



**HAL**  
open science

## Des solutions informatiques innovantes pour l'utilisation et le développement de biomarqueurs de diversité du microbiote.

Maria Bernard, Olivier Rué, Malo Le Boulch, Jean Mainguy, Laurent Cauquil, Claire Hoede, Sylvie Combes, Géraldine Pascal

### ► To cite this version:

Maria Bernard, Olivier Rué, Malo Le Boulch, Jean Mainguy, Laurent Cauquil, et al.. Des solutions informatiques innovantes pour l'utilisation et le développement de biomarqueurs de diversité du microbiote.. DefisPHASE 2019, 2019, Rennes, France. hal-03851900

**HAL Id: hal-03851900**

**<https://hal.science/hal-03851900>**

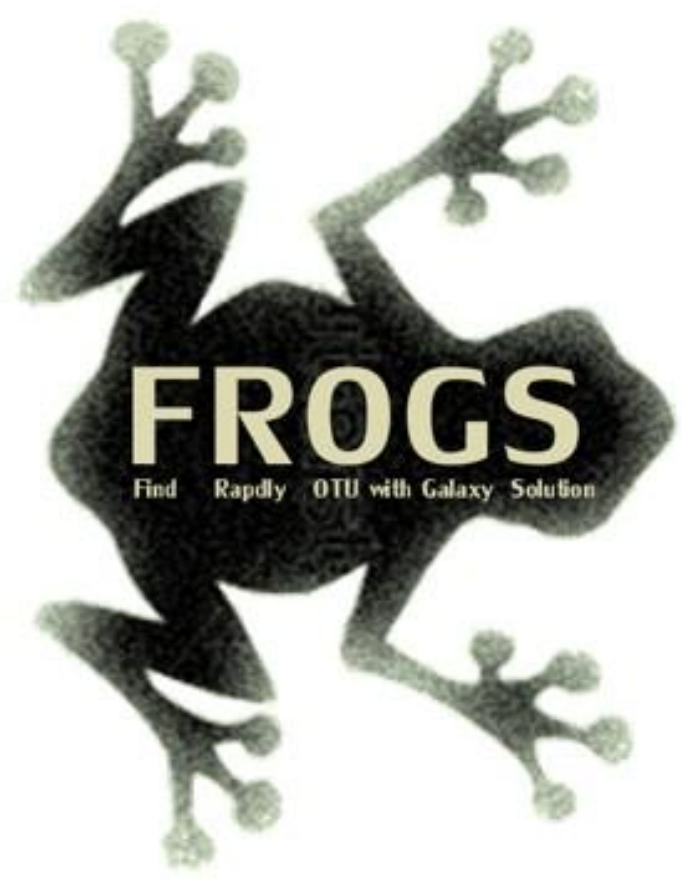
Submitted on 14 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



