une image = un projet

AG CATI BIOS4Biol 2022

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

Les productions du CATI se définissent par des productions qui

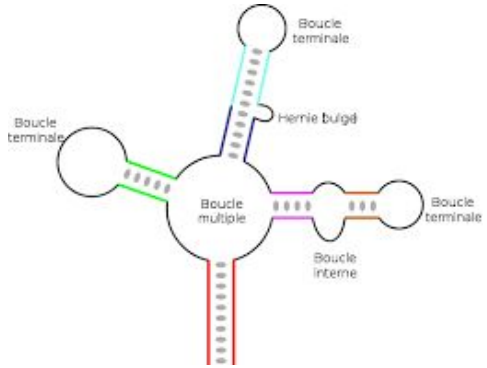
- émanent de discussions/idées/relations d'agents au sein du CATI
- sont des projets menés par plusieurs membres du CATI

– 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

pipelines, développements, benchmarking

Pipeline PAQMir

Chaîne de traitement dédiée à l'analyse et l'annotation des petits ARNs



Alignement → Comptage (UMI) → Annotation

Epigenetics Chromatin. 2021 May 24;14(1):24. doi: 10.1186/s13072-021-00397-5.

Dynamics of cattle sperm sncRNAs during maturation, from testis to ejaculated sperm

Eli Sellem [1], Sylvain Marthey [2 3], Andrea Rau [2 4], Luc Jouneau [5 6], Aurelie Bonnet [7], Chrystelle Le Danvic [7], Benoît Guyonnet [8], Hélène Kiefer [5 6], Hélène Jammes [5 6], Laurent Schibler [7]

RESEARCH Open Access

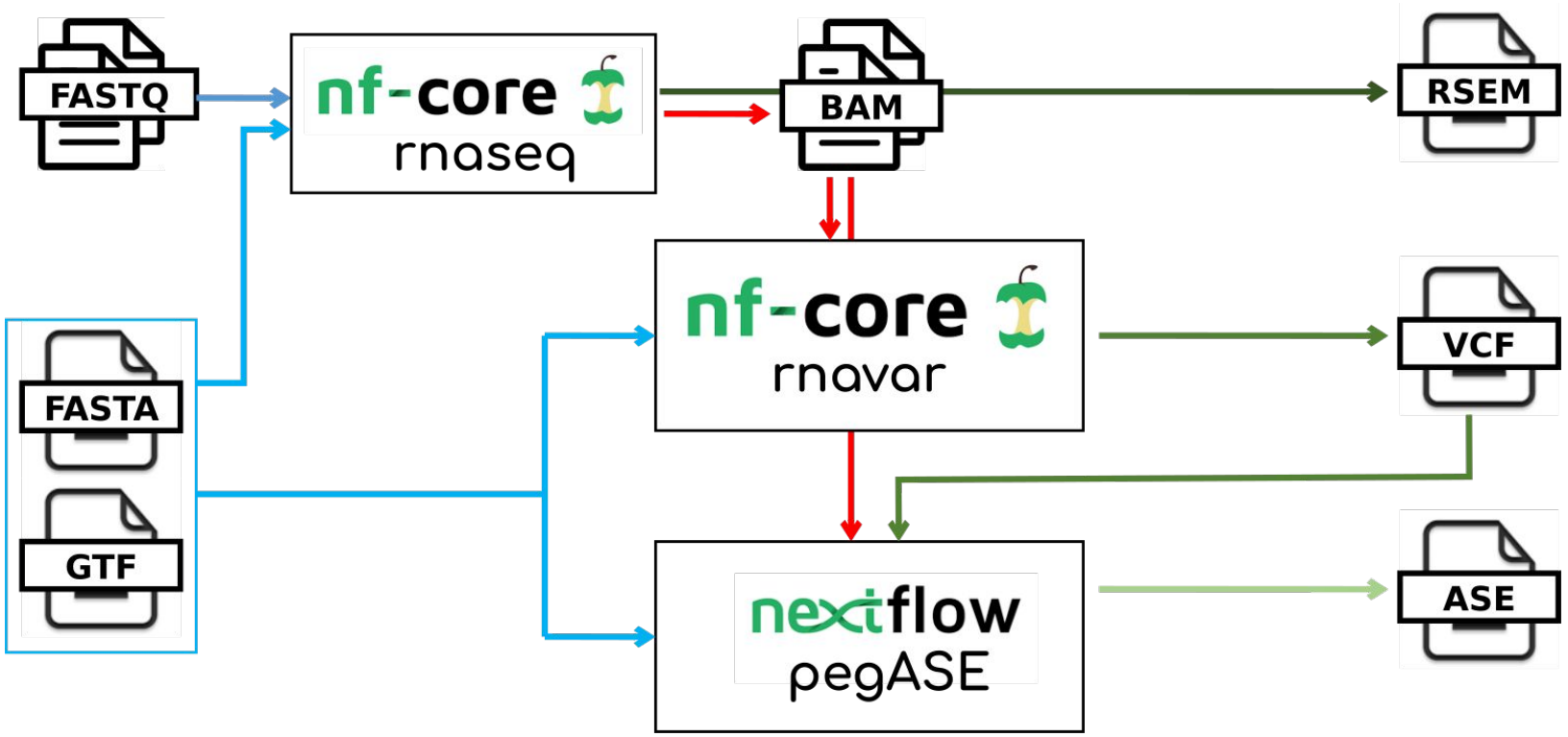
A comprehensive overview of bull sperm-borne small non-coding RNAs and their diversity across breeds

Eli Sellem[1*], Sylvain Marthey[2], Andrea Rau[2], Luc Jouneau[3,4], Aurelie Bonnet[1], Jean-Philippe Perrier[3,4], Sébastien Fritz[1,2], Chrystelle Le Danvic[1], Mekki Boussaha[2], Hélène Kiefer[3,4], Hélène Jammes[3,4] and Laurent Schibler[1]

Auteurs : Sylvain Marthey, **Anne Frambourg**, **Luc Jouneau**, Eli Sellem, Valentin Costes

Analyse de variants RNA-seq 100% Nextflow

Utilisation / développement / amélioration de 3 pipelines pour l'analyse de variants à partir de données RNA-seq



