



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

Jérôme Arnoux, Laura Bry, Quentin Fernandez De Grado, David Vallenet,
Alexandra Calteau

► To cite this version:

Jérôme Arnoux, Laura Bry, Quentin Fernandez De Grado, David Vallenet, Alexandra Calteau. PANORAMA: comparative pangenomics tools to explore interspecies diversity of microbial genomes. ISMB/ECCB 2023, Jul 2023, Lyon, France. <hal-04358457>

HAL Id: hal-04358457

<https://hal.science/hal-04358457v1>

Submitted on 21 Dec 2023

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.



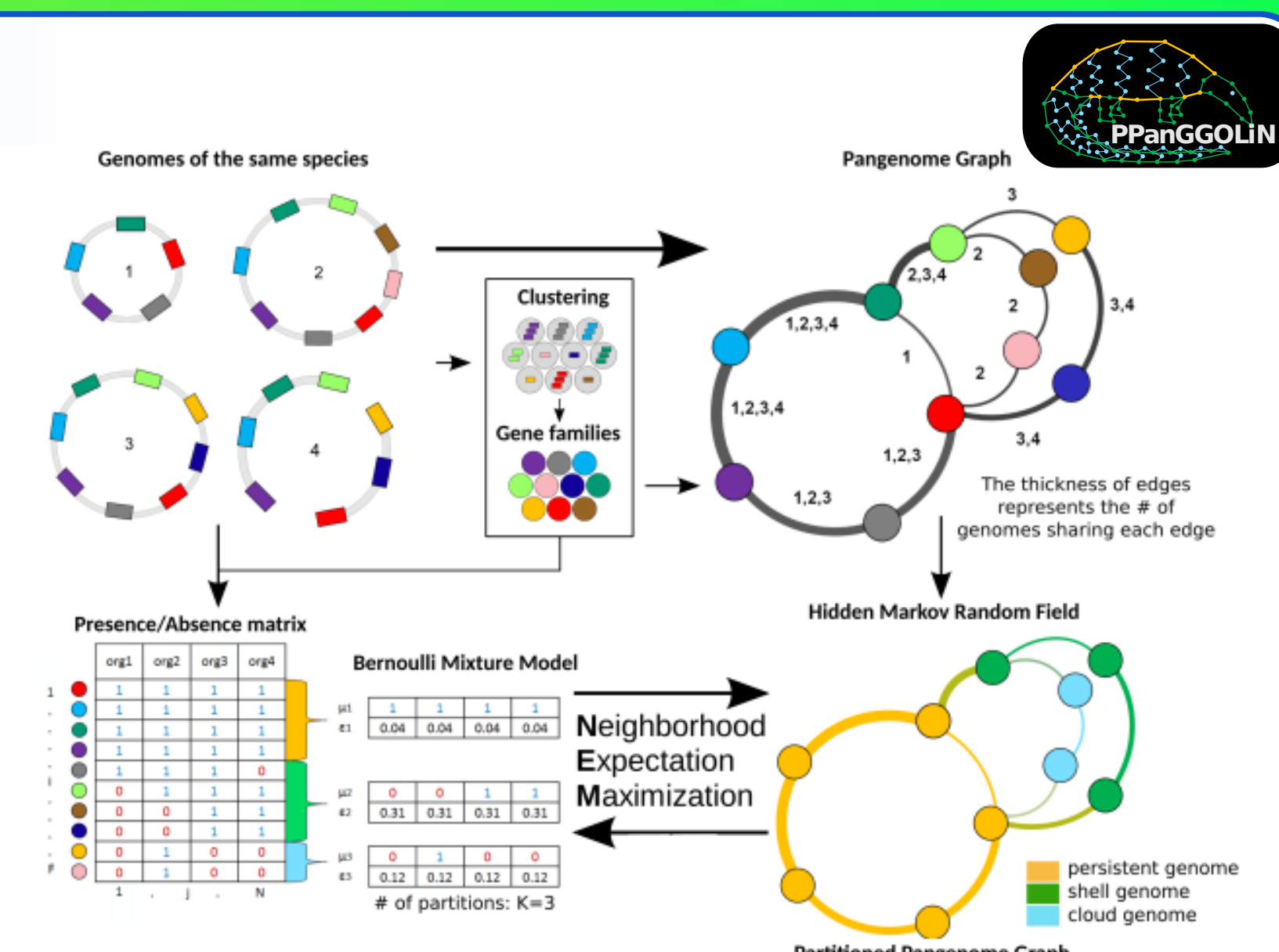
Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