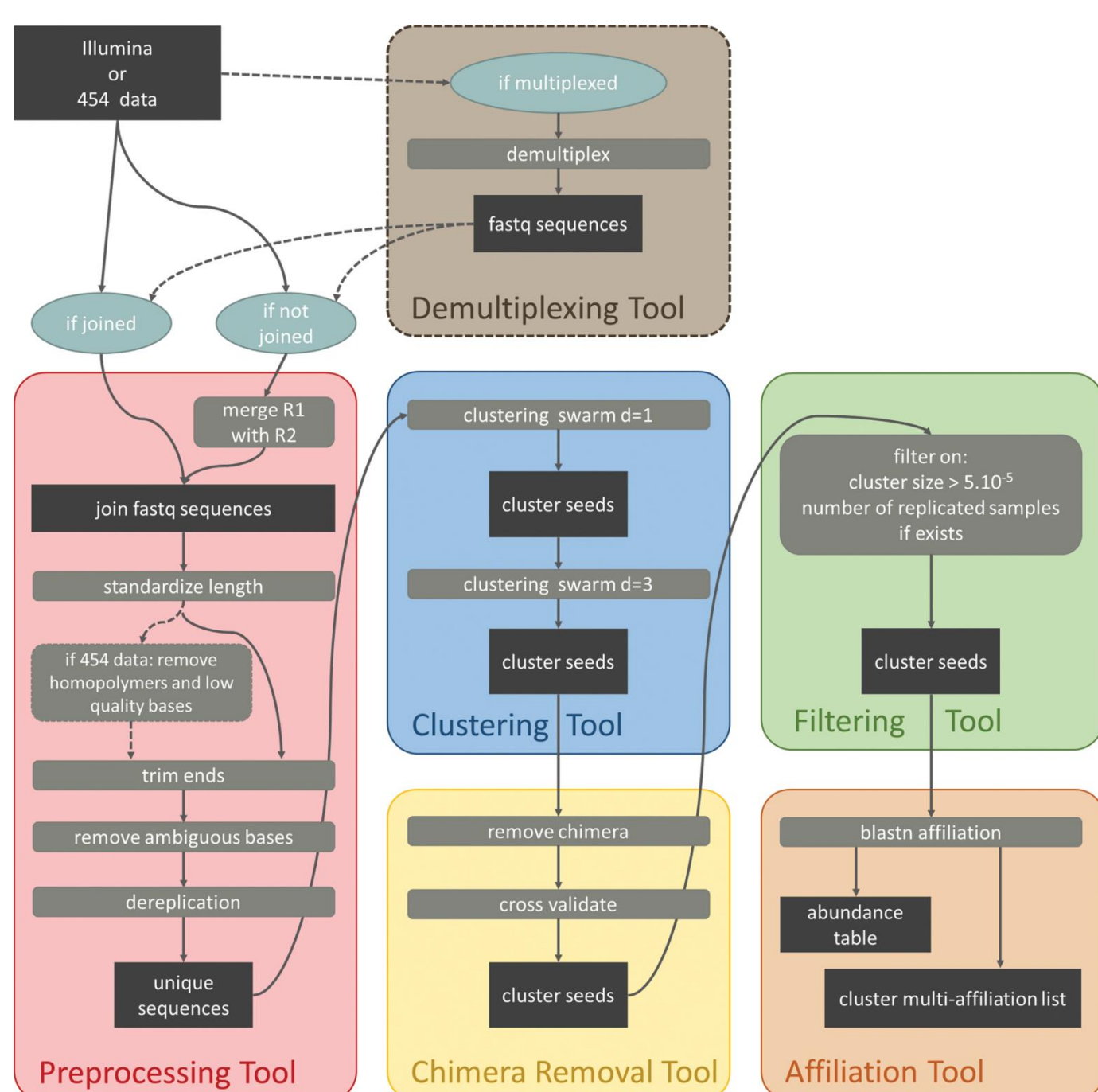
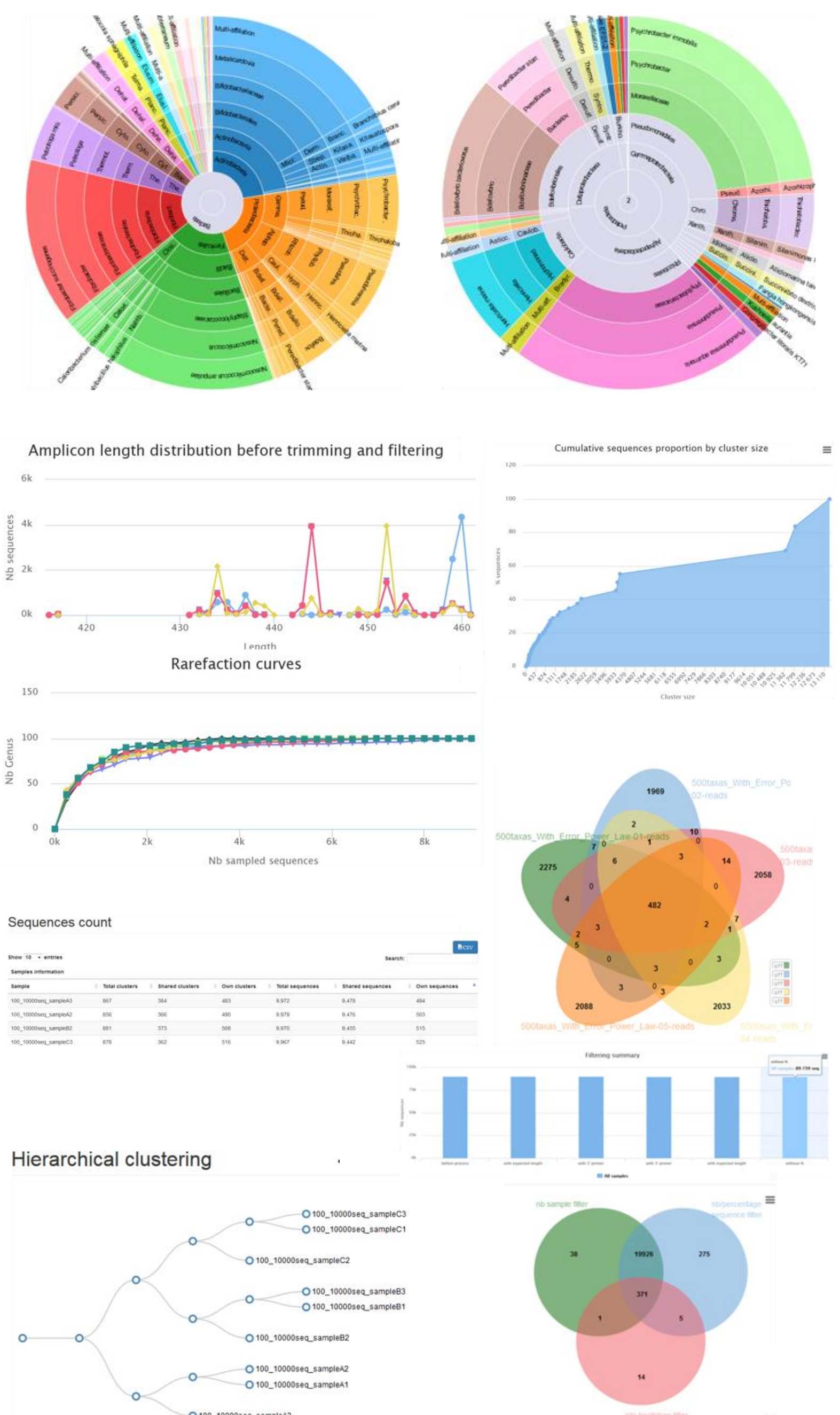
MARIA BERNARD <sup>(1)</sup>, OLIVIER RUÉ <sup>(2)</sup>, MALO LE BOULCH <sup>(3)</sup>, JEAN MAINGUY <sup>(4)</sup>, LAURENT CAUQUIL <sup>(3)</sup>, CLAIRE HOEDE <sup>(4)</sup>, SYLVIE COMBES <sup>(3)</sup> ET GÉRALDINE PASCAL <sup>(3)</sup>



