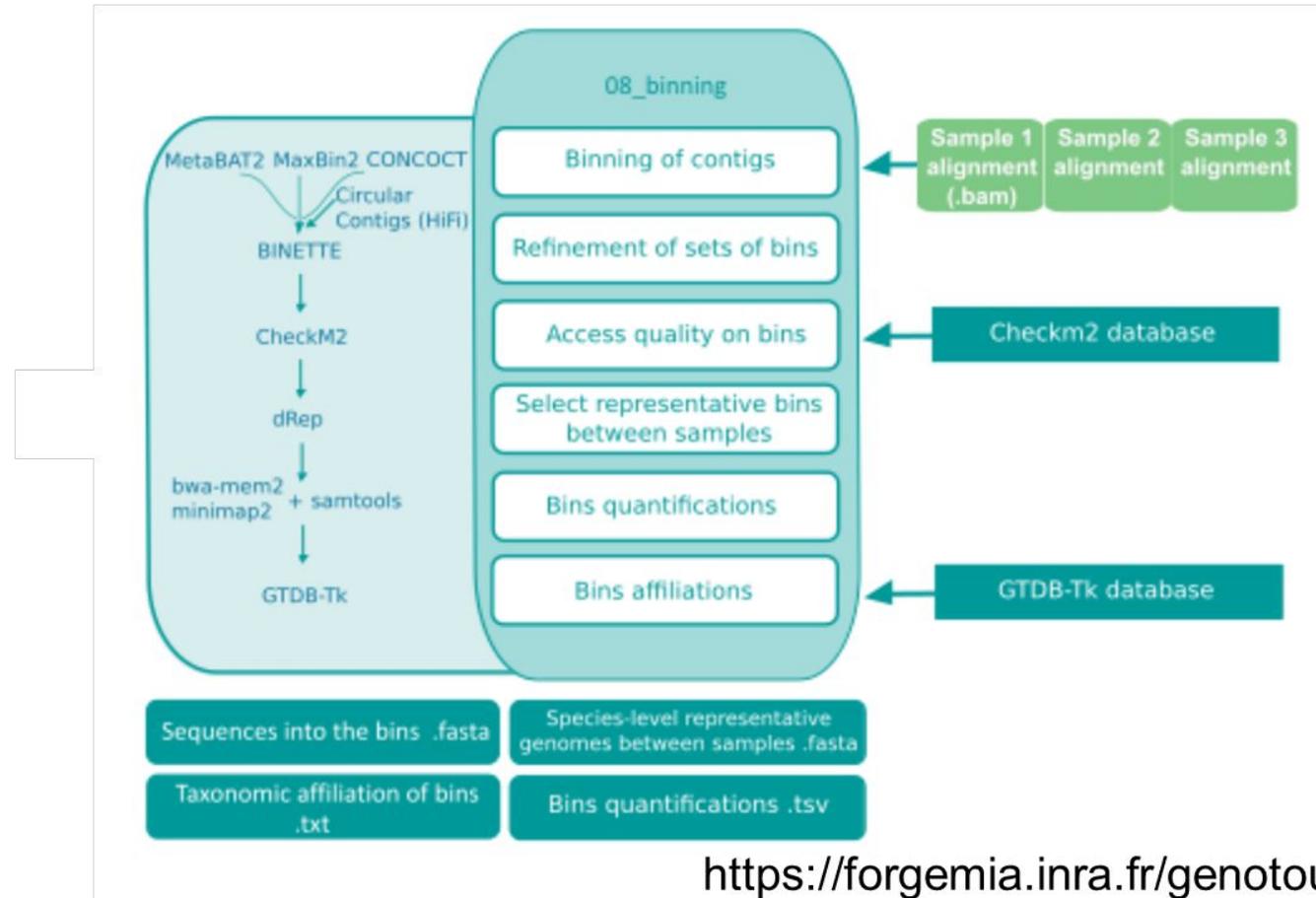
Batch mode in production!

Améliorations du workflow metagWGS



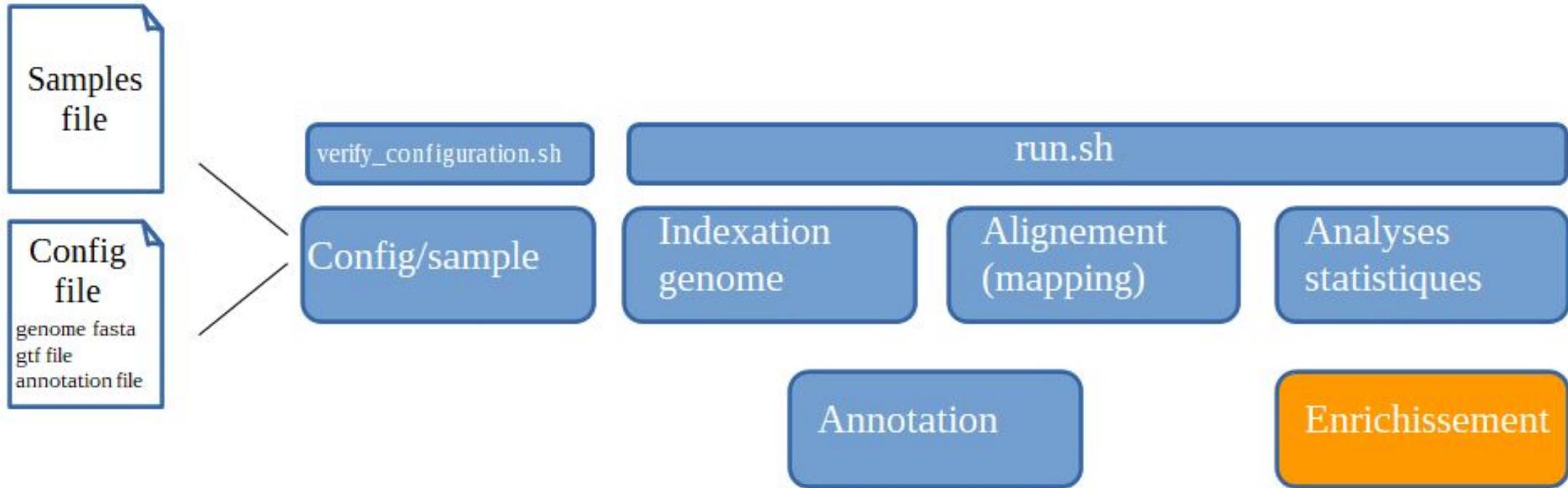
metagWGS 2.4 permet l'analyse des short-reads ou des long-reads HiFi PacBio.

