



HAL
open science

PANORAMA : comparative pangenomics tools to explore interspecies diversity of microbial genomes

Jérôme Arnoux, Alexandra Calteau, David Vallenet

► To cite this version:

Jérôme Arnoux, Alexandra Calteau, David Vallenet. PANORAMA : comparative pangenomics tools to explore interspecies diversity of microbial genomes. Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM) 2022, Jul 2022, Rennes, France. hal-04371243

HAL Id: hal-04371243

<https://hal.science/hal-04371243>

Submitted on 3 Jan 2024

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.

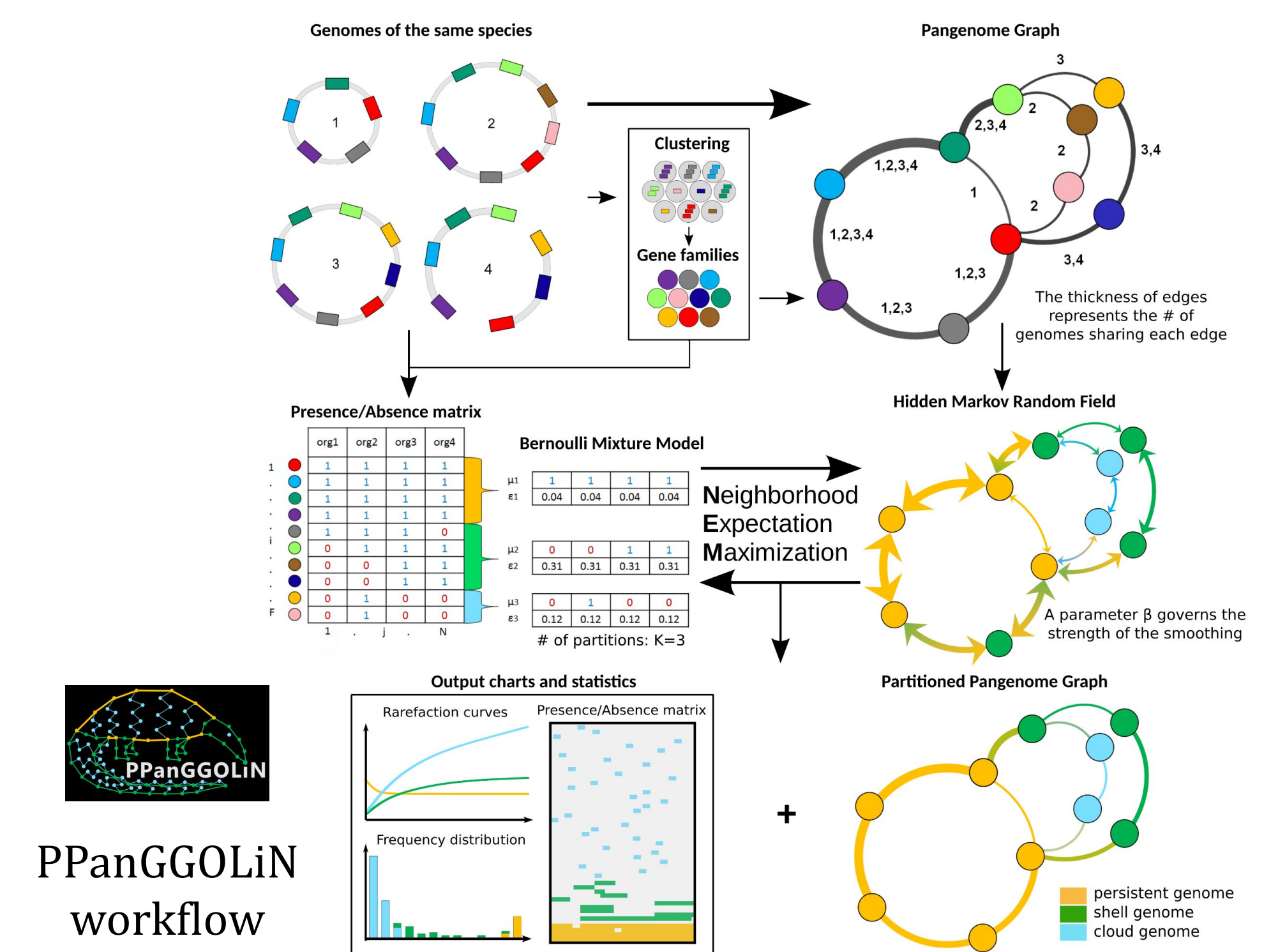


Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Introduction and objectives