## FROGS

Find, Rapidly, OTUs with Galaxy Solution  
Logiciel d'analyse des données du métabarcoding

- Analyse de grands ensembles de séquences d'amplicons d'ADN
- Précision et rapidité
- Table d'abondance des OTUs (Operational Taxonomic Units) et affiliations taxonomiques
- Interface graphique sous Galaxy (user friendly), ou ligne de commande
- 22 outils de bioinformatique et de statistiques
- Traitement des séquences ITS (levures et champignons)
- Gestion des paires de lectures non chevauchantes (ITS, RPOB, LSU ou tout autre marqueur de taille > 600 bases)
- Plébiscité par la communauté:
  - ✓ + de 2300 téléchargements de FROGS [anaconda.org/bioconda/frogs](https://anaconda.org/bioconda/frogs)
  - ✓ 100 citations de l'article Escudie *et al.* en 18 mois
  - ✓ Au moins 13 plateformes Galaxy dans le monde proposent FROGS
  - ✓ + de 15000 visites sur le [frogs.toulouse.inra.fr](https://frogs.toulouse.inra.fr)

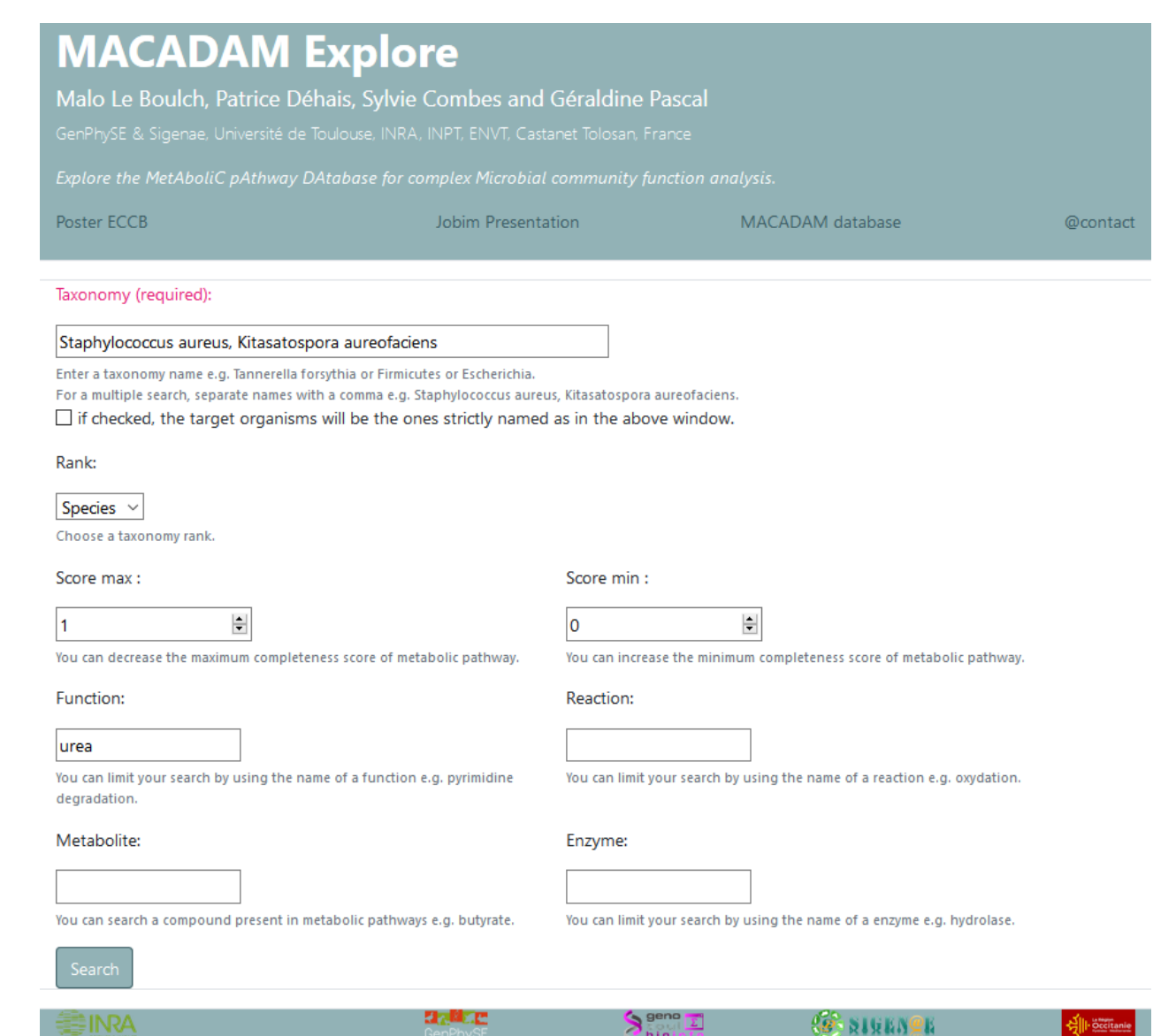
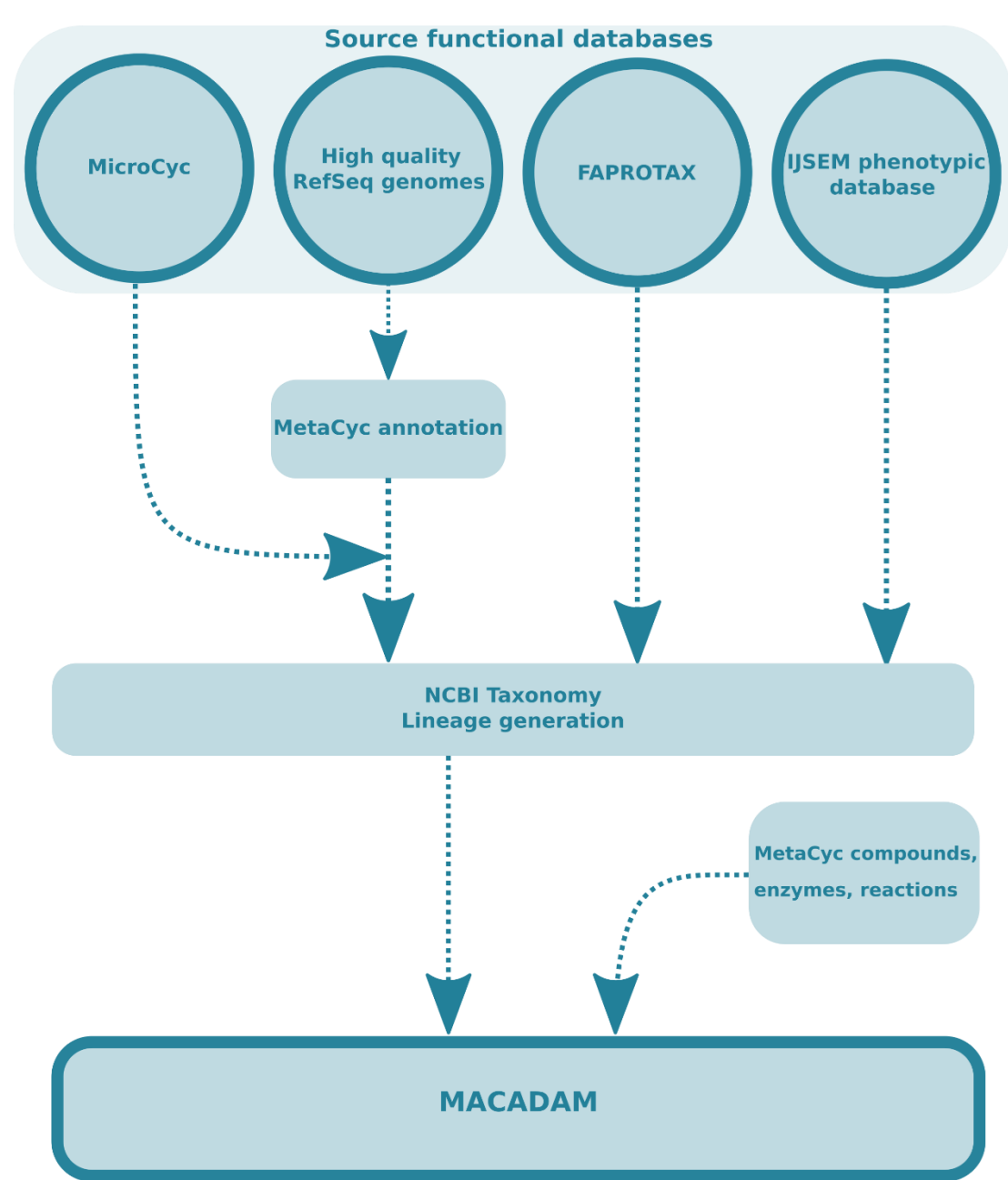