Utilisation de nf-core rnaseq

Amélioration de nf-core rnavar
- Mapping en 2 passes amélioré
- Sortie gvcf

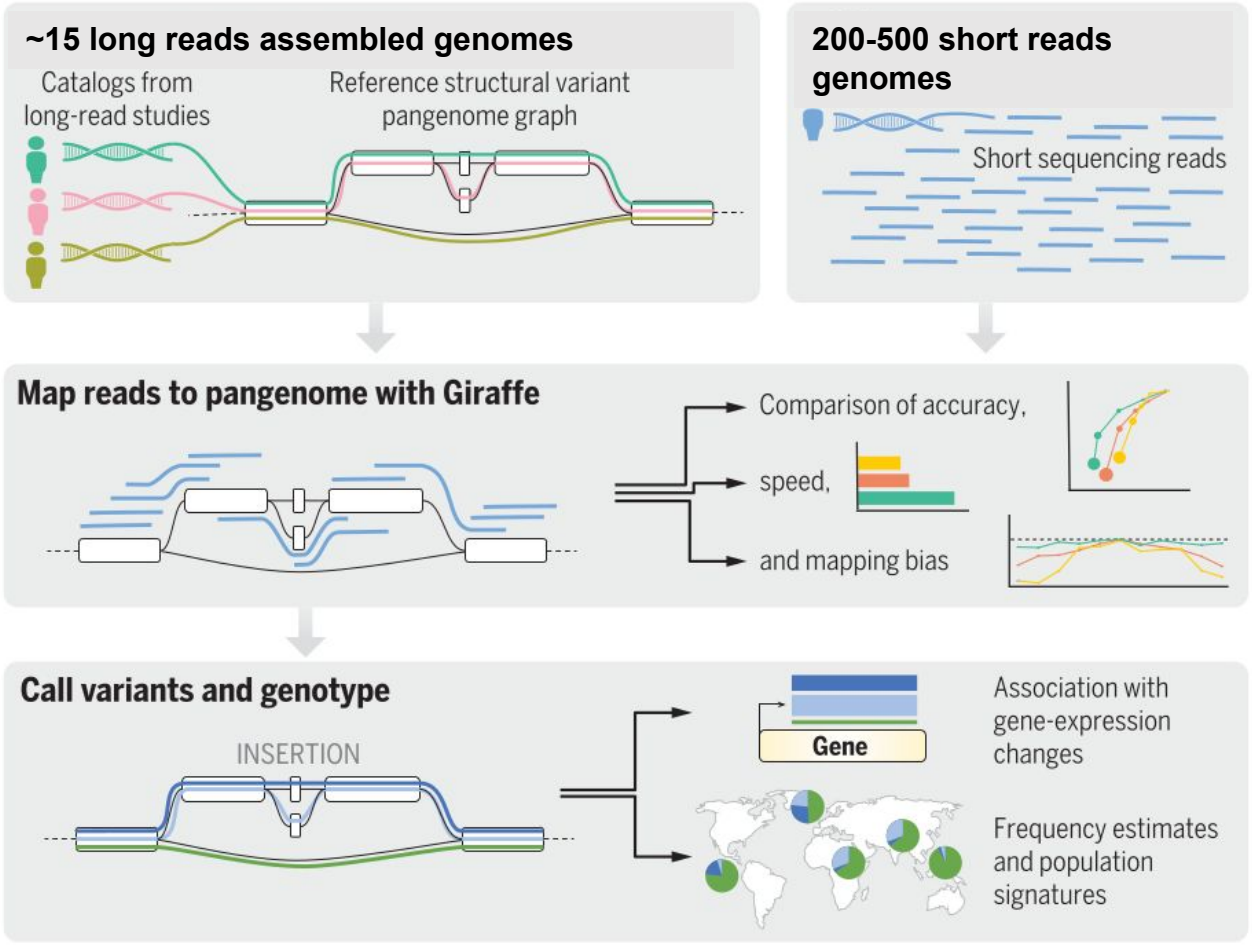
Création de pegASE
- Sur la base du travail (SnakeMake) de Maria Bernard
- Basé sur phASEr
- Au standard nf-core

Validation à grande échelle sur les données du projet EFFICACE et CO-LOcATION

Auteurs : **Maria Bernard, Mathieu Charles, Cervin Guyomar, Sarah Maman**

WoodySV - Pangenoak

Chaîne de traitement pour la construction de graphes de pangénomes afin de déterminer le rôle des variants structuraux dans l'adaptation locale chez le chêne et la domestication chez l'abricotier.

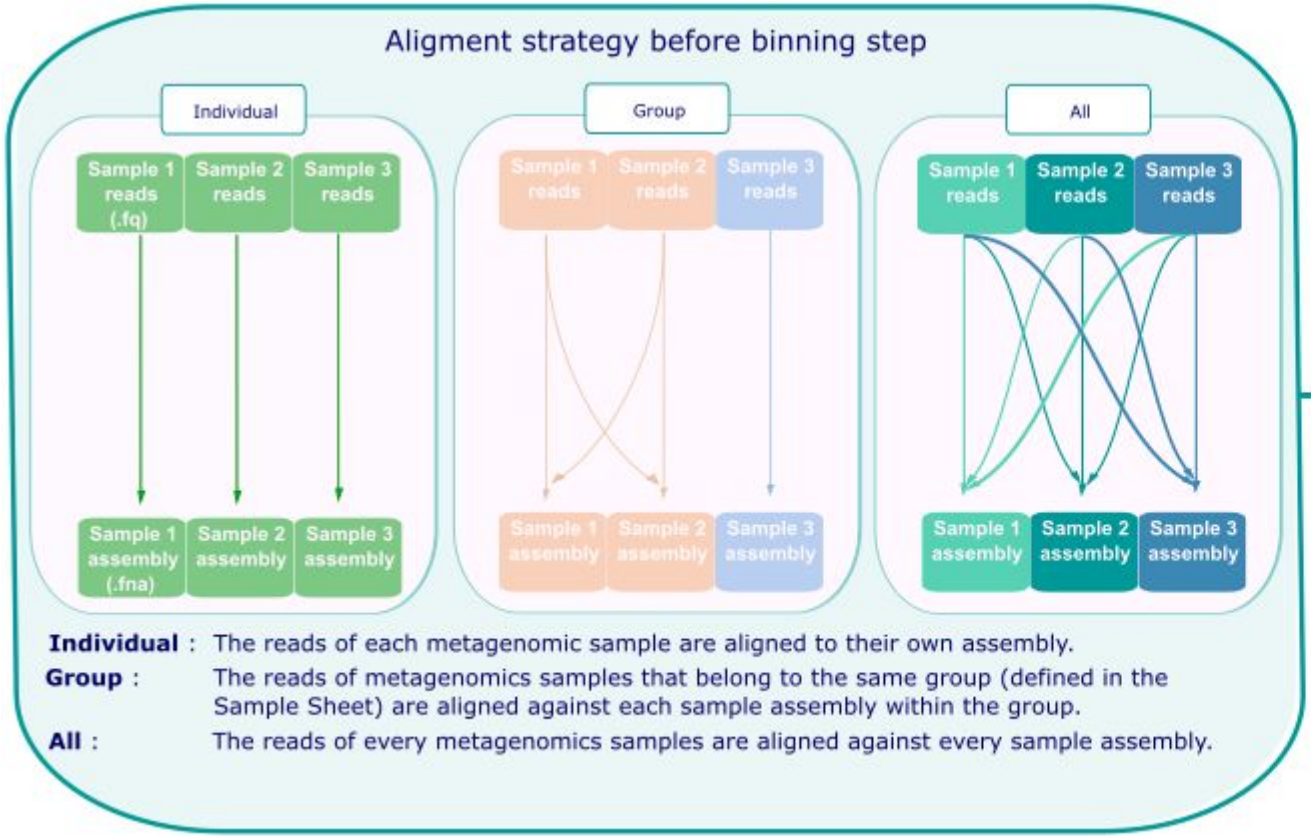


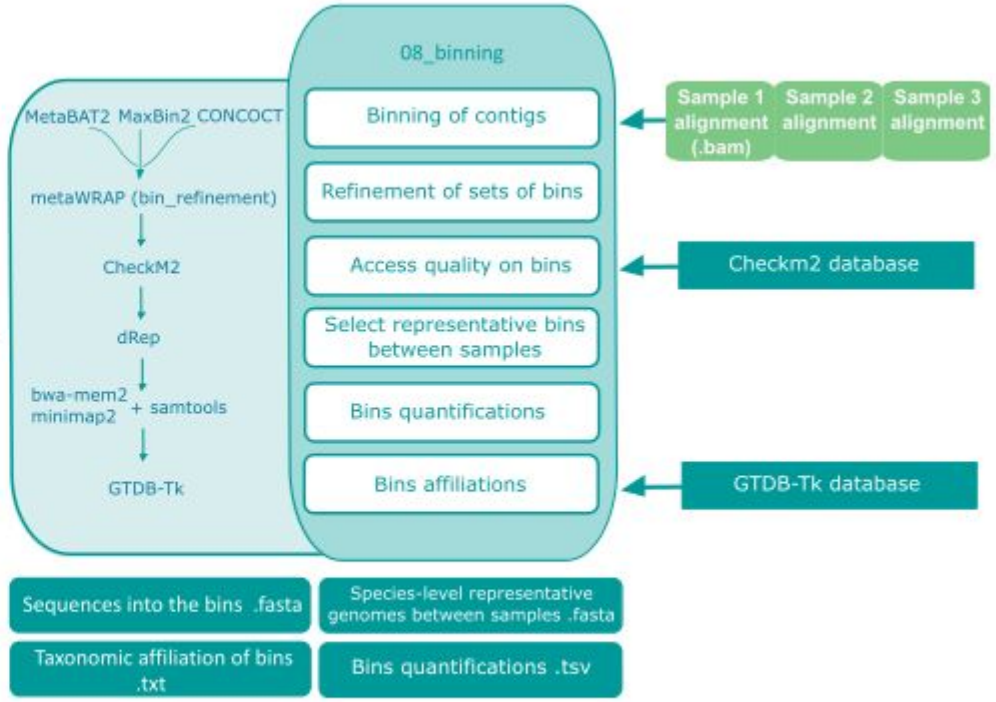
Modifié d'après Siren et al. 2021
10.1126/science.abg8871