Nouvelles fonctionnalités majeures : le co-assemblage, annotation des gènes plus rapide, et développement de binnette : <https://github.com/genotoul-bioinfo/Binette>



Auteurs : Jean Mainguy, Maïna Vienne, Vincent Darbot, Céline Noirot, Geraldine Pascal and Claire Hoede

Ajout module d'enrichissement au pipeline RNAseq



Trois méthodes proposées :

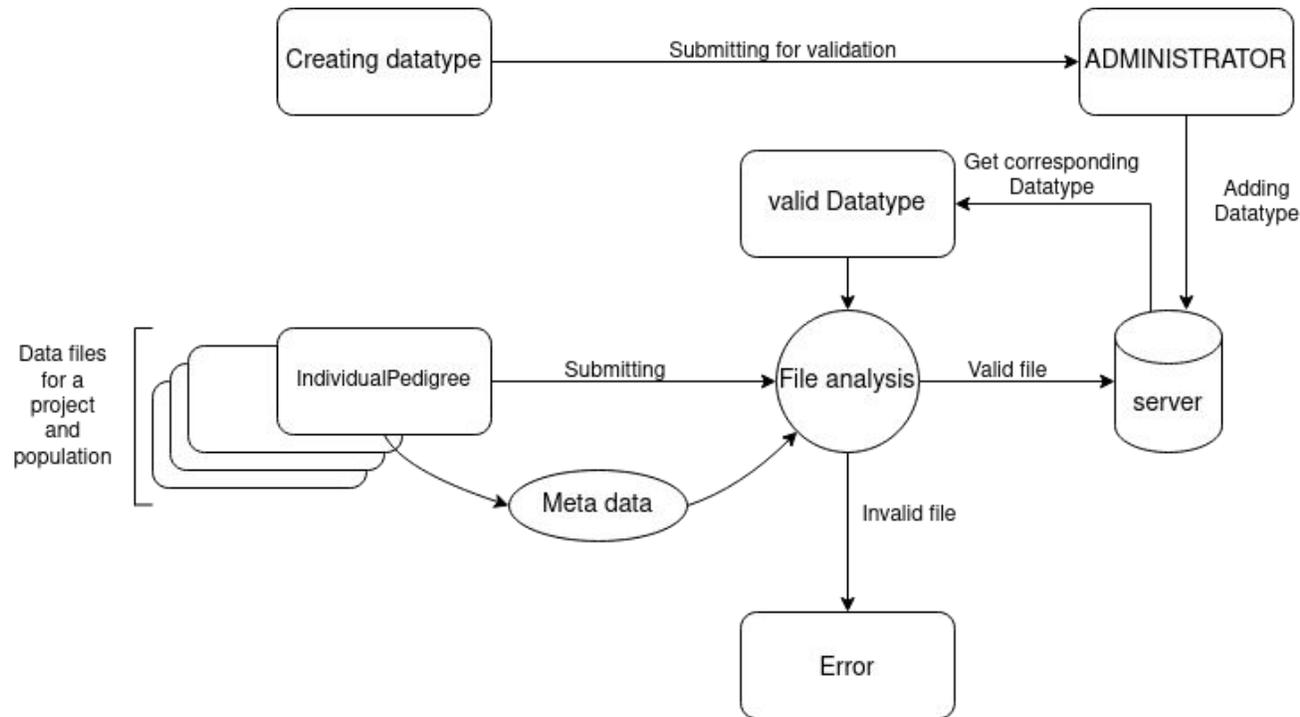
- GSEA_classique (Broad institute)
- GSEA_preranked (Broad institute)
- GSEA_DE (test hypergéométrique)



Auteurs : Anne Frambourg, Luc Jouneau

Efficace database

Efficace database est une application web de soumission et consultation de fichier respectant des contraintes pré-définies.



Home Submit Explore Datatype Contact

DATA

EXPLORE EFFICACE FILE DATABASE

Search
Filter on specific fields

Show/Hide columns
Configure the table below by selecting column name(s)

Facet
Explore data by facet

project	pop	batch	datatype	agefactor	feedfactor
<input type="checkbox"/> MDfastedFed	<input type="checkbox"/> A3A3	<input type="checkbox"/> 2016.1	<input type="checkbox"/> individualPedigree	<input type="checkbox"/> 0d	<input type="checkbox"/> CTvsFASTEDvsREFED
<input type="checkbox"/> novo	<input type="checkbox"/> com8	<input type="checkbox"/> 2017.2	<input type="checkbox"/> phenoBW	<input type="checkbox"/> 17w	<input type="checkbox"/> CTvsHF
<input type="checkbox"/> fatInteger	<input type="checkbox"/> Fayo	<input type="checkbox"/> 2019.2	<input type="checkbox"/> phenoR	<input type="checkbox"/> 26w	<input type="checkbox"/> CTvsLE
<input type="checkbox"/> chickStress	<input type="checkbox"/> FLL	<input type="checkbox"/> 2020.14255Shens	<input type="checkbox"/> phenoFatByUltraSound	<input type="checkbox"/> 30w	<input type="checkbox"/> NA
	<input type="checkbox"/> Lnu	<input type="checkbox"/> 2020.14667hens	<input type="checkbox"/> phenoCrest	<input type="checkbox"/> 31w	
	<input type="checkbox"/> N4N4	<input type="checkbox"/> 2020.1merge922	<input type="checkbox"/> phenoBodyTemperature	<input type="checkbox"/> 38w	
	<input type="checkbox"/> P2P2	<input type="checkbox"/> 2020.1merge922hens	<input type="checkbox"/> phenoBodyHematocrit	<input type="checkbox"/> 3w	
	<input type="checkbox"/> RpRm	<input type="checkbox"/> 2021.1	<input type="checkbox"/> phenoBloodGlycemiaCetonomia	<input type="checkbox"/> 4w	

Show 10 entries

T1	Project	T1	Pop	T1	Batch	T1	Datatype	T1	Age Factor	T1	Feed Factor	T1	Heat Factor	T1
<input type="checkbox"/>	MDfastedFed		com8		NA		individualPedigree		NA		NA		NA	
<input type="checkbox"/>	MDfastedFed		com8		NA		moleculeExtractionIs		4w		CTvsFASTEDvsREFED		NA	
<input type="checkbox"/>	chickStress		Fayo		NA		individualPedigree		NA		NA		NA	
<input type="checkbox"/>	chickStress		Fayo		NA		phenoBW		17w		CTvsLE		CTvsHS	
<input type="checkbox"/>	chickStress		Fayo		NA		phenoBW		26w		CTvsLE		CTvsHS	
<input type="checkbox"/>	chickStress		Fayo		NA		phenoBW		31w		CTvsLE		CTvsHS	
<input type="checkbox"/>	chickStress		Fayo		NA		phenoBW		38w		CTvsLE		CTvsHS	
<input type="checkbox"/>	chickStress		Fayo		NA		phenoBW		8w		CTvsLE		CTvsHS	
<input type="checkbox"/>	chickStress		Fayo		NA		phenoBloodHematocrit		NA		CTvsLE		CTvsHS	
<input type="checkbox"/>	chickStress		Fayo		NA		phenoBodyTemperature		NA		CTvsLE		CTvsHS	

Showing 1 to 10 of 131 entries

Select all Deselect all Download selection Join selection

Export the table in it's current state: Copy Excel CSV

Auteurs : Quentin Boone - Philippe Bardou - Cedric Cabau - Christophe Klopp

FROGS 4.1

Restructuration d'outils et ajout de nouvelles fonctionnalités dont :

FROGS_1 Pre-process merging, denoising and dereplication (Galaxy)