## MACADAM

MetAboliC pAthways DATabase for Microbial taxonomic groups

Une base de données pour l'exploration des fonctions métaboliques des microbiotes à partir de la taxonomie

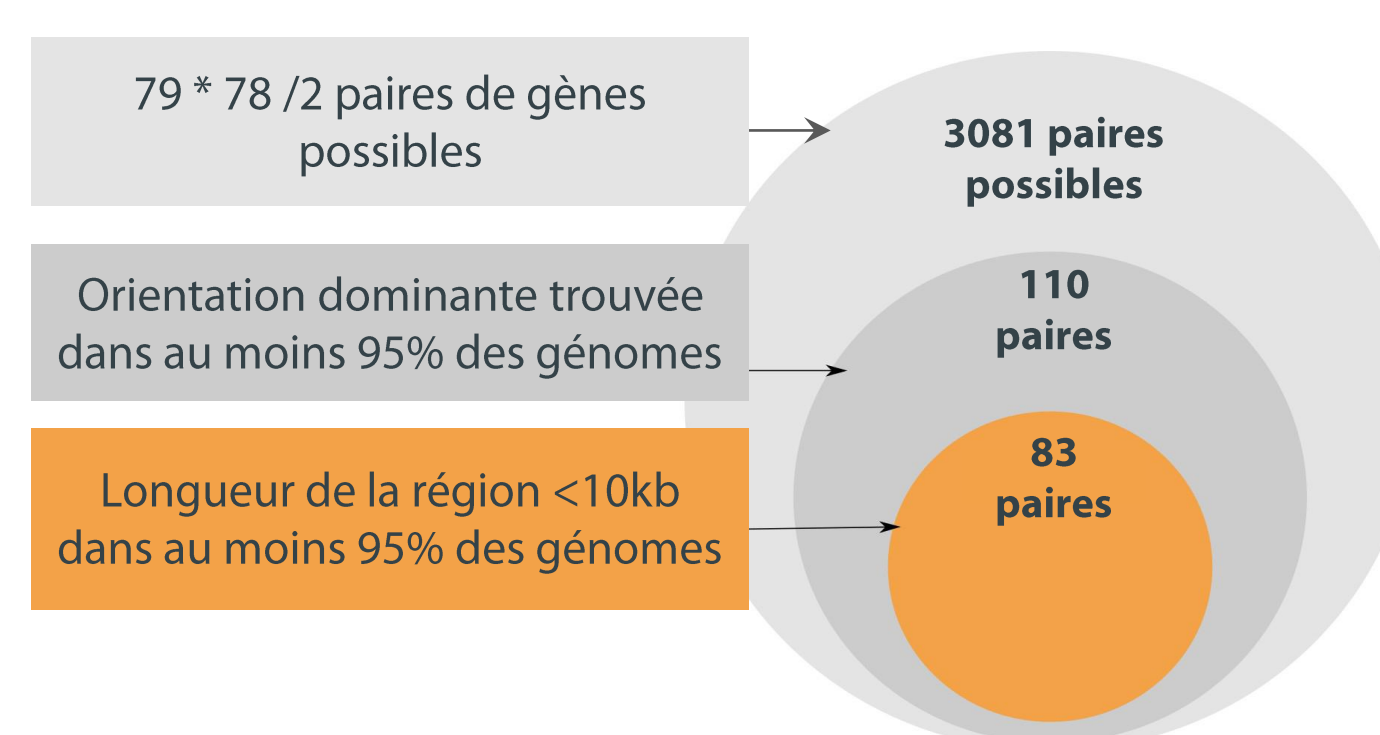
MACADAMExplore - application informatique d'exploration de MACADAM - permet de relier les affiliations taxonomiques des microbes à leurs annotations fonctionnelles – [macadam.toulouse.inra.fr](https://macadam.toulouse.inra.fr)