OUTLINE

PANORAMA is an open-source bioinformatics toolbox, including new methodological developments for the comparative study of pangenomes. It benefits from methods for the **reconstruction and analysis of pangenome graphs**, thanks to the **PPanGGOLIN**¹ software suite. PANORAMA integrates multiple features, such as the possibility to **compare genomic context** between pangenomes or the **annotation of biological systems** at the pangenome level.

PANORAMA allows comparative analysis of pangenomes using thousands of genomes. The software is developed in python and can be install with conda. It generates many tables and graphical outputs, some of them being compatible with external bioinformatic tools.

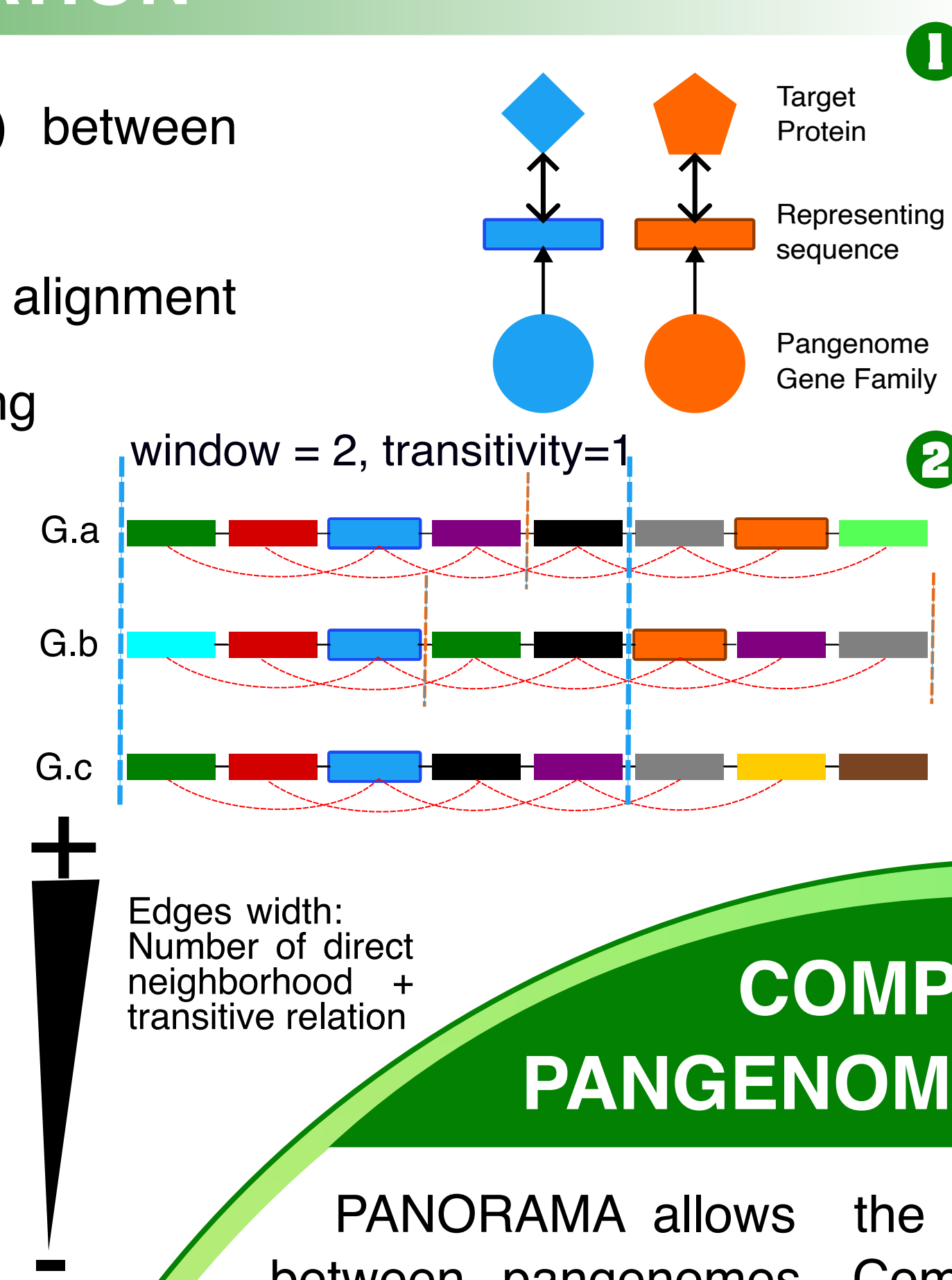
PANORAMA aims to help microbiologists understand the adaptive potential of bacteria and, through the exploration of functional modules in different species, to better understand the evolutionary dynamics behind the metabolic diversity of microorganisms.