Auteurs : Sukanya Denni, Quynh Trang-Bui, Véronique Decroocq, **Ludovic Duvaux**

Ajout du binning dans le workflow metagWGS

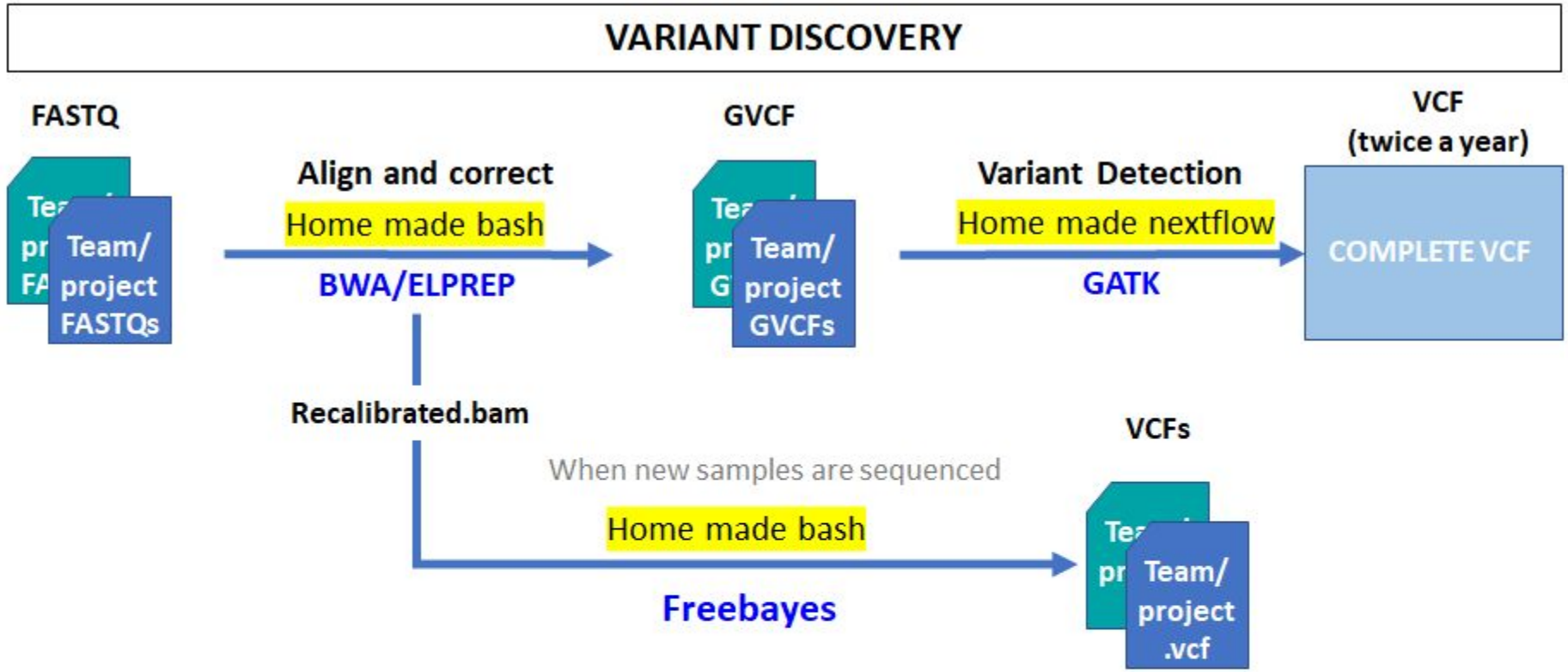
metagWGS 2.3 permet l'analyse des short-reads ou des long-reads HiFi PacBio.
Nouvelles fonctionnalités majeures : le binning des contigs couplé à la possibilité d'aligner plusieurs échantillons sur plusieurs assemblages.
Amélioration des performances de l'étape de raffinement des bins : temps d'exécution divisé par 7.





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

VarPipeline workflow description



Recherche de variants et génotypage sur de larges et nombreux jeux de données (nb animaux incrémentés en fonction de l'arrivée de nouveaux lots, ex: porc >100 animaux actuellement) de façon récurrente et complète à chaque nouvelle version d'assemblage.

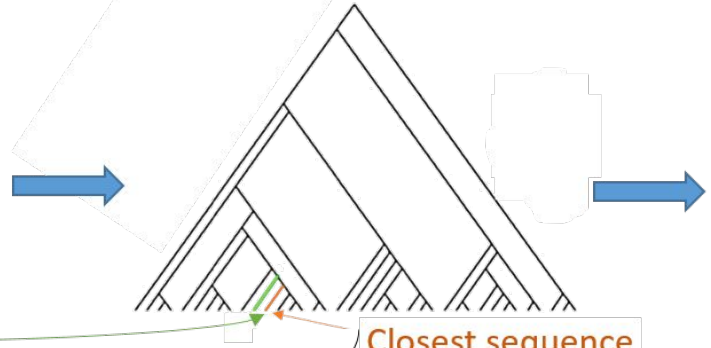
Du fait de la taille des lots à traiter, il a été nécessaire de générer de nouvelles chaînes de traitement en s'inspirant de celles existantes, en mettant l'accent sur l'économie de l'espace de stockage phagocyté par les fichiers temporaires de ces dernières, et en optimisant le parallélisme.

Scripts bash / Nextflow / Slurm / Genotoul / Singularity (portabilité)

Patrice Dehais, Bertrand Servin, **Sabrina Legoueix**

+ Picrust2 ➔ FROGSFUNC

	Taxonomic affiliation	sequence	Sample1	Sample2	Sample3
OTU1	Species A	>ACCGC…	3500	6300	210
OTU2	Species B	>TGCGG…	0	460	36
OTU3	Species C	>ATTGT…	400	700	500



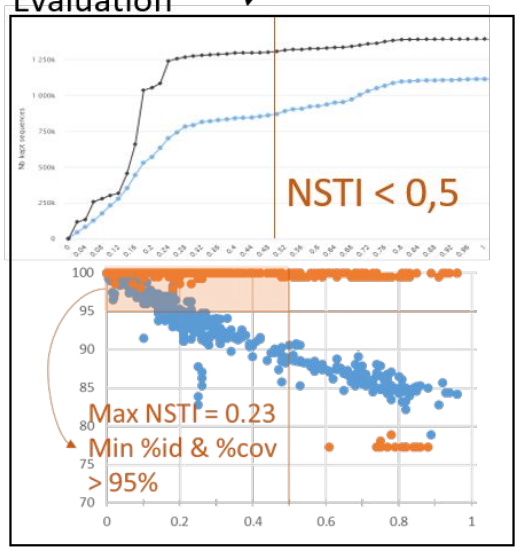
Closest sequence

	Name	Sample1	Sample2	Sample3
Func1	Func A	1301	1695	2206
Func2	Func B	31	19	2
Func3	Func C	10221	7684	10419

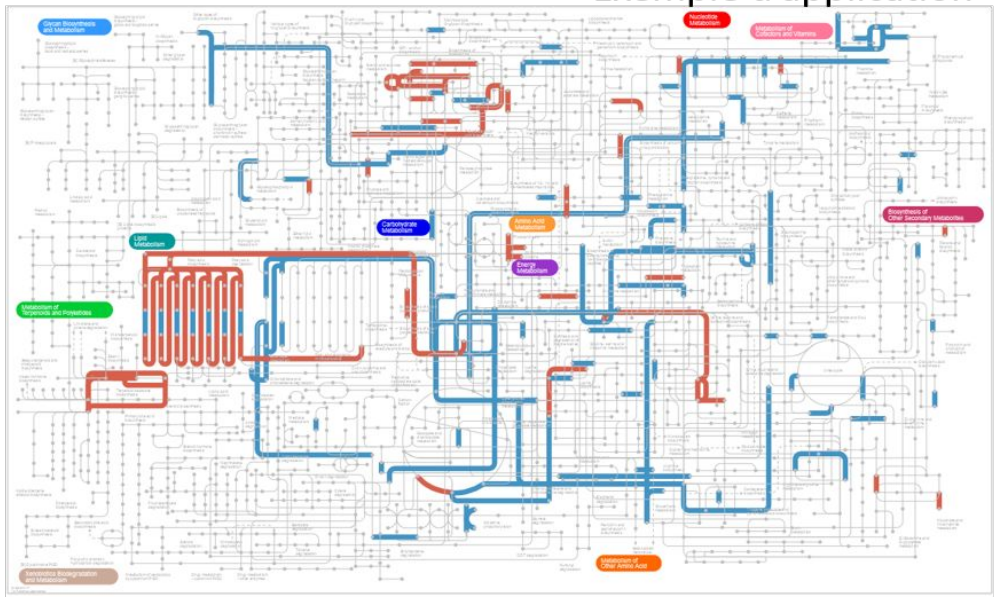
- Intégration des outils PICRUSt2 (4 outils) pour la prédiction et quantification fonctionnelle des écosystèmes microbiens.