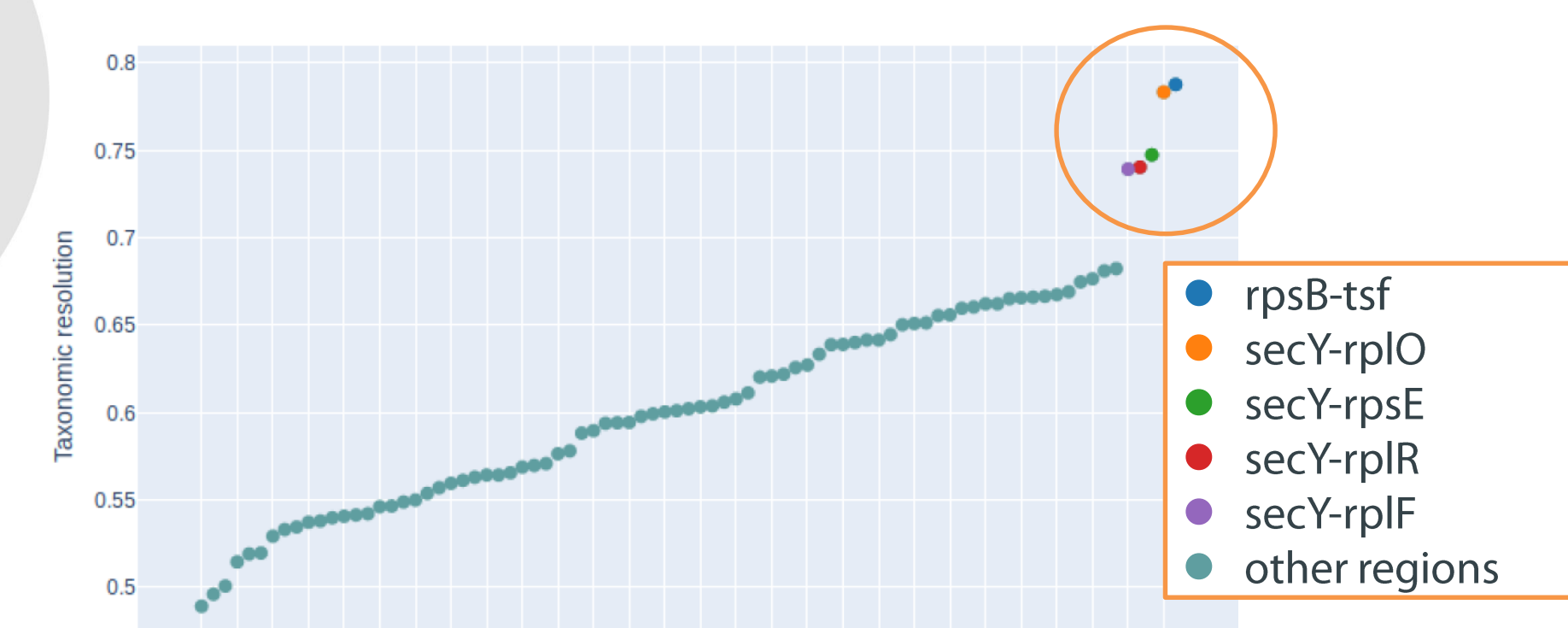
## Les longues lectures de séquençage, l'avenir des marqueurs de biodiversité ?

- Contraintes actuelles du marqueur ARN16S séquençé via Illumina:
  - lectures de 450 nucléotides
  - multi-copies → problème pour la quantification
  - faible résolution taxonomique
- Objectif: Recherche de nouveaux biomarqueurs avec:
  - haute résolution taxonomique au niveau de l'espèce
  - mono copie
  - longues lectures de séquençage *i.e.* fragment de génome de 5000 à 8000 bases pouvant contenir plusieurs gènes

✓ 83 paires de gènes sélectionnées



✓ 5 paires de gènes ont une résolution taxonomique élevée



→ 79 gènes bactériens sont sélectionnés

Critères de sélection des paires de gènes :

- Gènes universels en simple copie
  - Orientation des paires homogène
  - Distances entre les paires
  - Constance de ces distances
- 

Perspectives:

- Exploration/validation de ces 5 régions sur tous les génomes de Refseq *i.e.* ~171 000 génomes.
- Comparaison avec les données sur l'opéron de l'ARN ribosomique.
- Design d'amorces des 5 régions : *In silico* validation des amorces et évaluation de l'impact de ces amorces sur la résolution taxonomique.