GENOMIC CONTEXT EXPLORATION

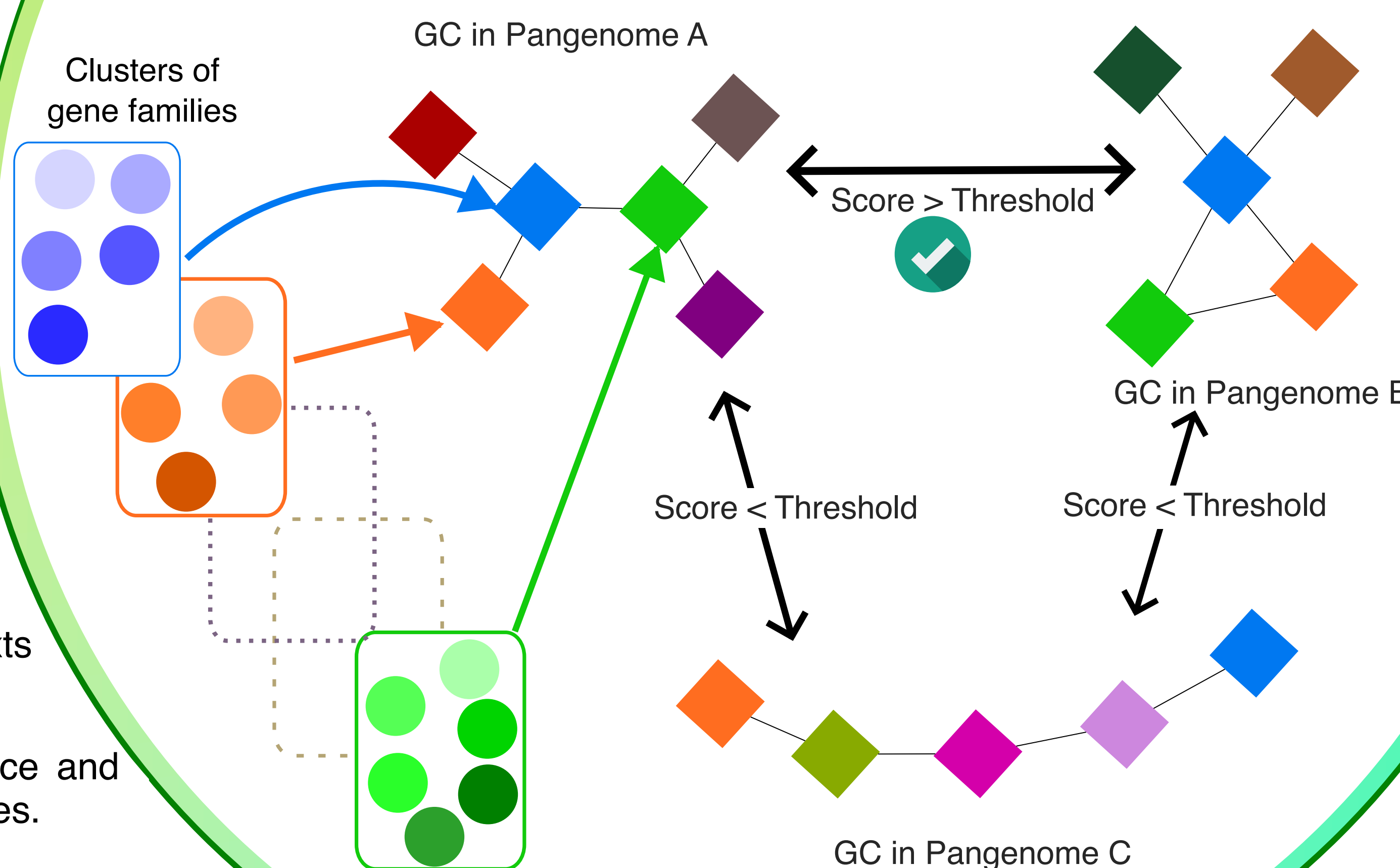
Identify conserved **genomic contexts** (GC) between pangenomes from a set of proteins of interest.