Sequencer

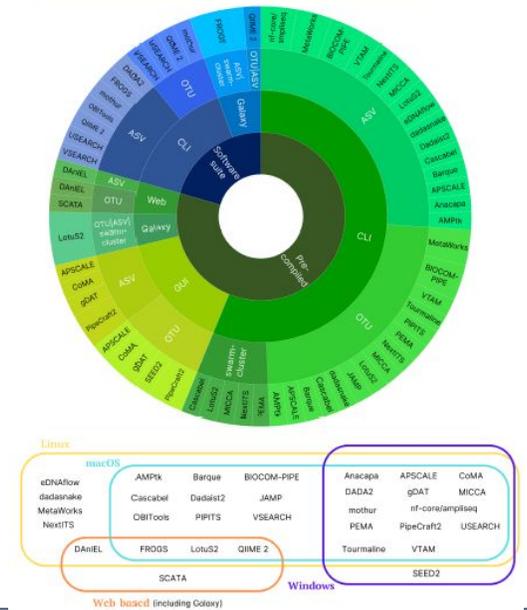
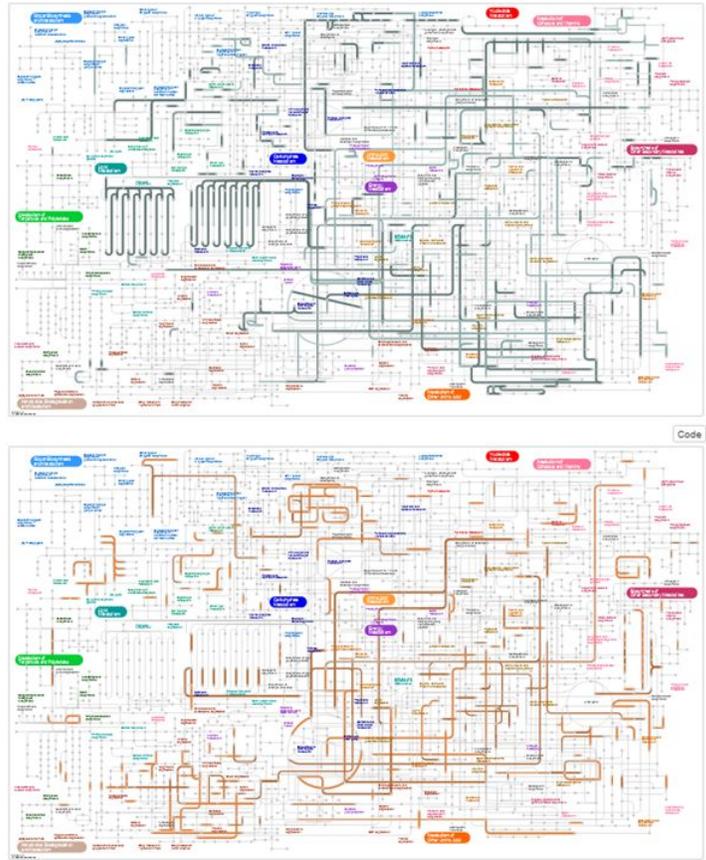
Longreads (PACBIO, ONT)

454

TAR archive file

prise en charge long reads

intégration de l'inférence fonctionnelle en analyse différentielle (Picrust + DESeq2)



FROM THE COVER

MOLECULAR ECOLOGY RESOURCES | WILEY

A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses

Ali Hakimzadeh¹ | Alejandro Abdala Asburn² | Davide Albanese³ | Maria Bernard^{4,5} | Dominik Buchner⁶ | Benjamin Callahan⁷ | J. Gregory Caporaso⁸ | Emily Curd⁹ | Christophe Djemiel¹⁰ | Mikael Brandström Durling¹¹ | Vasco Elbrecht⁶ | Zachary Gold¹² | Hyun S. Gweon^{13,14} | Mehrdad Hajibabaei¹⁵ | Falk Hildebrand^{16,17} | Vladimir Mikryukov¹ | Eric Normandeau¹⁸ | Ezgi Özkurt^{16,17} | Jonathan M. Palmer¹⁹ | Géraldine Pascal²⁰ | Teresita M. Porter¹⁵ | Daniel Straub²¹ | Martti Vasar¹ | Tomáš Větrovský²² | Haris Zafeiropoulos²³ | Sten Anslan¹

mutation OTU ⇨ ASV

Auteurs : Maria Bernard, Vincent Darbot, Olivier Rué, Lucas Auer, Géraldine Pascal

YapOTU: librairie python pour construire, visualiser et analyser des OTUs en barcoding & métabarcoding

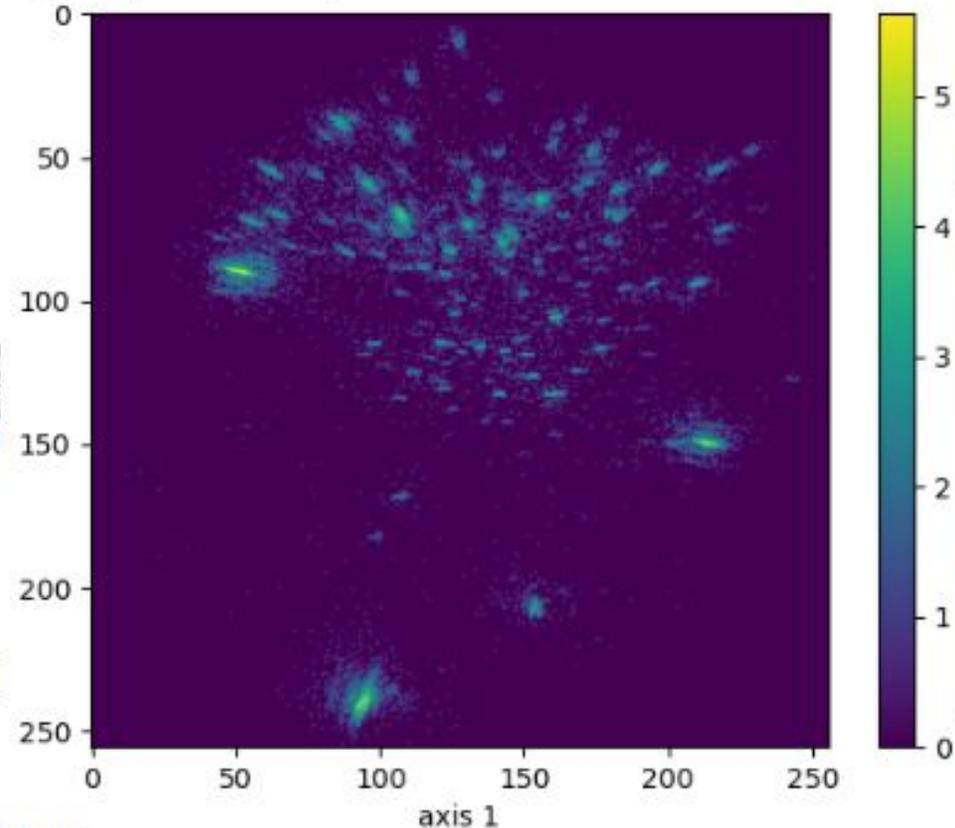
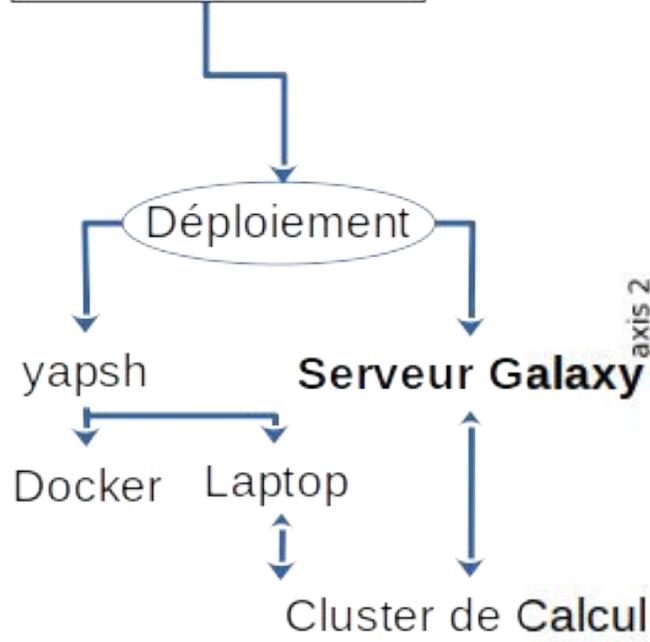