- Application au microbiote ceacal de poules divergentes sur l'efficience alimentaire.

Evaluation



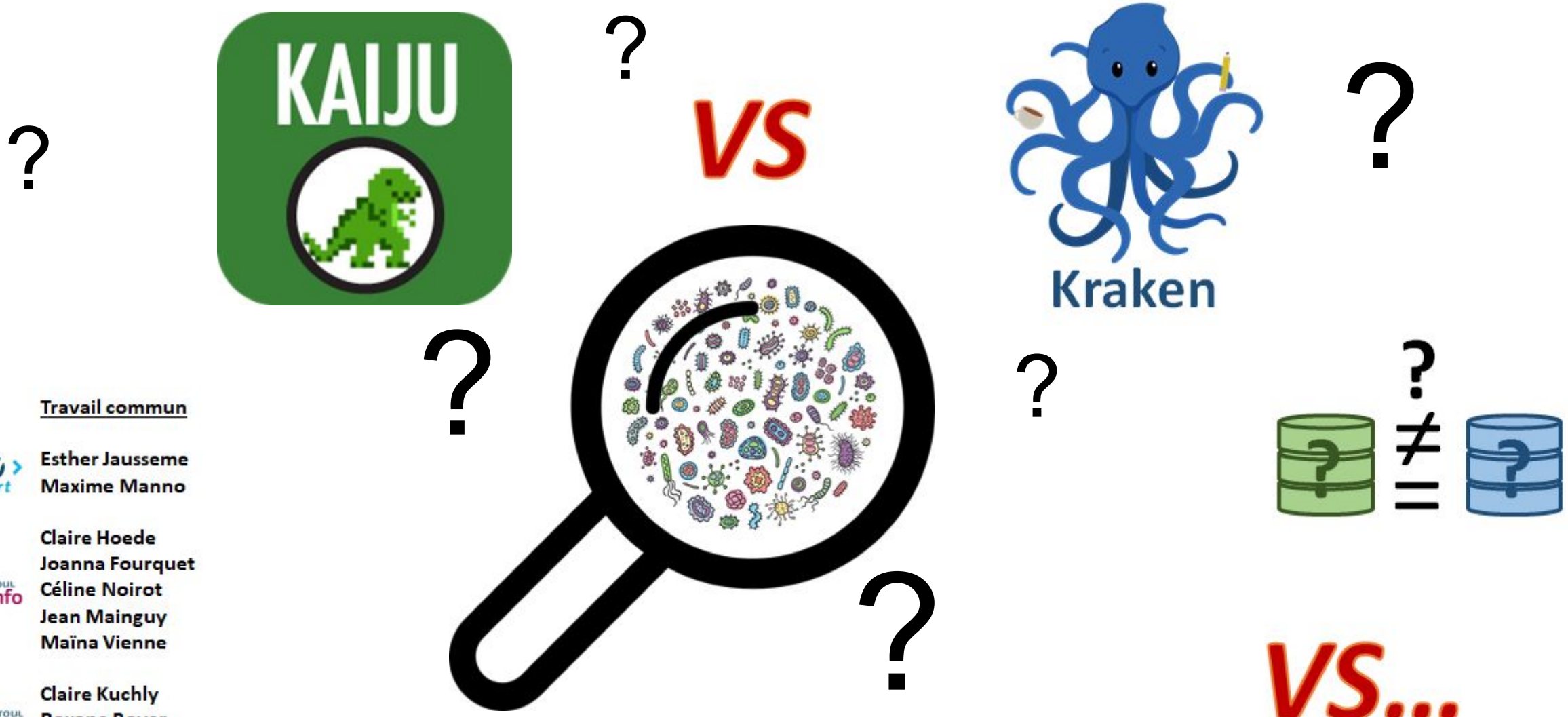
Exemple d'application



Auteurs : **Vincent Darbot,** Moussa Samb, **Maria Bernard,** Olivier Rué, **Géraldine Pascal**

Tests d'outils pour l'affiliation taxonomique de données WMS ou WGS

-> Échanges autour de l'affiliation taxonomique de données de **Whole Metagenome Sequencing** ou **Whole Genome Sequencing** en vue de faire de la recherche de contamination ou étudier la composition d'une communauté complexe.



Travail commun

INRAE transfert
Esther Jausseme
Maxime Manno

Genotoul Bioinfo
Claire Hoede
Joanna Fourquet
Céline Noirot
Jean Mainguy
Maïna Vienne

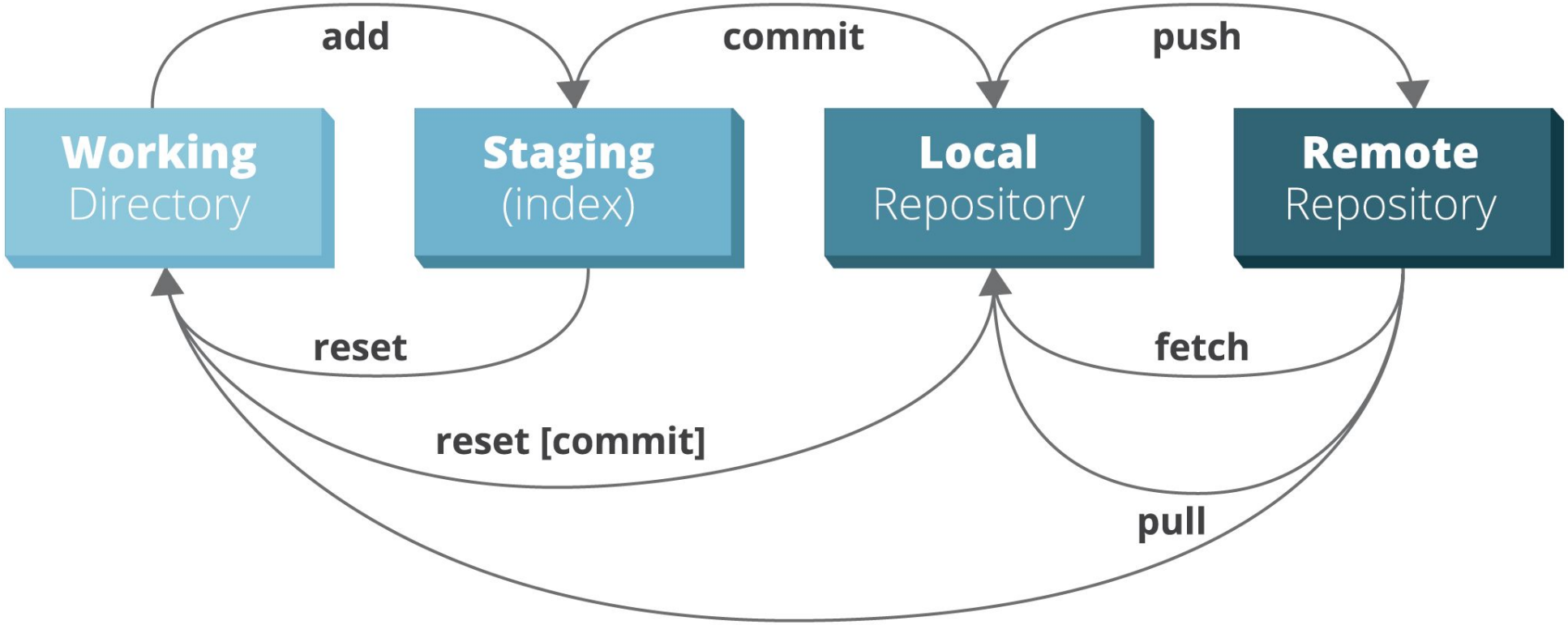
Genotoul GeT
Claire Kuchly
Roxane Boyer
Eden Darnige