1. Detect homologous families using sequence alignment
2. Search for connected components applying transitive closure and window size parameters
3. GC graph construction.
4. Edges filtering according to a Jaccard index



COMPARATIVE PANGENOMICS ANALYSES

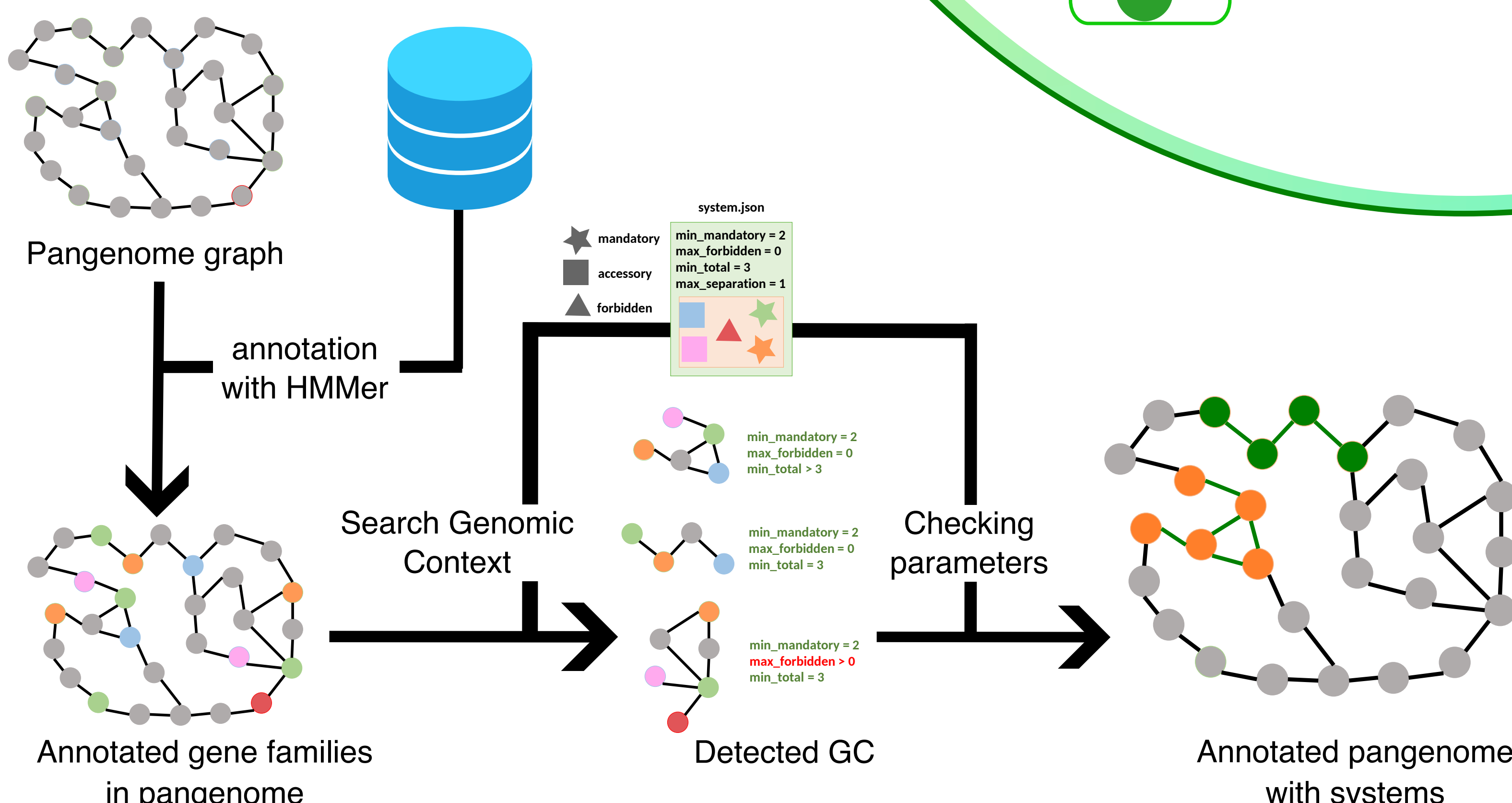
PANORAMA allows the identification of conserved GCs between pangenomes. Comparisons are based on grouping pangenome families and computing common connected components. A conservation score is calculated for each pair of GCs to obtain those that are common and may correspond to pathways or cellular processes conserved between species..