With the increase of prokaryotic genomes in public databases, comparative genomics approaches now use hundreds of genomes to analyse species diversity. Many studies focus on the overall species gene content, the **pangenome**, to understand its evolution in terms of common and variable genes with regard to epidemiological or environmental data. In this context, we have been working on genomic data representation as **pangenome graphs**. We have developed methods for **pangenome reconstruction and partitioning (PPanGGOLiN¹)**, **regions of genomic plasticity identification (panRGP²)** and **module detection (panModule³)**.

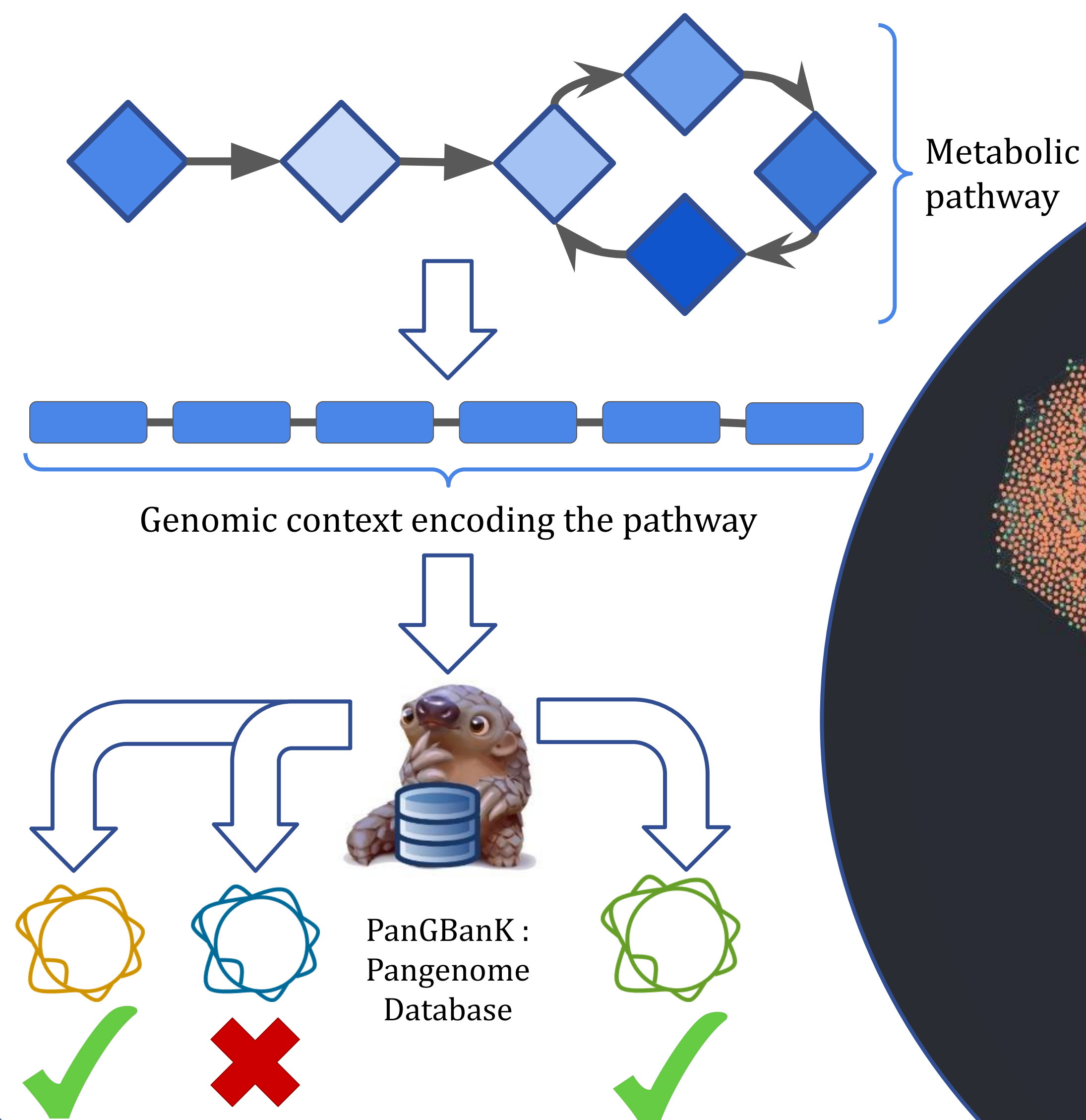
With **PANORAMA**, we will achieve new methodological developments for the **comparative study of pangenomes**. It will help to study the **adaptive potential of bacteria** and to better understand the **evolutionary dynamics** behind the metabolic diversity of microorganisms.



