



# Bilan 2023

## Effectifs du CATI

- Sorties 2023
  - Pierre Maud (statisticienne Bordeaux),
  - CDDs SeqOcln
- Entrées 2023 :
  - Quentin Boone (CDD Sigenae),
  - Julien Henry (IR GenoToul Bioinfo),
  - Gabryelle Agoutin (CDD GenPhySE),
- 43 membres actifs

## Réunions mensuelles

- 17 Janvier : Jérôme Montfort : “FEVER, a tool to explore expression evolution”
- 03 Février : J.Tourmayre - A.Frambourg : “Point site web du CATI”
- 09 Mars : Sylvain Foissac : “Différence de conformation 3D de la chromatine par comparaison de données Hi-C
- 18 Avril : Michèle Halstead : “Analyses Cut&Tag et Cut&Run”
- 12 Mai : Jérémy Tourmayre : “Microannot et assemblage d’un génome de microsporidie”
- 13 Juin : M.Bergevin (Itrium) : “Datacenter - Technique de refroidissement par immersion”
- 03 Octobre : Jean-Marc Frigerio : “Outil de metabarcoding YapOTU”
- 07 Novembre : Géraldine Pascal : “Metabarcoding : Utilité des longreads dans l'affiliation taxonomique”
- 12 Décembre : C.Noiroit - P.Bardou - P.Bordron - L.Jouneau : “Développements d’applications web”

## Actions inter-CATI

- 4<sup>ème</sup> Hackathon inter-CATI Omiques 2023 qui s’est tenu à Roscoff du 18-21 septembre. 7 participants BIOS4Biol: Vincent Darbot, Claire Hoede, Philippe Bordron, Ludovic Duvaux, Cervin Guyomar, Maria

Bernard. Financement 100% départements à hauteur de 15 500€. Frais de déplacements des agents du CATI pris en charge par le CATI.

- Projet CAPIOT (porteur Jean-Marc Frigerio) : Création et animation d'une communauté d'agents de terrain, électroniciens et informaticiens autour des solutions IoT. Financement SAPI reçu (2000€) de trop faible hauteur et la somme perçue a été rendue. L'achat de l'imprimante 3D n'ayant pu se faire pour le moment, ce projet est en pause.

## Avancement projets CATI

- MultiQC:

Sujet de stage déposé mais pas stagiaire. Une seule réunion en 2023. Réflexions pour redynamiser le projet de la part des animateurs (S. Maman, Y. Lippi, C. Hoede)

- Site web pour le CATI:

Principaux acteurs: Jeremy Tournayre, Anne Frambourg et Marie-Stéphane Trottard.

Site web en ligne: <https://bios4biol.toulouse.inrae.fr/>

Utilisation pour la présentation aux nouveaux arrivants

Difficultés à remplir les fiches agents, peu de retour des membres du CATI.

- **Journée Fête de la Science**

Cette journée s'est tenue le mardi 10 octobre 2023. Nos ateliers portaient sur rôle du microbiote dans la performance des chevaux d'endurance. Trois ateliers ont été construits, ils étaient dédiés aux collégien.ne.s :

1. Assemblage de séquences
2. Initiation à la programmation
3. Introduction à la biostatistique

On a produit des supports partageables et réutilisables. On a reçu 2 classes de 3ème.

Ont participé à ces journées: Gabryelle Agoutin, Lucas Auer, Cédric Cabau, Vincent Darbot, Cervin Guyomar, Luc Jouneau, Géraldine Pascal, Didier Laborie, Elise Maigné, Sabrina Rodriguez

<https://www.fetedelascience.fr/decryptage-desbacteries-de-l-intestin-des-chevaux-de-courses-les-metiers-de-la-bioinformatique-et>