SYSTEM DETECTION

System detection in PANORAMA is based on a functional annotation of gene families, GC searching and system modeling.

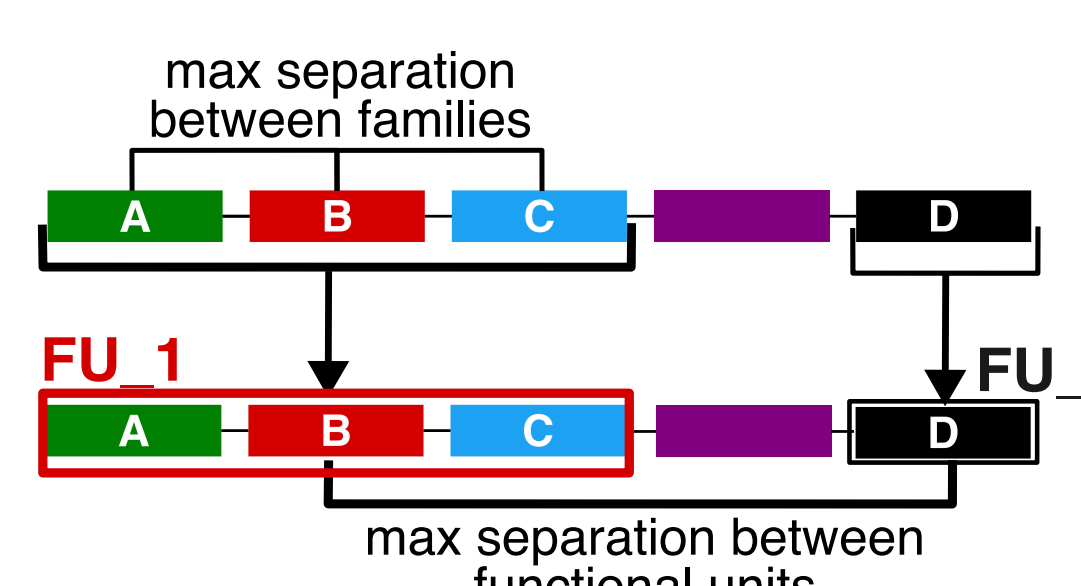
1. Annotation of gene families with HMMs associated to system-models
2. Detection of genomic contexts corresponding to the models
3. PANORAMA verifies that presence/absence and synteny parameters fit with the associated rules.