Exploration of genomic context

To study metabolic pathways of interest, we developed a method to search for a **genomic context** (a set of genes or gene families) in pangenomes. This method could be applied to multiple pangenomes to search for similar genomic contexts in a taxonomic group or an environment.

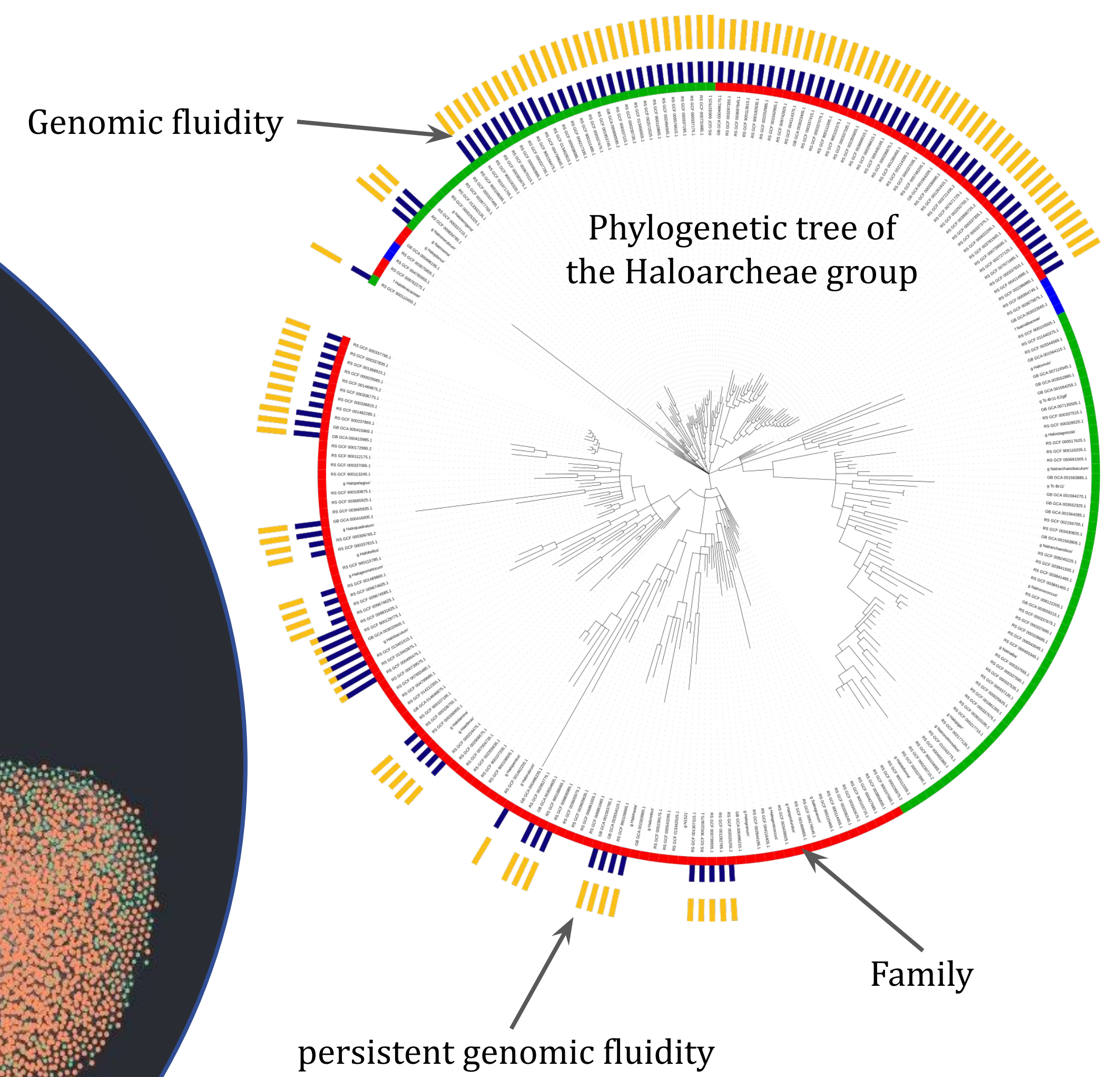
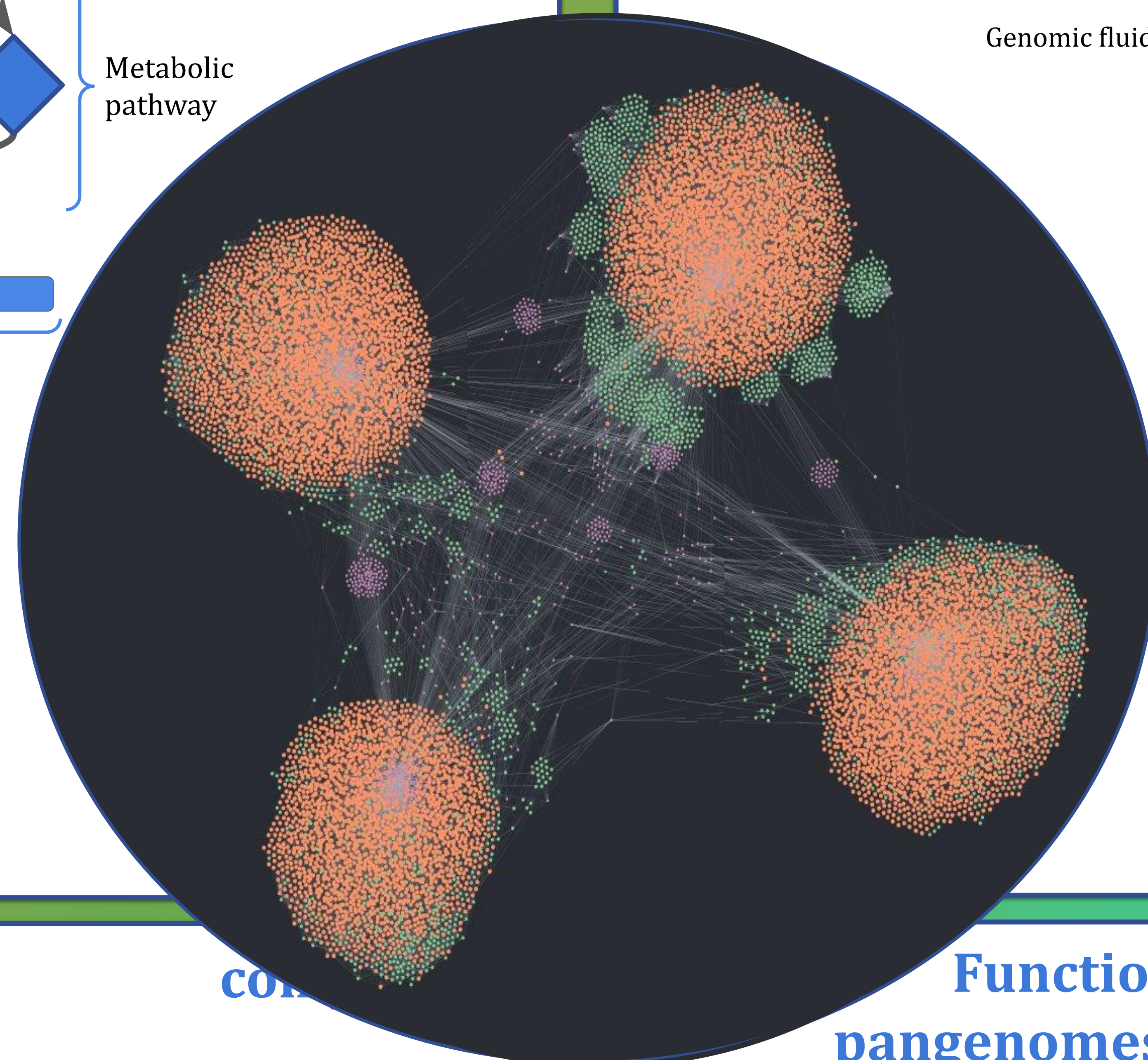
Genomic context detection in multiple pangenomes