Code libre:
<https://gitlab.inria.fr/biodiversiton/yap>

Productions accessibles:
<https://entrepot.recherche.data.gouv.fr/dataverse/Metabarcoding>

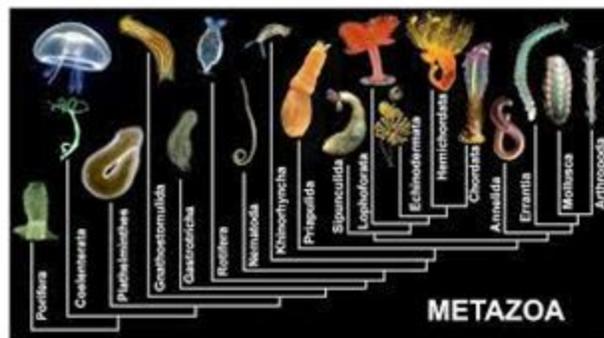
Autres librairies scientifiques
 PhyML, seaview, ...

Utilisateurs / Testeurs	
Valérie Laval	INRAE Bioger
Maria Kahlert	SLU Uppsala
Diego Fontaneto	CNR Verbana



Projets
R-syst
Bases de données
Guyane
Champignons
Diatomées
Métabarcoding
MALABAR (Arcachon)

Ilôts CpG et génomes d'animaux



Création d'un pipeline Nextflow

<https://forgemia.inra.fr/lpgp/cgi>

Auteur: Aurélien
BRIONNE

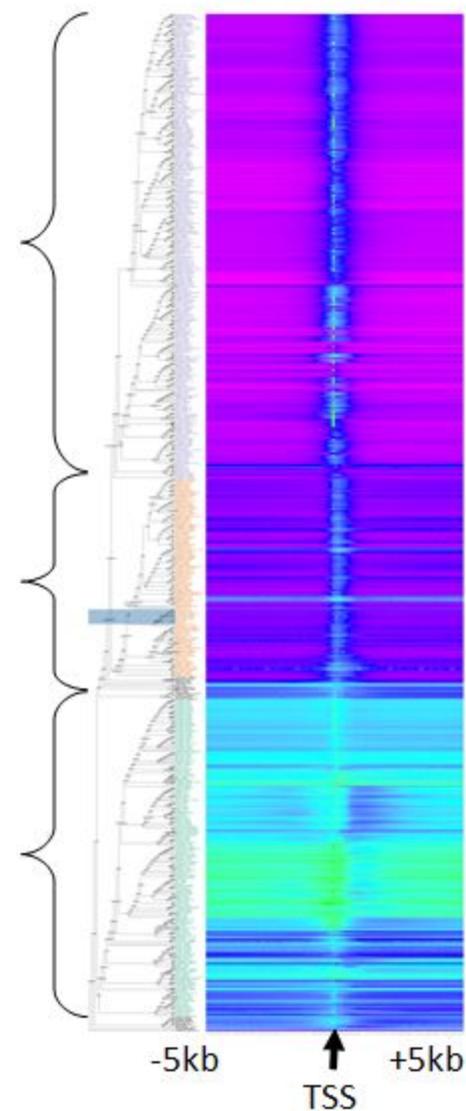
Exploitation en cours

Tétrapodes

Actinopterygii

Protostomiens

Arbre de la vie



autour de la formation

Présentation des métiers de bioinformaticien.ne et biostatisticien.ne à la fête de la science 2023



Atelier 4 : goûter

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

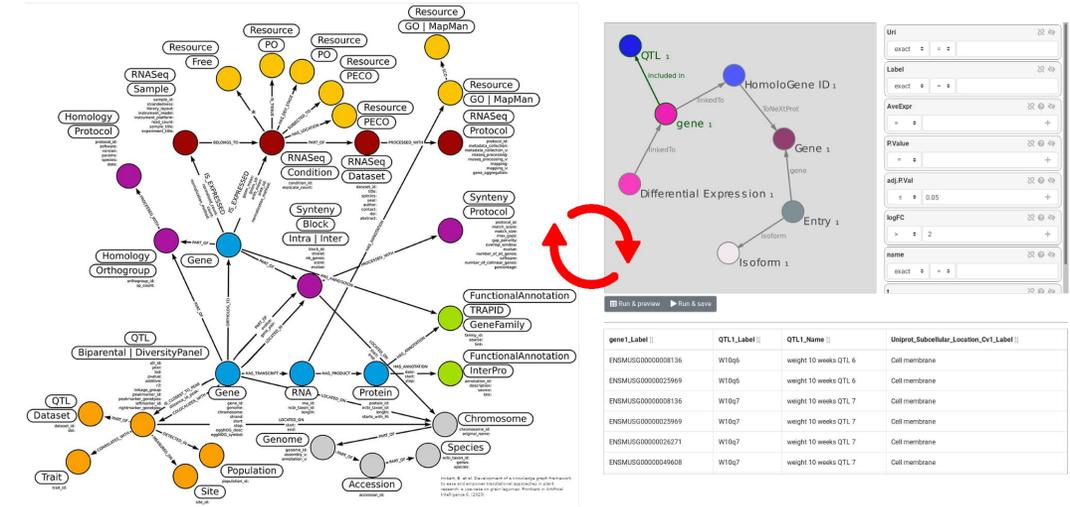
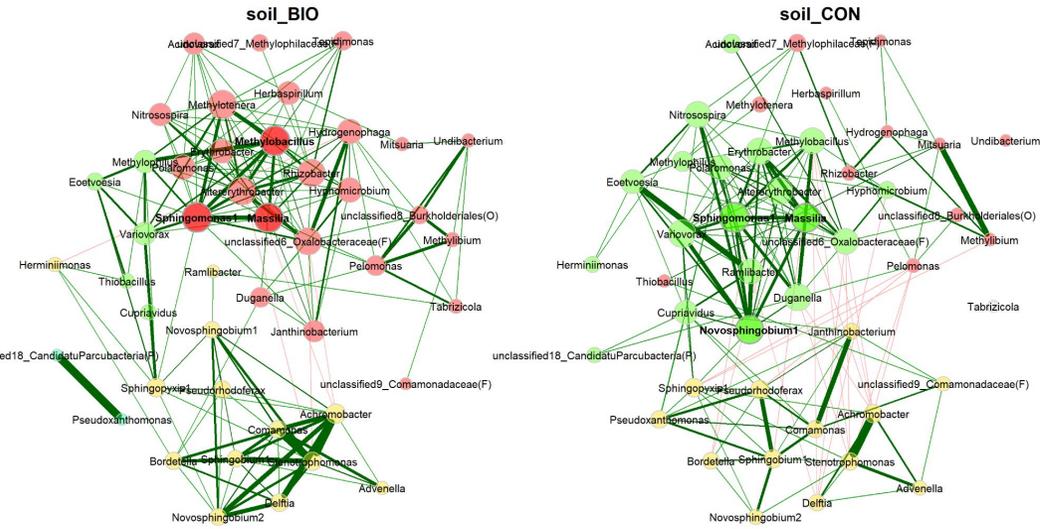
Supports partageables / réutilisables