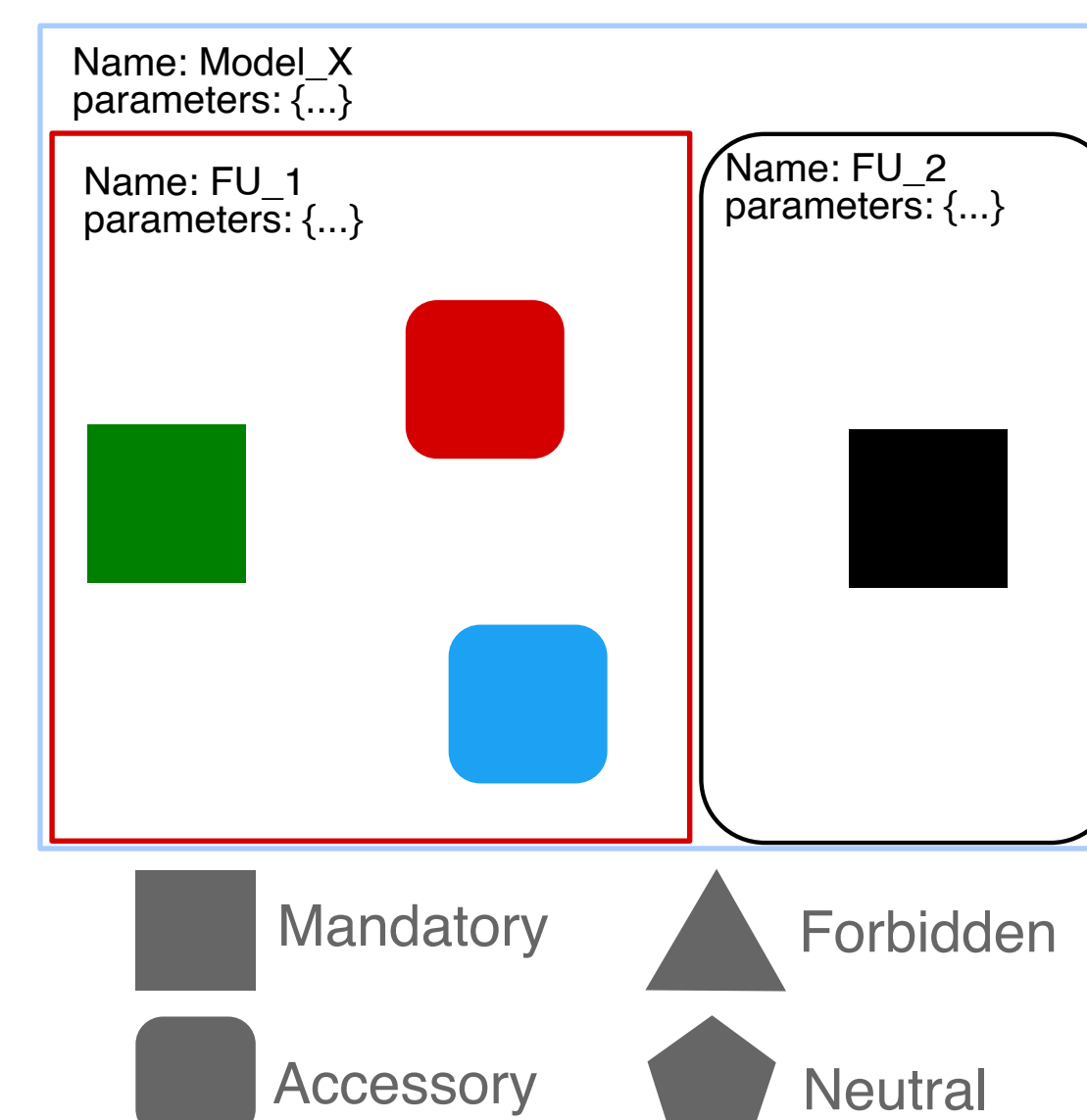
SYSTEM MODELING

PANORAMA models **biological systems** based on rules which describes gene family presence/absence and synteny conservation, similarly to MacSyFinder⁴.

Model representation in a genome



Model representation schema



Model structure in JSON grammar

```
{
  "name": "model_X",
  "parameters": {
    "max_separation": 1,
    "min_mandatory": 1,
    "min_total": 1
  },
  "func_units": [
    {
      "name": "FU_1",
      "type": "mandatory",
      "parameters": {
        "min_total": 2
      }
    },
    {
      "name": "famA",
      "type": "mandatory"
    },
    {
      "name": "famB",
      "type": "accessory"
    },
    {
      "name": "famC",
      "type": "accessory",
      "parameters": {
        "max_separation": 2
      }
    }
  ]
}
```