Moyens : Plusieurs réunions effectuées, discussions informelles et échanges de scripts

autour de la formation

Formation GIT

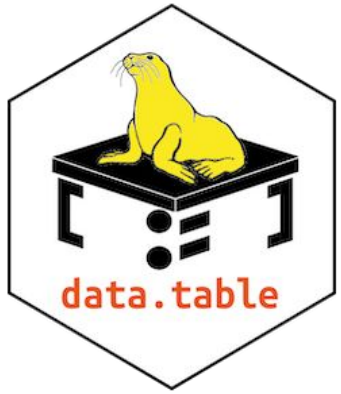
Organisation de 9 sessions de formation à GIT, à Toulouse et Jouy en Josas. 3 sessions débutant et 6 session avancée. 54 personnes formées dont 30 du CATI BIOS4Biol.

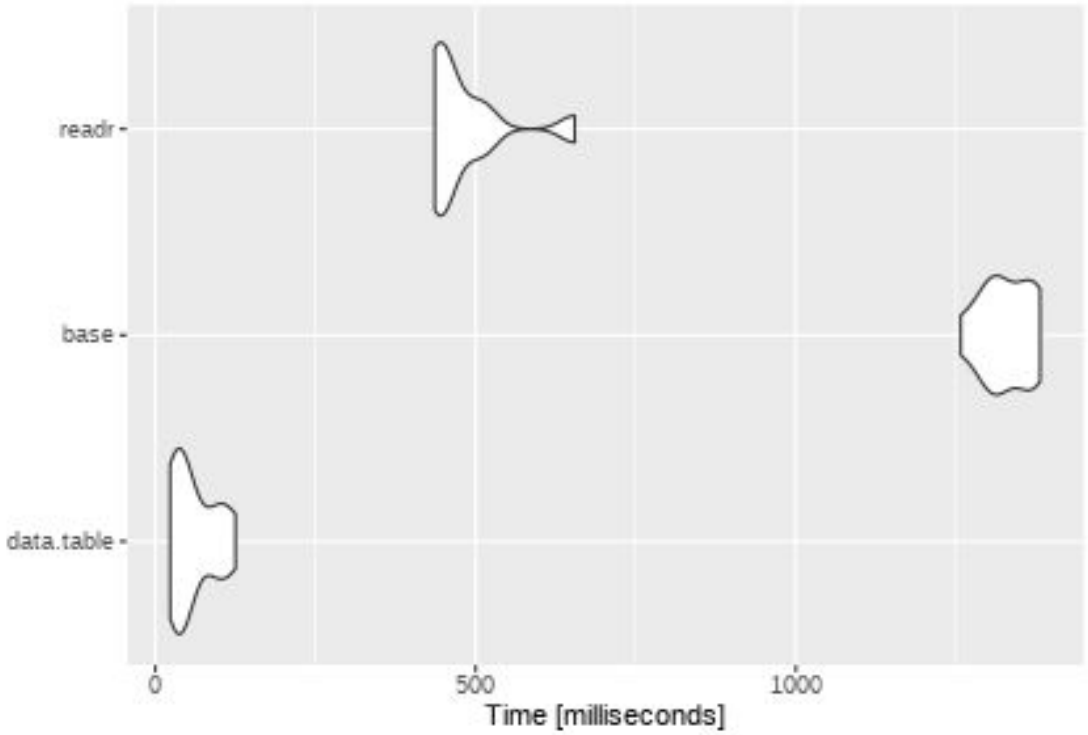


Auteurs : **Céline Noirot**, Estelle Ancelet, **Maria Bernard, Luc Jouneau**, formation donnée par Makina Corpus.

Session de formation R avancé : le package data.table

Organisation d'une formation interne (MIAT) sur 1 jour sur le package R data.table.
data.table = « high performance data.frame ». 7 personnes formées dont 3 du CATI bios4biol





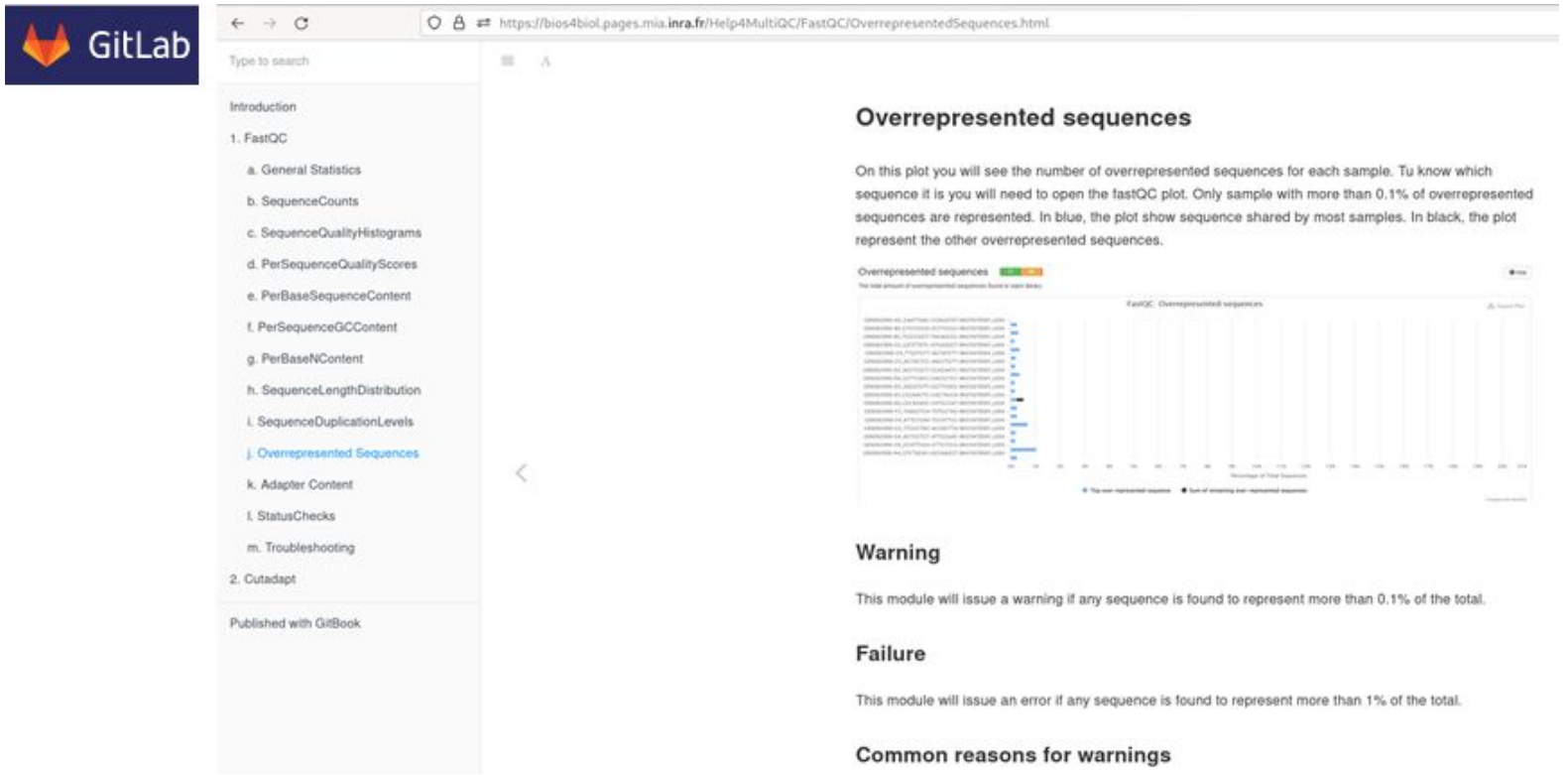
Auteurs : **Paul Tierzan**, Julien Henry et **Elise Maigné**