General Metrics on the pangenome graph

General metrics are based on **pangenome content**. One of the defined metrics is the **genomic fluidity⁴**, which is a measure of genome diversity assessed at the gene level in a pangenome. To improve pangenome comparison, we are working on genomic fluidity **normalization**.

Another possibility is to use metrics based on graph content in nodes and edges or its topology.



Inter-pangenome comparison with graph database

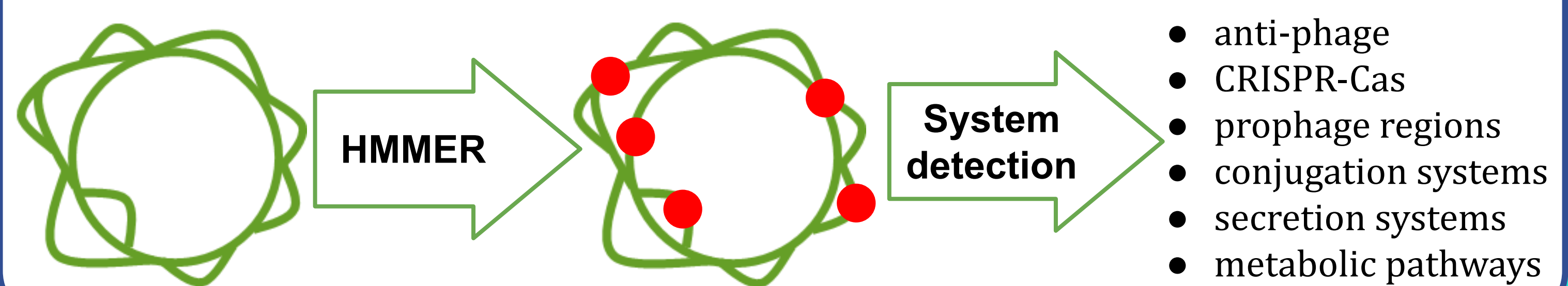
We currently focused our work on **inter-pangenome graph comparison** by identifying **common modules** between pangenomes. In the framework of the **D4Gen Hackathon**, we created our first **pangenome graph database** to make this comparison.

As a proof of concept, we generated a **Neo4J** pangenome graph database of 4 pangenomes from the *Acinetobacter* genus (central image of the poster) connected together by common modules (pink nodes).



Functional annotation of pangenomes families

In order to allow **pangenome comparisons at the functional level**, we are integrating methods to annotate gene families with **HMMs** and **detect biological systems** using rules^{6,7}.



Conclusion & perspectives

PANORAMA will offer rapid and easy to use comparative analysis of pangenomes on several thousands genomes from different species. Thus, it will help to understand the adaptive potential of bacteria and, with the exploration of functional modules in different species, provide a better understanding of the evolutionary dynamics behind the metabolic diversity of microorganisms. Our methods offer a new way of studying microbial communities and can be applied in many fields, such as ecology or health.

All these methods will be integrated in the PANORAMA analysis toolbox, a python software, freely and publicly available on GitHub.

References

1 G. Gautreau et al. PLOS Comp. Biol., 2020 doi: 10.1371/journal.pcbi.1007732
2 A. Bazin et al. Bioinformatics, 2020 doi: 10.1093/bioinformatics/btaa792
3 A. Bazin et al. Biorxiv, 2022 doi: 10.1101/2021.12.06.471380
4 A. O Kislyuk, et al. BMC Genomics 2011doi: 10.1186/1471-2164-12-32

5 B. Paten et al. Journal of Computational Biology 2011 doi: 10.1089/cmb.2010.0252
6 SS Abby et al. PLoS ONE 2014 doi:10.1371/journal.pone.0110726
7 L. J. Payne, et al. Nucleic Acids Research 2021 doi: 10.1093/nar/gkab883