## Organisation des journées annuelles 25-27 septembre 2023 (en présentiel à Toulouse)

- Lundi:
  - Accueil + nouveaux arrivants
  - ouverture à de nouveaux pans scientifiques et techniques : Laure Luno CNES - « Le programme Cospas-Sarsat (balises de détresses) et sa mise en œuvre en France »
  - ChatGPT présentation par P. Bardou
  - Présentation des ateliers C. Guyomar « Fête de la science »
- Mardi:
  - Journée à MétéoFrance
  - Sarah Berthet météoFrance: Modéliser la biogéochimie marine dans un modèle de climat avec PISCES
  - Pierre Amato CNRS - Microbiologie de l'atmosphère – Clermont Ferrand: Bioaérosols: des problématiques aux interfaces disciplinaires - Apports de la bioinformatique et de la météorologie à la compréhension de l'aéromicrobiome
  - Marc Pontaud MétéoFrance: la prévision numérique du temps : des fondements à l'IA
  - Visite du site de MétéoFrance : centre calcul ; centre de prévision ; parc instrumenté
  - restaurant
- Mercredi matin:
  - bilan des productions de l'année avec les responsables scientifiques du CATI
  - organisation de deux ateliers : IA et rédaction d'une revue dans Kids Frontiers
  - restaurant

31 agents ont participé aux journées.

## Bilan des productions CATI 2023

Les productions du CATI se définissent par des productions qui émanent de discussions/idées/reliations d'agents au sein du CATI et qui sont des projets menés par plusieurs membres du CATI (une production CATI peut être une production issue d'une plateforme)

N.B : Un document en annexe ("Une image - Un projet - AG CATI BIOS4Biol 2023.pdf") détaille les différentes productions.

### **Pipelines, développements, benchmarking**

D-Genies: Piste de visualisation supplémentaire sur D-Genies en batch mode. (Philippe Bordron, Christophe Klopp, Bioinfo Genotoul)

metagWGS : co-assemblage et assemblage de tous les échantillons ou par lot + amélioration de l'annotation des gènes. Création et intégration de l'outil Binette permettant le merging de bin. (M.Vienne, V.Darbot, J.Mainguay, C.Noiro, G.Pascal, C.Hoede)

Pipeline RNASeq: 3 nouvelles méthodes de GSEA (enrichissement) + formation associées d'une journée (initiation à unix + utilisation du pipeline) (Anne Frambourg, Luc Jouneau)

Efficace Database: Efficace database est une application web de soumission et consultation de fichier respectant des contraintes prédéfinies (Quentin Boone - Philippe Bardou - Cedric Cabau - Christophe Klopp)

FROGS 4.1 : Restructuration d'outils et ajout de nouvelles fonctionnalités - prise en charge longreads intégration de l'inférence fonctionnelle en analyse différentielle - mutation OTU -> ASV (Maria Bernard, Vincent Darbot, Olivier Rué, Lucas Auer, Géraldine Pascal)

YapOTU: librairie python pour construire, visualiser et analyser des OTUs en barcoding et métabarcoding (Jean-Marc Frigerio, Alain Franc)

Ilots CpG et génomes Animaux: pipeline NextFlow (Aurélien Brionne)

### **Infrastructure, Système et Base de données**

Passage de Genologin à GenoBioinfo (MS.Trotard, P.Deiais, D.Laborie)

Progression du projet SK8 (serveur Shiny) : ~54 demandes d'hébergement depuis l'ouverture du service ( ~20 ITA INRAE, 10 CATI représentés. Porté par Jean-François Rey (CATI IMOTEP), pour le CATI Bios4Biol : Elise Maigné, Yannick Lippi)

1000G chicken: recherche de variants - adaptation pour un cluster de calcul non adapté à la bioinfo (Philippe Bardou - Mathieu Charles - Christophe Klopp)

Projet de GetPlage: Passage de toutes les données de Get-Plage vers GenoBioinfo et passage de NG6 à NGL-BI Evolution LIMS NGL et de JFlow en NextFlow (E.Darnige, C.Kuchly, G.Salin, J.Sabban, C.Noiro, R.Therville)

### **Traitements de données**

Assemblage & annotation de génomes de "poissons " Production de 7 nouveaux génomes d'Elopomorphes <https://dx.doi.org/10.1101/2022.04.07.487469> (Jérôme Montfort, Cédric Cabau, Christophe Klopp)

### **Formations**

Formations FROGS: 8 pages de tutoriels <http://frogs.toulouse.inrae.fr/html/tuto.html> + 2 formations de 4 jours (avril et octobre 2023) (L. Auer, V. Darbot, G. Agoutin, G. Pascal)

Alignement de lectures courtes et recherche de variants de petite taille ( P. Bardou, C. Cabau)