Help4MultiQC

→ Rédaction collaborative d'un GitBook d'aide à l'interprétation des résultats compilés par multiQCreport à partir de pipelines NGS standards tels que RNAseq, 16S metaGenomics, …

→Ajout du lien vers l'aide dans multiQCreport.

https://bios4biol.pages.mia.inra.fr/Help4MultiQC



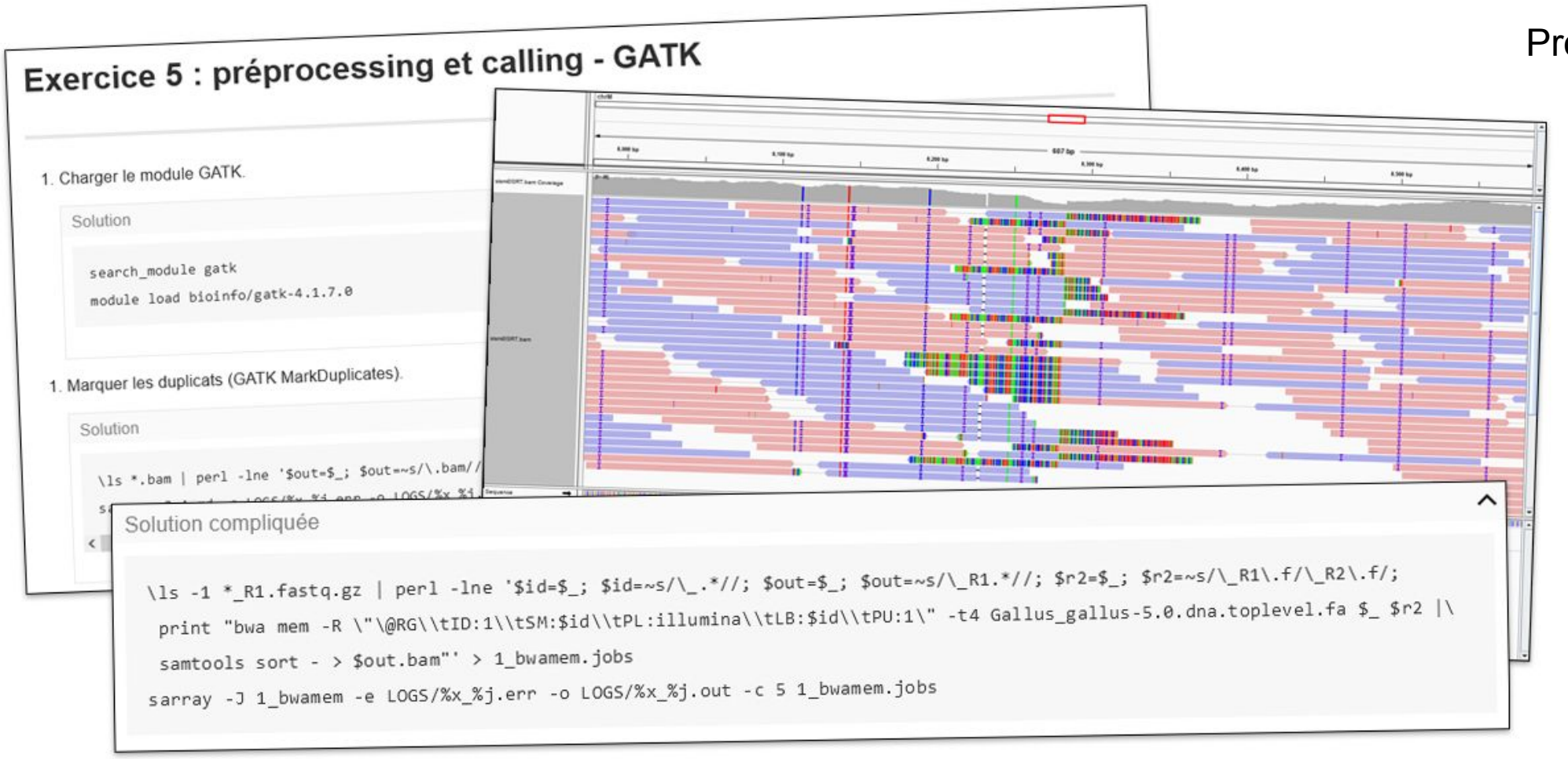
Auteurs : **Cervin Guyomar, Claire Hoede, Yannick Lippi, Sarah Maman.**

Le groupe FROGS s’agrandit

- Lucas Auer a rejoint « officiellement le groupe de travail FROGS ».
  support et support-utilisateurs en statistiques, formation en binôme (3 webinaires en 2022), développement du site web et production de pages de tutoriels des outils.
- Perspectives de développements de nouvelles méthodes à intégrer dans FROGS

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

Formation alignement de lectures courtes et recherche de variants de petite taille



Programme sur 2 jours :

- qualité fastQC
- aln BWA
- manipulation SAM/BAM
- visualisation IGV
- calling GATK
- formats VCF/gVCF
- annotation SNPsift/SNPeff
- pipeline nf-core/sarek

Auteurs : **Céline Noirot**, **Philippe Bardou**, **Cédric Cabau**

Participation à l'hackathon bioinformatique inter-CATI omiques

- du 5 au 7 octobre 2021 à Sète
- 5 CATIs : BARIC, BIOS4Biol, BOOM, GREP, PlantBreed
- 4 ateliers :