Auteurs : Gabryelle Agoutin, Lucas Auer, Cédric Cabau, Vincent Darbot, Cervin Guyomar, Luc Jouneau, Géraldine Pascal

InterCATI - Roscoff - 18-21 septembre 2023

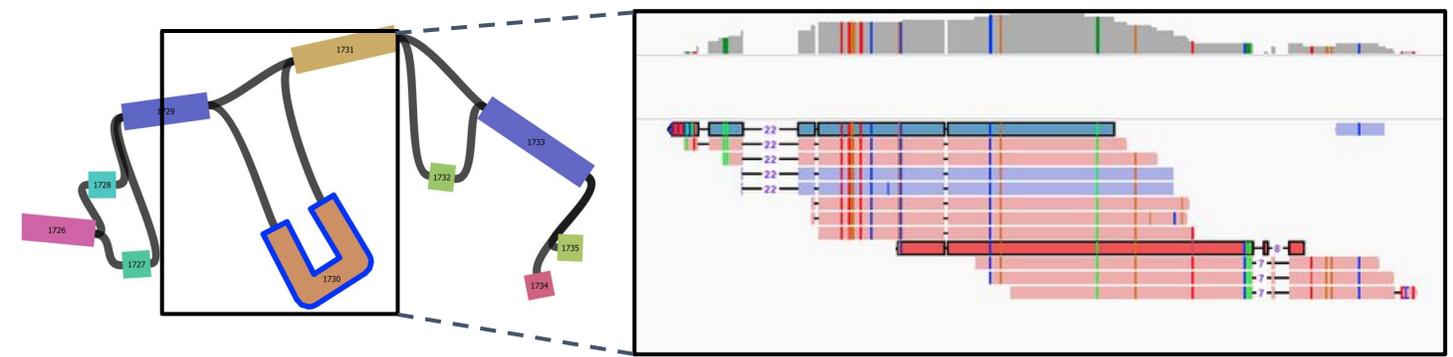
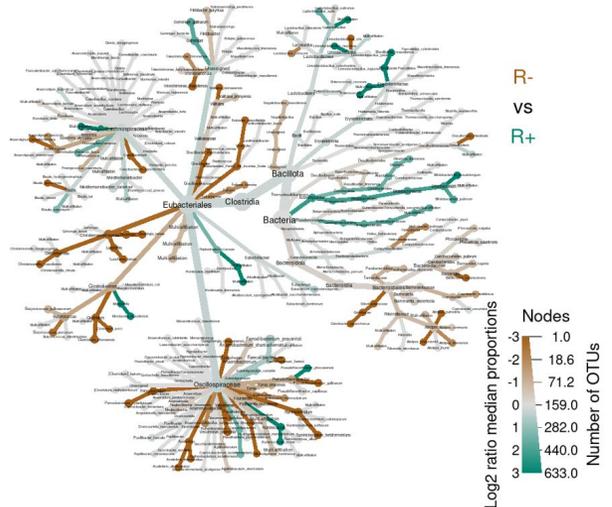
Atelier 1 : inférence de réseaux en métagénomique

Atelier 4 : Interopérabilité entre graphes RDF et graphes de Neo4j



gene1_Label	QTL1_Label	QTL1_Name	UniProt_Subcellular_Location_Ov1_Label
ENSMUSG00000008136	W10q6	weight 10 weeks QTL 6	Cell membrane
ENSMUSG00000023969	W10q5	weight 10 weeks QTL 6	Cell membrane
ENSMUSG00000008136	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG00000023969	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG00000026271	W10q7	weight 10 weeks QTL 7	Cell membrane
ENSMUSG00000049608	W10q7	weight 10 weeks QTL 7	Cell membrane

Atelier 3 : workflow de graphes de pangénomique



Auteurs : Vincent Darbot, Claire Hoede, Philippe Bordron, Ludovic Duvaux, Cervin Guyomar, Maria Bernard, +36 participants