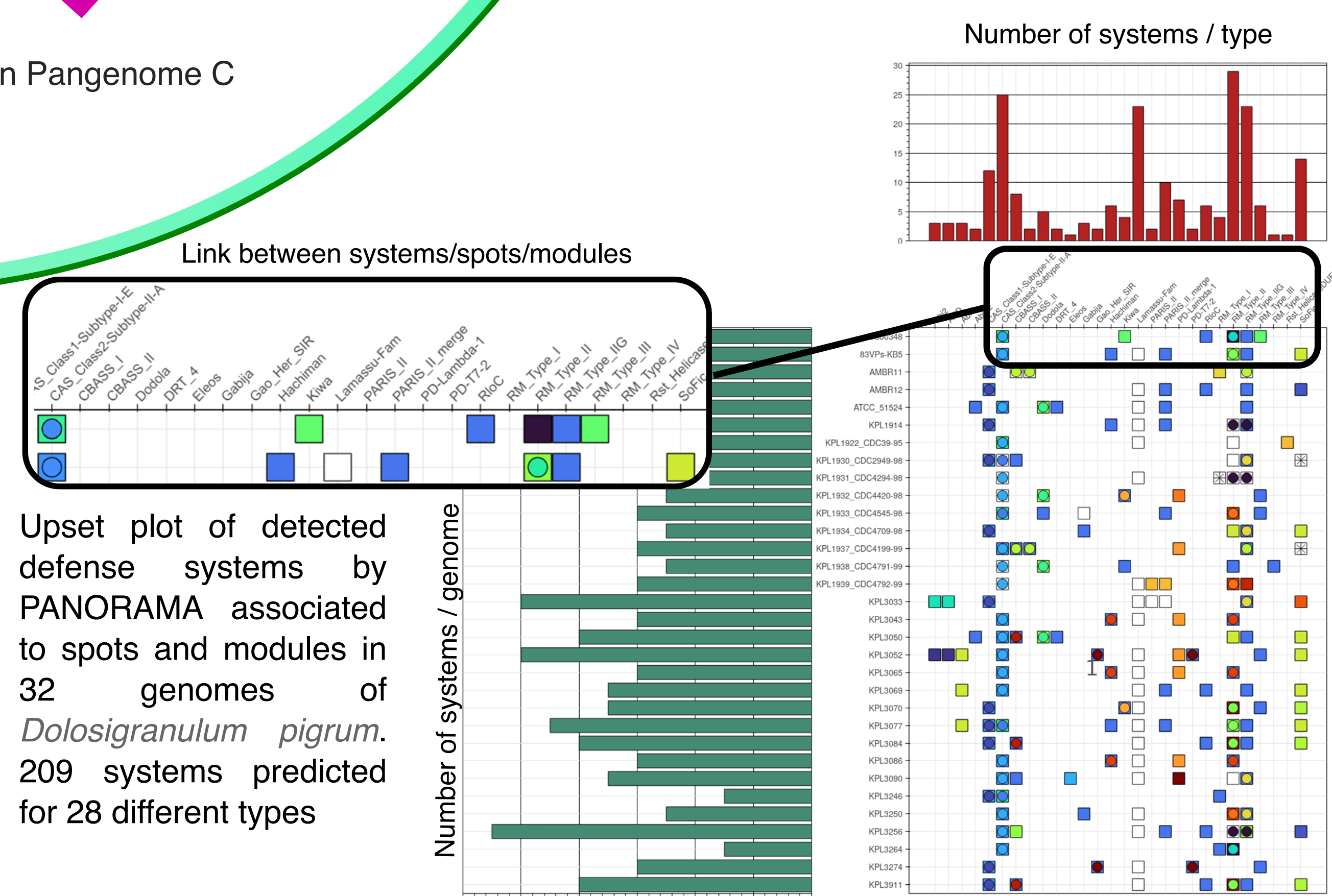
461 models integrated:

- defense(435)
- secretion(26)
- Ongoing systems:
- KEGG modules
- Polysaccharides Utilization Loci

SYSTEM-ASSOCIATION

Systems detected at the pangenome level can be projected to individual genomes and associated to pangenome graph components (partition, region of genomic plasticity², spots² and modules³) and give them a functional annotation.

Using PANORAMA to predict defense systems allow to identify regions of genomic plasticity corresponding to anti-phage defense islands in genomes



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 N. Bertrand et al. Peer Community Journal, 2023, doi: 10.24072/pcjournal.250
- 5 F. Tesson et al. Nat Commun 2022, doi: 10.1038/s41467-022-30269-9
- 6 L.J. Payne et al. Nucleic Acids Res 2021, doi: 10.1093/nar/gkab883
- 7 M. Kanehisa et al. Nucleic Acids Res 2022, doi: 10.1093/nar/gkac963
- 8 J.M. Grondin et al. J Bacteriol 2017, doi: 10.1128/JB.00860-16

ACKNOWLEDGMENT



Contact

jarnoux at genoscope.cns.fr



@jpjarnoux