Reproductibilité des workflows

DataScience

Text-mining

Développement d'API

Auteurs : 69 participants dont **21** du CATI Bios4Biol

https://forgemia.inra.fr/inter_cati_omics/hackathon_octobre2021

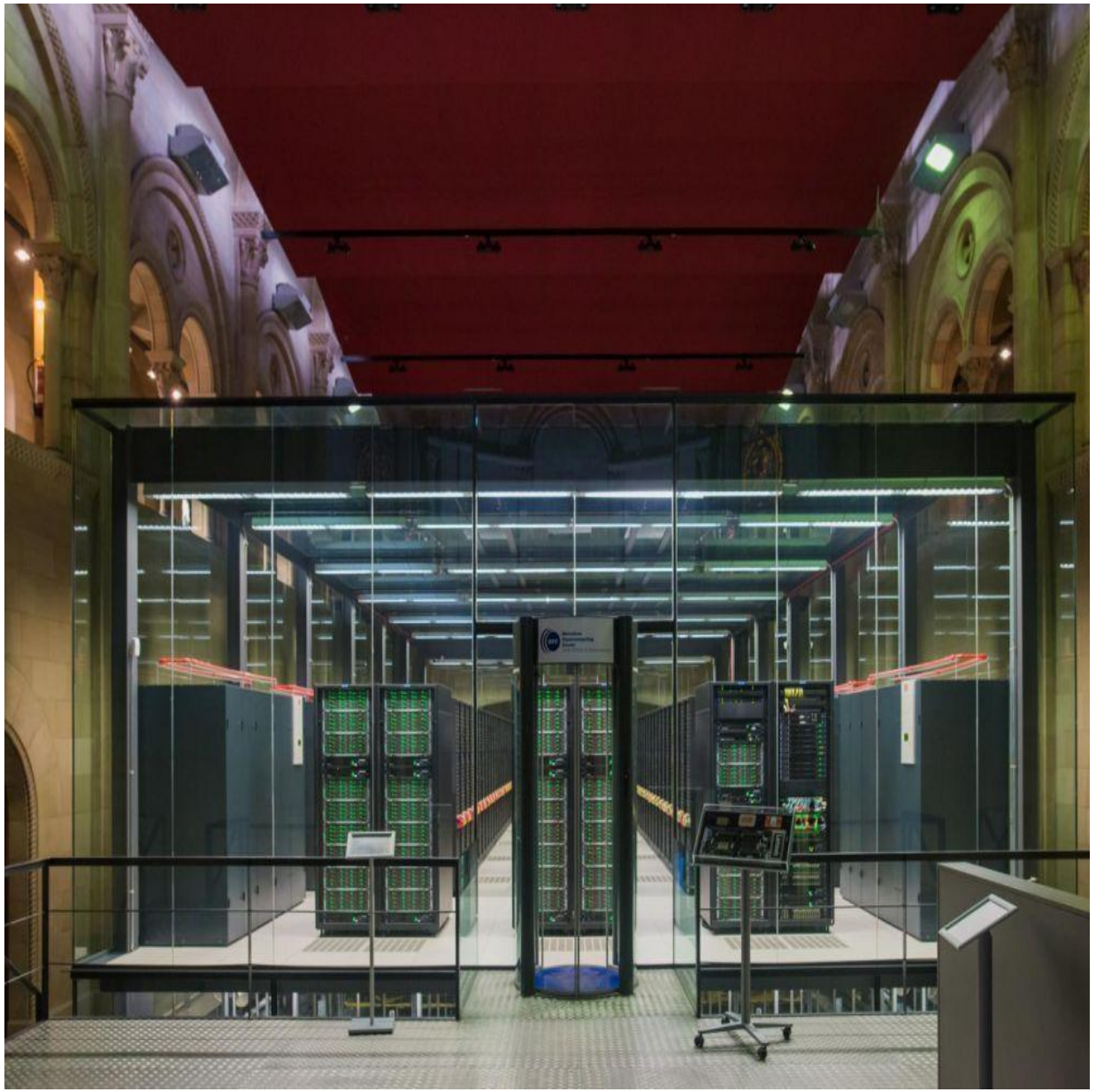
infra/systeme/BD

Acquisition cluster de calcul 2022

- Consultation des fournisseurs info.
- Rédaction du cahier des charges
- Analyse et comparaison des offres
- Signature du marché mi-juin
- Livraison / installation octobre 2022

- Matériel LENOVO
- Intégration AXIANS
- Hébergement DROCC

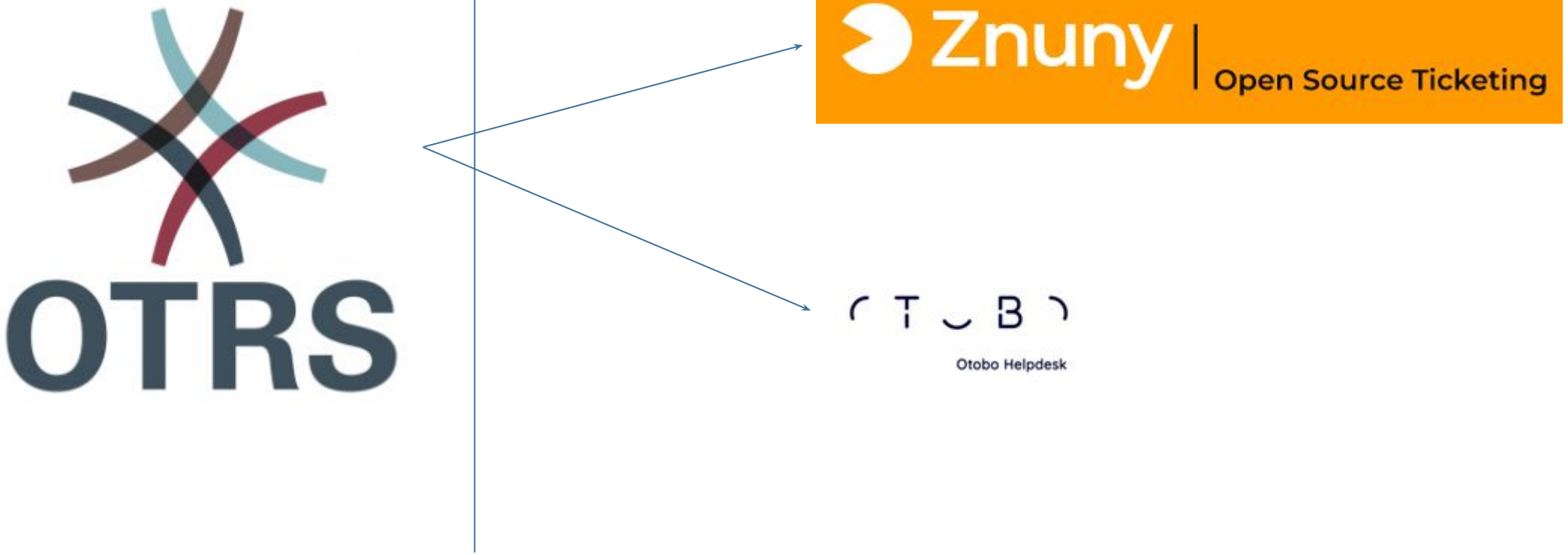
- 5000 cœurs, 16G RAM / cœur
- Stockage Spectrum (/work) 2.5 Po
- Nouveaux services : GPU, OOD



Auteurs : **Marie-Stéphane Trotard, Patrice Dehais, Didier Laborie**

Remplacement OTRS (gestionnaire de tickets)

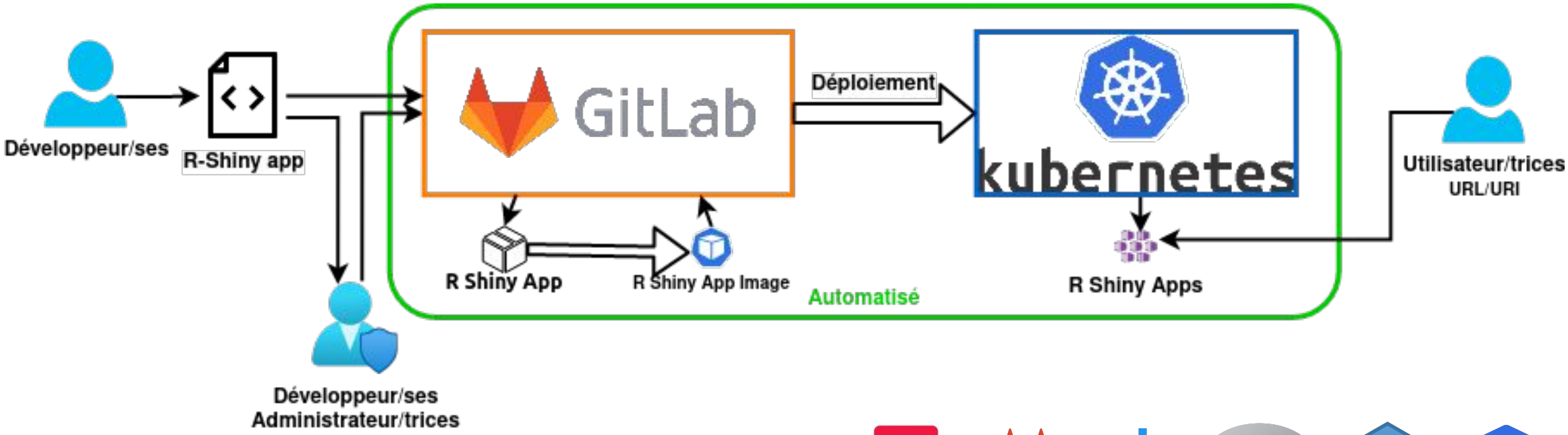
- Outil libre community edition non maintenu depuis plusieurs mois, soumis à licence
- Utilisé par PF Bioinfo/SIGENAE/PlaGe (2 instances)
  - Evaluation / mise en place forks OTRS



Auteurs : **Gérald Salin, Didier Laborie, Marie-Stéphane Trotard**

Lancement du service SK8

Infrastructure pour déployer des application R shiny.
16 demandes d'hébergement depuis ouverture à INRAE en avril 2022 :



Auteurs : ~30 personnes, 9 CATI représentés.
Porté par Jean-François Rey (CATI IMOTEP), pour le CATI Bios4Biol : **Elise Maigné**