infra/systeme/BD

De genologin à genobioinfo

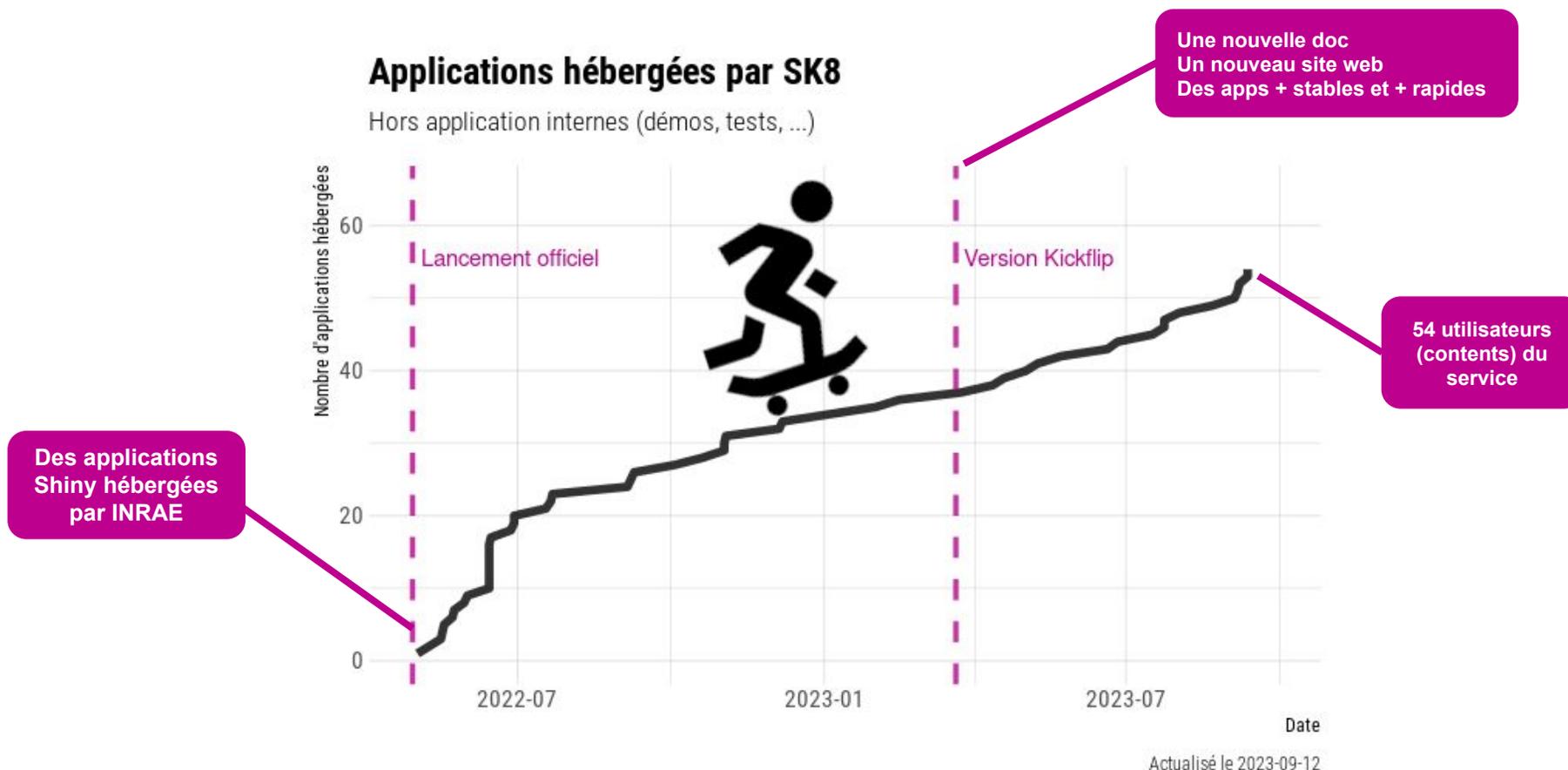


- Migration sur le nouveau cluster : utilisateurs, logiciels, projets, sites web
 - Point sur les 2 infrastructures de calcul : INRAE et Clement ADER
 - Point sur les nouveaux investissements en cours et à venir
-
- Auteurs : **Marie-Stéphane Trotard, Patrice Dehais, Didier Laborie**

Des nouvelles de SK8 « un projet qui roule comme sur des roulettes »



Projet inter-CATI, infrastructure INRAE pour déployer des application R shiny.
~54 demandes d'hébergement depuis l'ouverture du service



CI/CD



docker



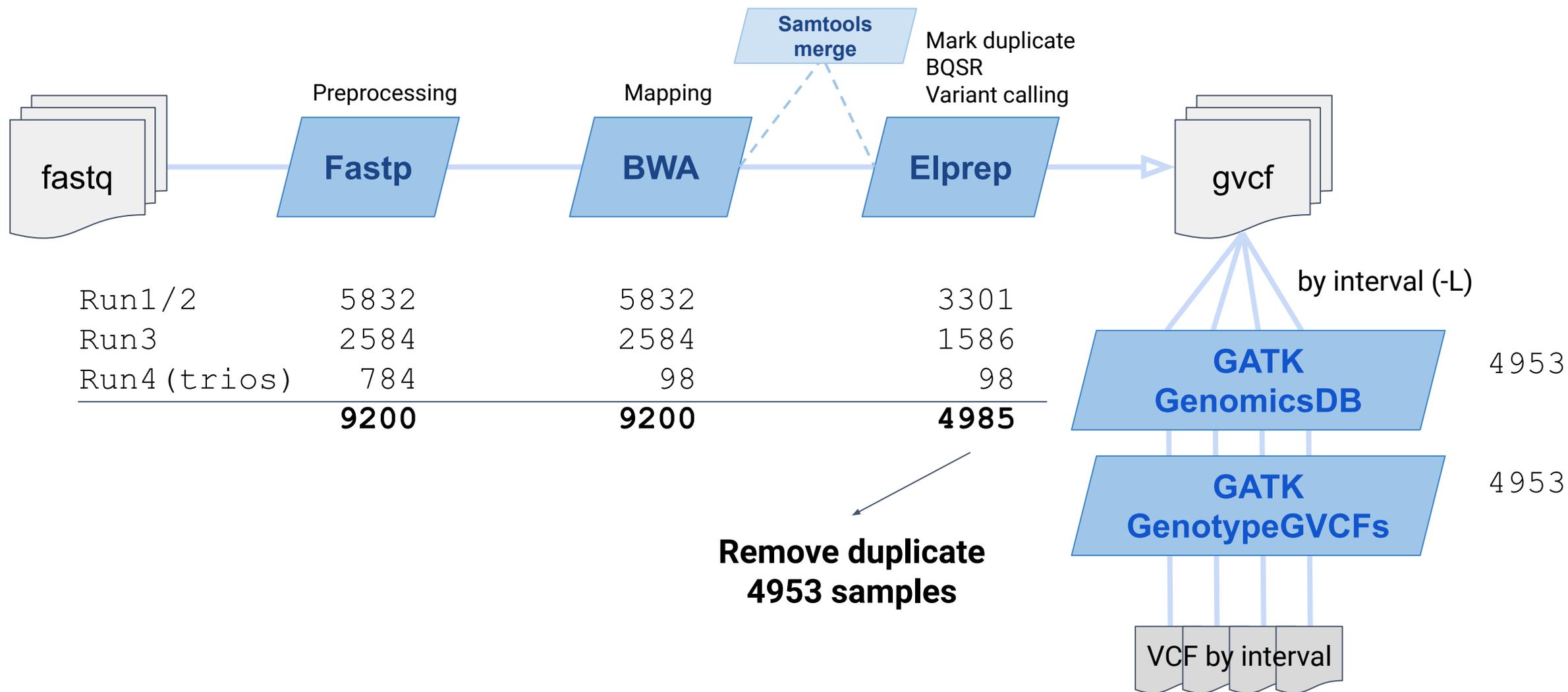
openstack.



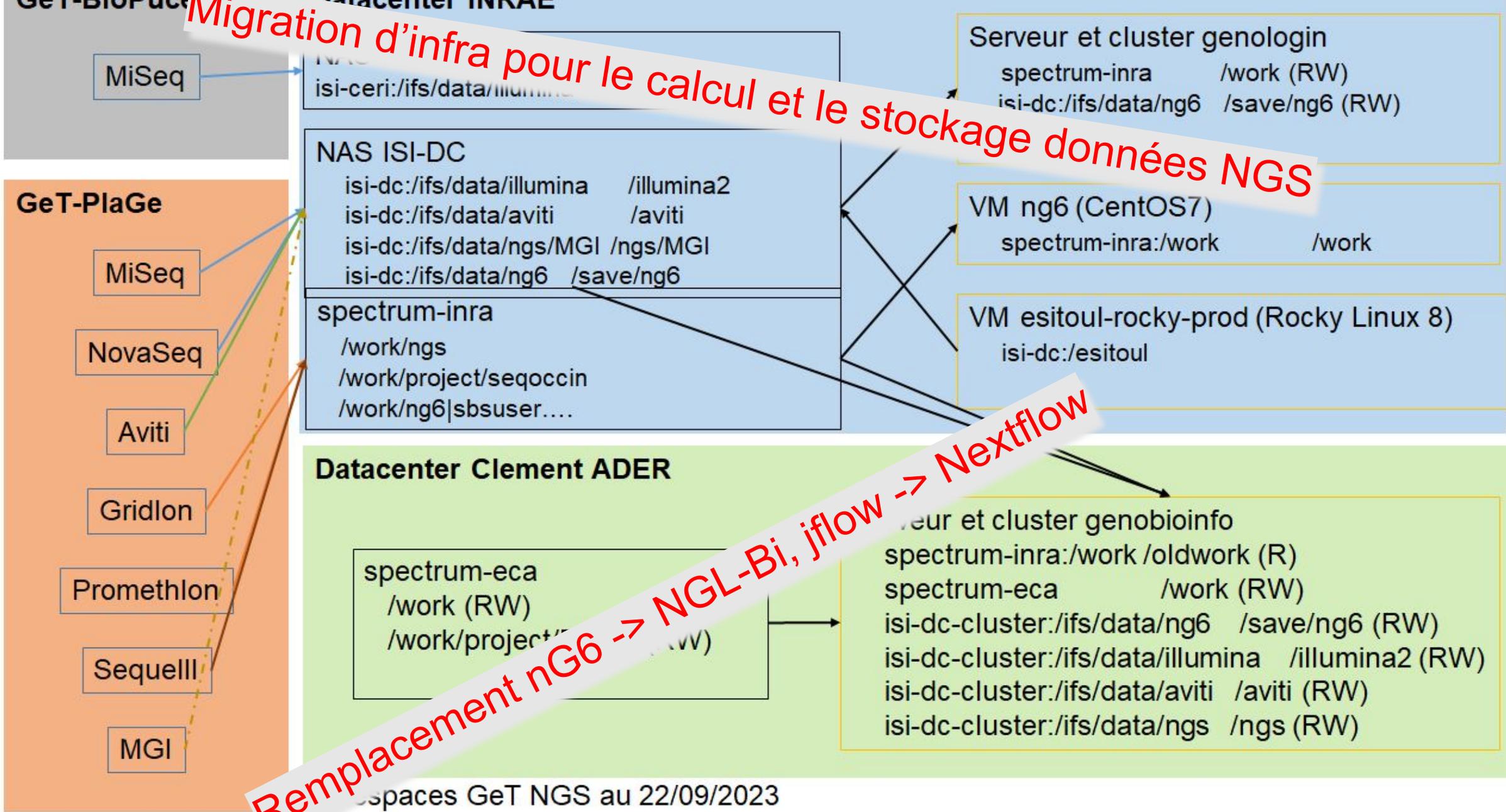
Auteurs : ~20 ITA INRAE, 10 CATI représentés.