Initiation à l'uniligne perl (P. Bardou, C. Cabau)

## Présentations du CATI à d'autres instances

- Présentation du CATI à la nouvelle cheffe de département adjointe PHASE Florence Gondret

## Préparation des CATI4G

- Avril/mai : sondage auprès des agents du CATI
- Mai: retour auprès des responsables scientifiques du CATI et auprès des membres du CATI
- Juin: séminaire à Sètes avec la DIPSO
- Juillet: conclusion par la DIPSO du séminaire de Sète
- Septembre: Jeudi 7 : réunion Départements et DIPSO - Mardi 12 : visio avec notre tutelle MathNum et GA
- Décembre: réunion départements, DIPSO, resp CATI.

documents et calendriers : <https://pepi2g.wiki.inrae.fr/doku.php/communaute:catis:catis4g>

## Perspectives

- Préparations des **journées annuelles 2024**, et du hackathon 2024.
- Soumission demande de soutien aux projets à la DIPSO - **SAPI 2024**
  - Projet PANANNOT (porteur Ludovic Duvaux - Tester et implémenter des solutions pour produire, décrire et manipuler des graphes de pangénomomes et les annotations associées) -> reçu : 5 500 € en 2024 et 11 500 € en 2025 pour les trois workshops
  - Projet SK8 (porteur Elise MAIGNÉ - service d'hébergement d'applications R-Shiny) -> reçu : 3 100 € pour 2024 et 4 100 € pour 2025
  - 5<sup>ème</sup> Hackathon inter-CATIs Omiques 2024 (cati Omiques) Ateliers: inférence de réseaux en métagénomique, Interopérabilité entre graphes RDF et graphes de Neo4j et workflow de graphes de pangénome -> accepté - financement reçu de 20 000€ - complément de budget par les 5 CATIs
- Kids Frontiers : A la suite des visios mensuelles, certains membres du CATI (Anne Frambourg, Elise Maigne, Luc Jouneau) ont poursuivi la visio-conférence (3 réunions) pour discuter du projet de rédaction d'un article pour le journal Kid Frontiers. Le but est de présenter de manière

synthétique et pédagogique ce qu'est la bio-informatique et en quoi elle est devenue nécessaire dans de multiples domaines de recherche de la biologie. Anne Frambourg a fait une proposition de paragraphe pour l'historique et Luc Jouneau pour présenter ce qu'est une analyse différentielle de transcriptomique. On devrait discuter de ce début de contenu lors de notre prochain échange

- Construction du **mini-symposium "l'IA, impact sur nos métiers"** qui se tiendra lors du congrès JOBIM 2024 <https://jobim2024.sciencesconf.org/resource/page/id/6>. (Cedric Cabau , Cervin Guyomar, Sarah Maman Haddad , Gabryelle Agoutin, Patrice Dehais, Lucas Auer, Philippe Bardou, Géraldine Pascal)

Objectifs du symposium :

Ce symposium vise à créer un espace de discussion stimulant, où les participants pourront explorer les façons dont l'IA transforme leurs pratiques et les recherches à venir. Nous aimerions discuter des opportunités émergentes, des défis et des implications éthiques liées à l'IA en biologie et en bioinformatique.

Programme/objectifs :

Mot d'ouverture : Introduction aux enjeux actuels de la bioinformatique et de l'IA.

Conférences : Interventions de Vincent Guigé et Guillaume Cabanac. Table ronde :

Discussions interactives sur les implications, les défis et les opportunités. Réseautage :

Occasion d'échanger des idées et de nouer des collaborations.

- **Réunions mensuelles 2024:**
  - 14 Janvier : Hugo Chauvet et Christophe Chevalier - Présentation du synchrotron Soleil
  - 6 Février : Aurélien Brionne - Ilôts CpG et génomes d'animaux
  - 5 Mars : Olivier Ridoux (IRISA) - Numérique et sobriété
- A venir
  - 2 Avril : Jérémy Tournayre - Présentation de la manipulation de graphe avec cytoscape
  - 7 Mai : Laurent Bréhélin - TFScope: Une approche de Machine Learning pour l'analyse des caractéristiques génomiques impliquées dans les préférences de liaison des facteurs de transcription
  - 4 Juin : Arnaud Liehrmann et Guillem Rigau - Package DiffSegR