Déploiement de l'app matrix sur SK8

Pipeline automatisé dépôt gitlab SK8

https://forgemia.inra.fr/sk8/sk8-apps/get-trix/matrixapp

✓ **Explore** data/results

✓ **Functional analysis**

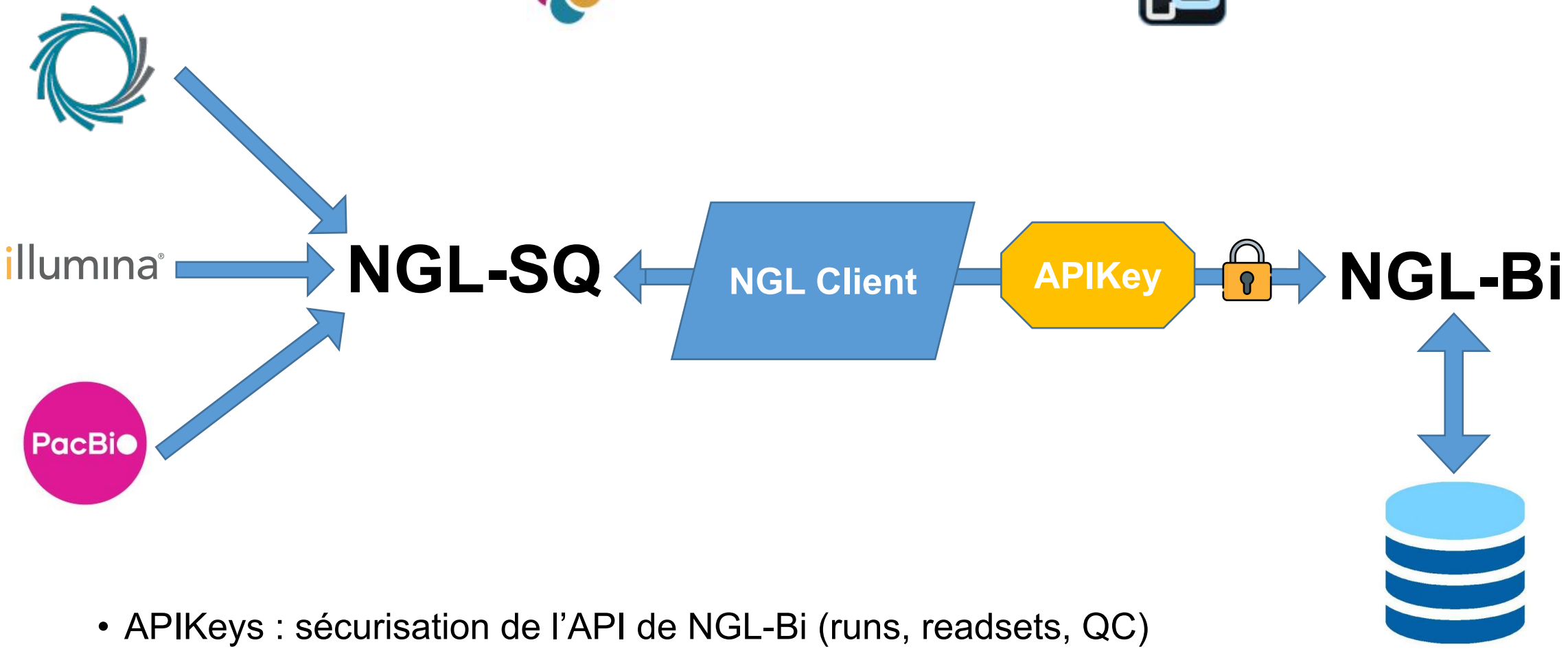
Enrichr

Actions :
- Adaptation du Dockerfile
- Aide debug

Tests à venir :
- Authentification utilisateurs
- Montage volume de données

Auteurs :**Yannick Lippi**, Jean-François Rey (CATI IMOTEP)

Évolution de NGL succession à

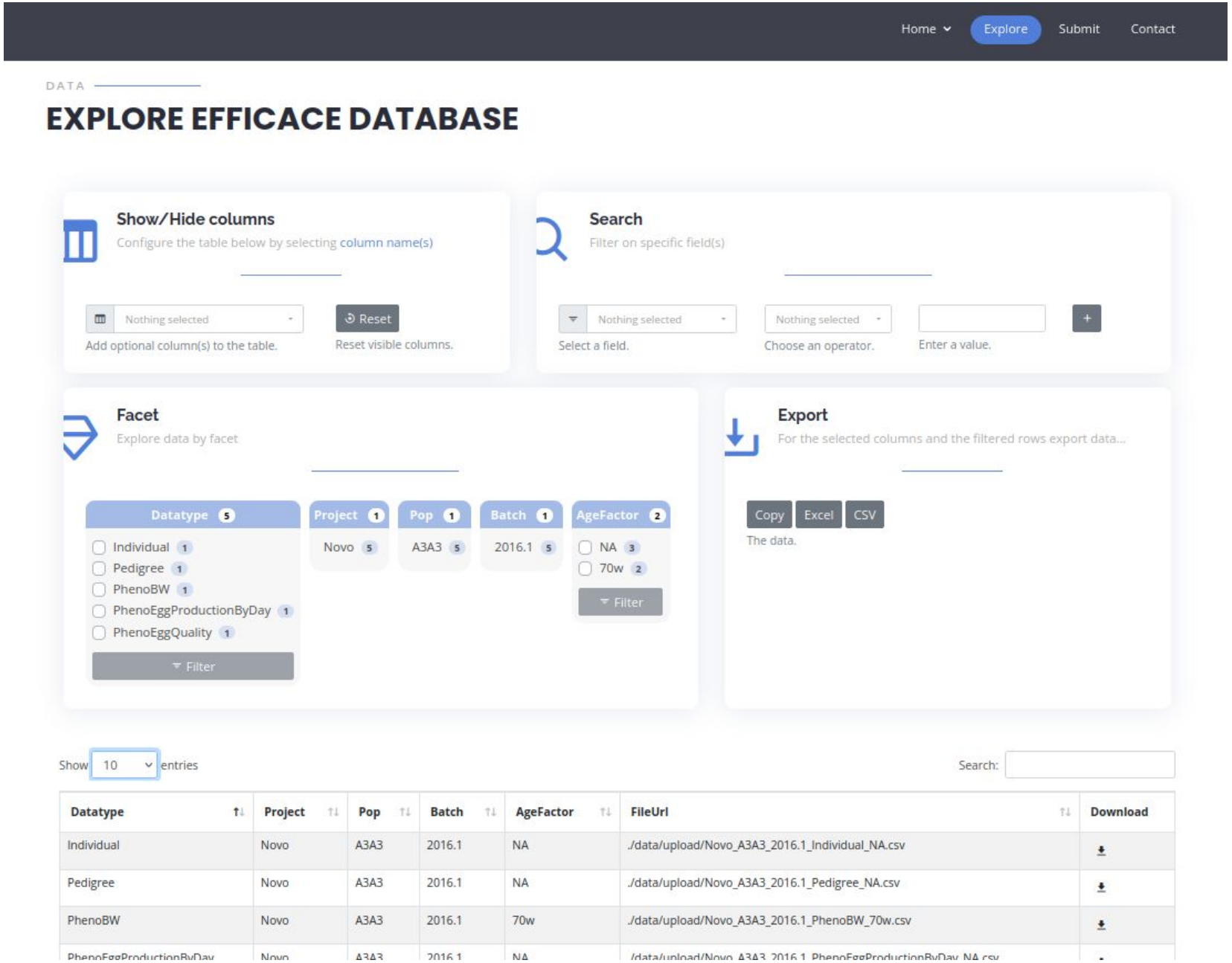


- APIKeys : sécurisation de l'API de NGL-Bi (runs, readsets, QC)
- Intégration de QC : MultiQC, Krona, MinionQC, SMRTLink nextflow

Auteurs : **Eden Darnige, Claire Kuchly, Gerald Salin, Archimede Abdias Towe-Patipe, Jules Sabban, Céline Noirot, Romain Therville**

Une base de données de fichiers

Gérer des fichiers de données hétérogènes (phénotypes, séquences, ...)



Auteurs : **P. Bardou**, **C. Cabau**, **M. Gachet**, **C. Klopp,** Sandrine Lagarrigue, F. Lecerf