Porté par Jean-François Rey (CATI IMOTEP), pour le CATI Bios4Biol : **Elise Maigné, Yannick Lippi**

1000G Chicken



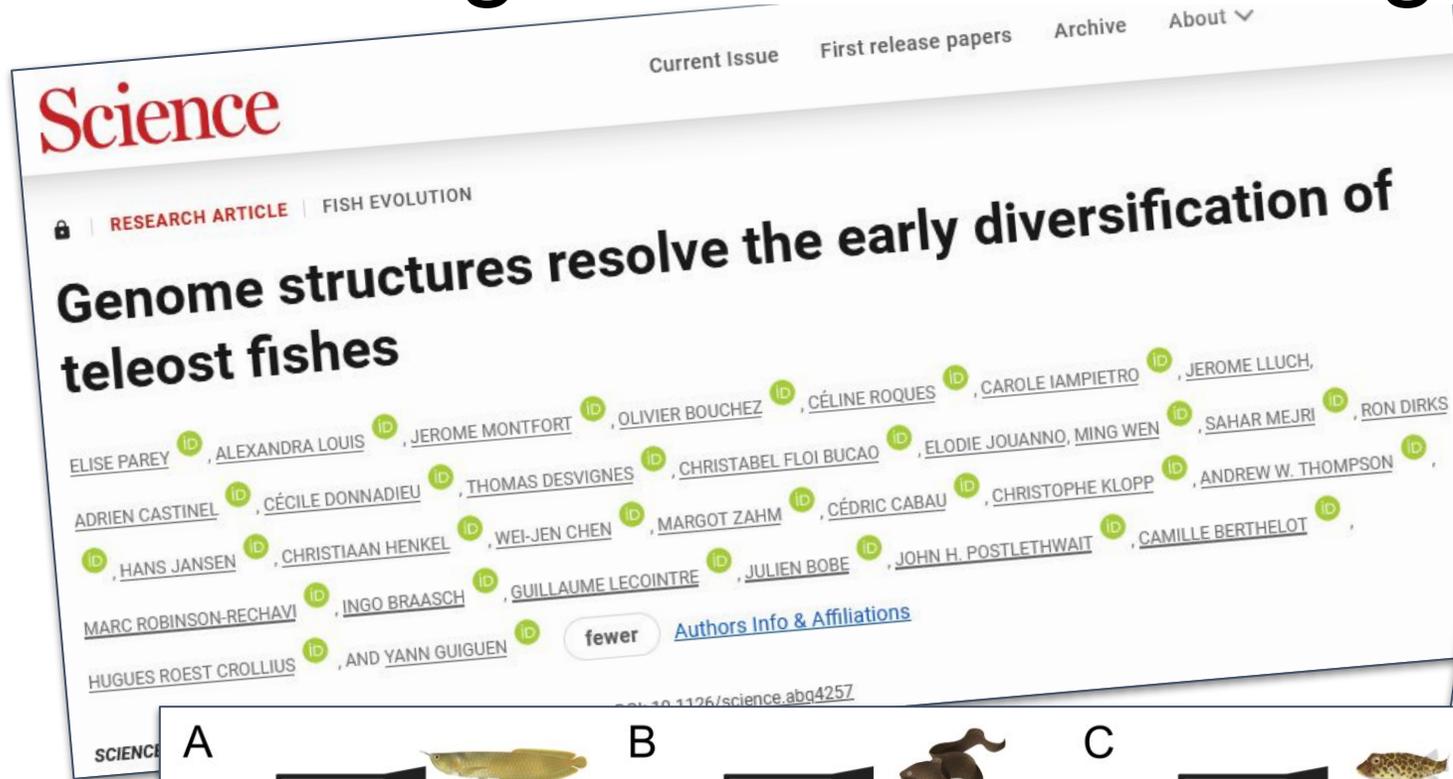
Migration d'infra pour le calcul et le stockage données NGS



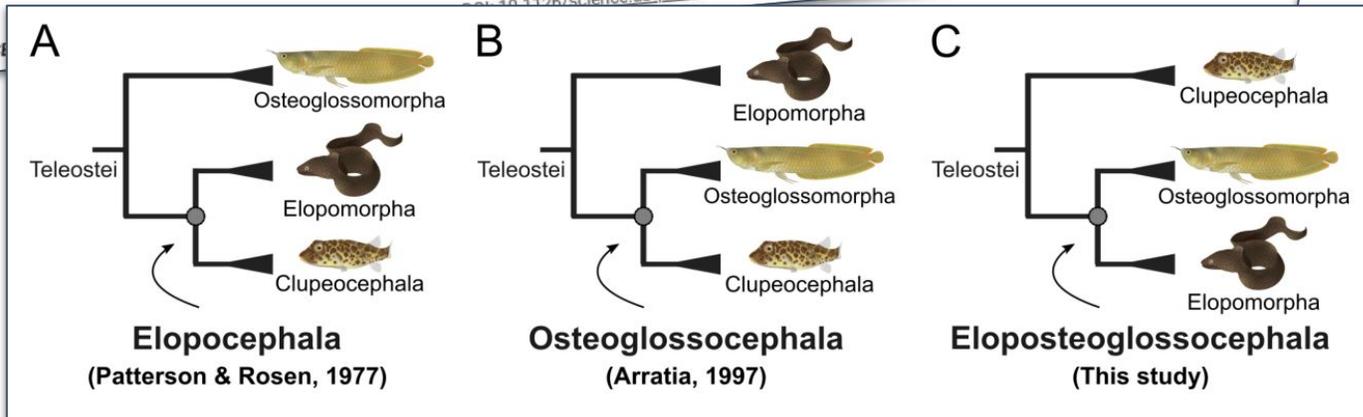
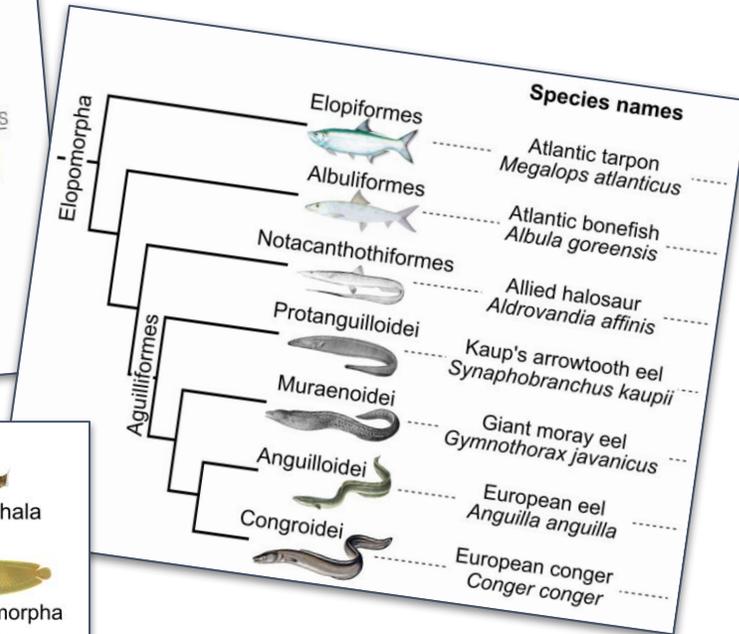
espaces GeT NGS au 22/09/2023

autres thèmes/domaines

Assemblage & annotation de génomes



Production de 7 nouveaux génomes d'Elopomorphes



Auteurs : Jérôme Montfort, Cédric Cabau, Christophe Klopp

Site web du CATI BIOS4Biol

Site web développé avec wordpress.

Description du CATI, des membres, des événements

CATI BIOS4Biol
Bioinformatique et Statistique pour la Biologie.

ACCUEIL MEMBRES QUI SOMMES NOUS? EVENEMENTS LOGIN

Accueil

B/B Evénements à venir

SEP septembre 18 - septembre 21
18 Hackathon Inter-CATI omics 2023
[Voir Le Calendrier](#)

> 42 agents

- **7 départements** : GA (11), PHASE (10), MathNum (8), ECODIV (7), SA (3), INRAE Transfert (2), AQUA (1)
- **15 laboratoires** : BioForA, BIOGECO, BREED, DECOD, GABI, GenPhySE, GeT-PlaGe, laM, LPGP, MIAT, PRC, ToxAlim, UMRH, UR EABX, VIM
- **8 centres INRAE** : Toulouse, Jouy, Rennes, Theix, Val de Loire (Nouzilly et Orleans), Nancy, Bordeaux
- **5 plates-formes** (Get-PlaGe, Genotoul Bioinfo, Get-Trix, Get-IT, Sigenae)

Map of France showing agent locations: Lille, Amiens, Rouen, Reims, Nancy/Strasbourg, Dijon, Clermont-Ferrand, Toulouse, Bordeaux, Nantes, Poitiers, Cherbourg, Caen, Marseille, Ajaccio.



Auteurs : Jérémy Tournayre, Marie-Stéphane Trotard, Anne Frambourg

Rappel : qu'est-ce qu'une production CATI ?

Les productions du CATI se définissent par des productions qui

- émanent de discussions/idées/rerelations d'agents au sein du CATI
 - sont des projets menés par plusieurs membres du CATI
-
- Une production CATI peut être une production issue d'une plateforme
 - Une personne/un groupe peut indiquer dans ce document plusieurs projets mais une diapo = une production
 - Modèle demandé = diapo 3
 - Exemple = diapo 4

Lancement du service SK8

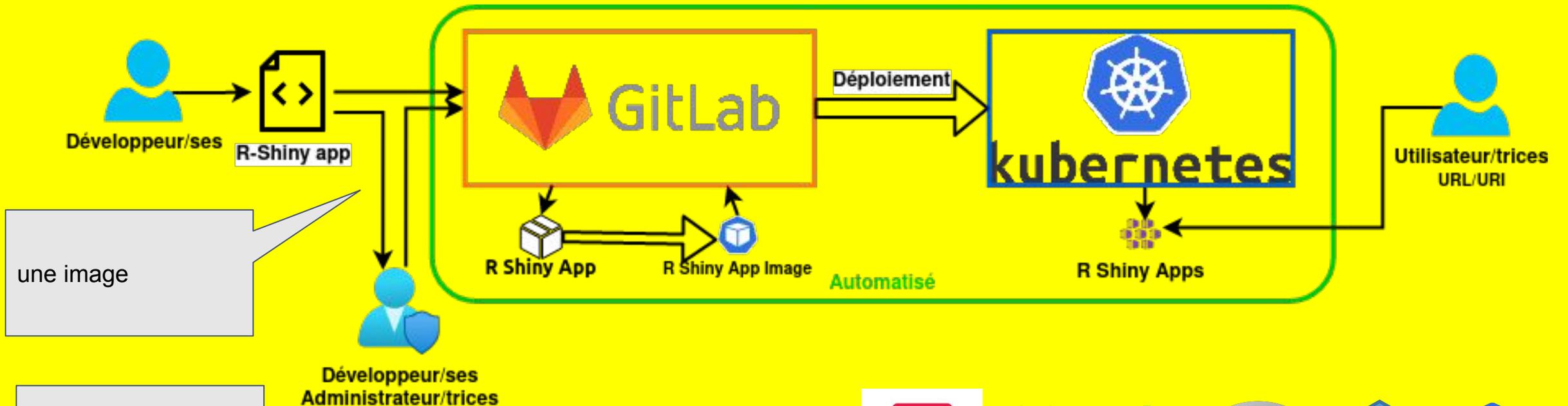
Titre du projet/outil/formation

SK8

Infrastructure pour déployer des application R shiny.

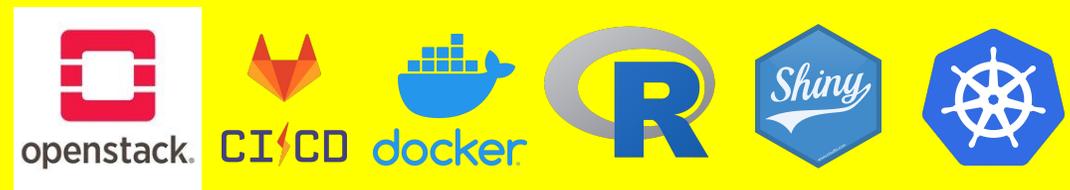
16 demandes d'hébergement depuis ouverture à INRAE en avril 2022 :

bref descriptif



une image

les auteurs avec membres du CATI en gras



Auteurs : ~30 personnes, 9 CATI représentés.

Porté par Jean-François Rey (CATI IMOTEP), pour le CATI Bios4Biol : **Elise Maigné**