

# PanExplorer: a web application for exploratory analysis and visualization of microbial pan-genomes

**Easily explore the pan-genomic landscape of your favorite microbes.**

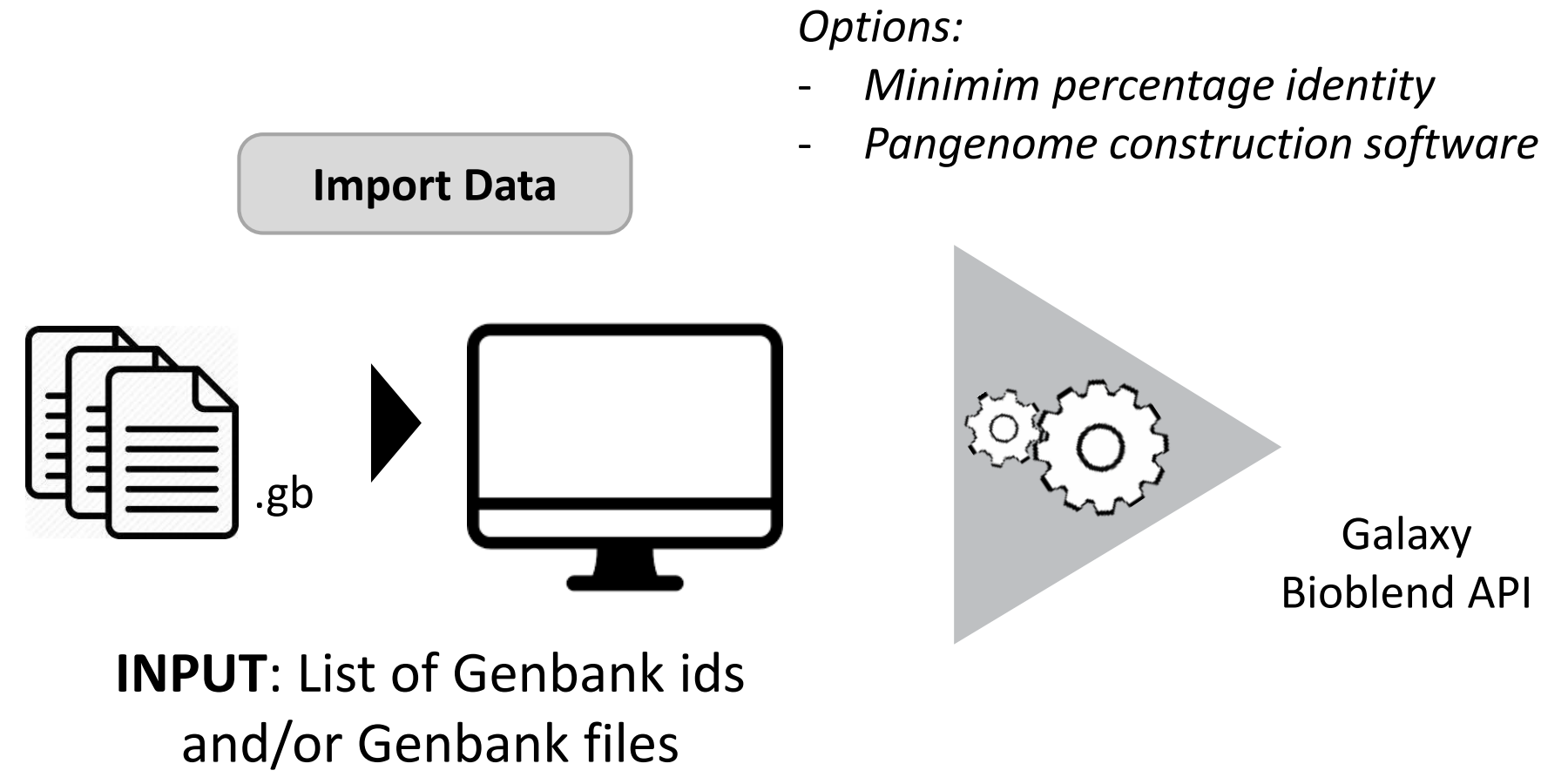
Alexis Dereeper<sup>1,2,\*</sup>, Sébastien Cunnac<sup>1</sup>, Ralf Koenig<sup>1</sup>, Marilyne Summo<sup>3,2</sup>, Damien Meyer<sup>4</sup>  
 (1) PHIM Plant Health Institute, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France  
 (2) French Institute of Bioinformatics (IFB) - South Green Bioinformatics Platform, Bioversity, CIRAD, INRAE, IRD, F-34398 Montpellier, France  
 (3) CIRAD, UMR AGAP, F-34398 Montpellier, France  
 (4) CIRAD, UMR ASTRE, 97170, Petit-Bourg, Guadeloupe, France  
 \* email: alexis.dereeper@ird.fr

<https://panexplorer.southgreen.fr/>

## 1) Import your genomes

Import your selection of microbial genomes. Genomes must be completely assembled and annotated. Input can be:

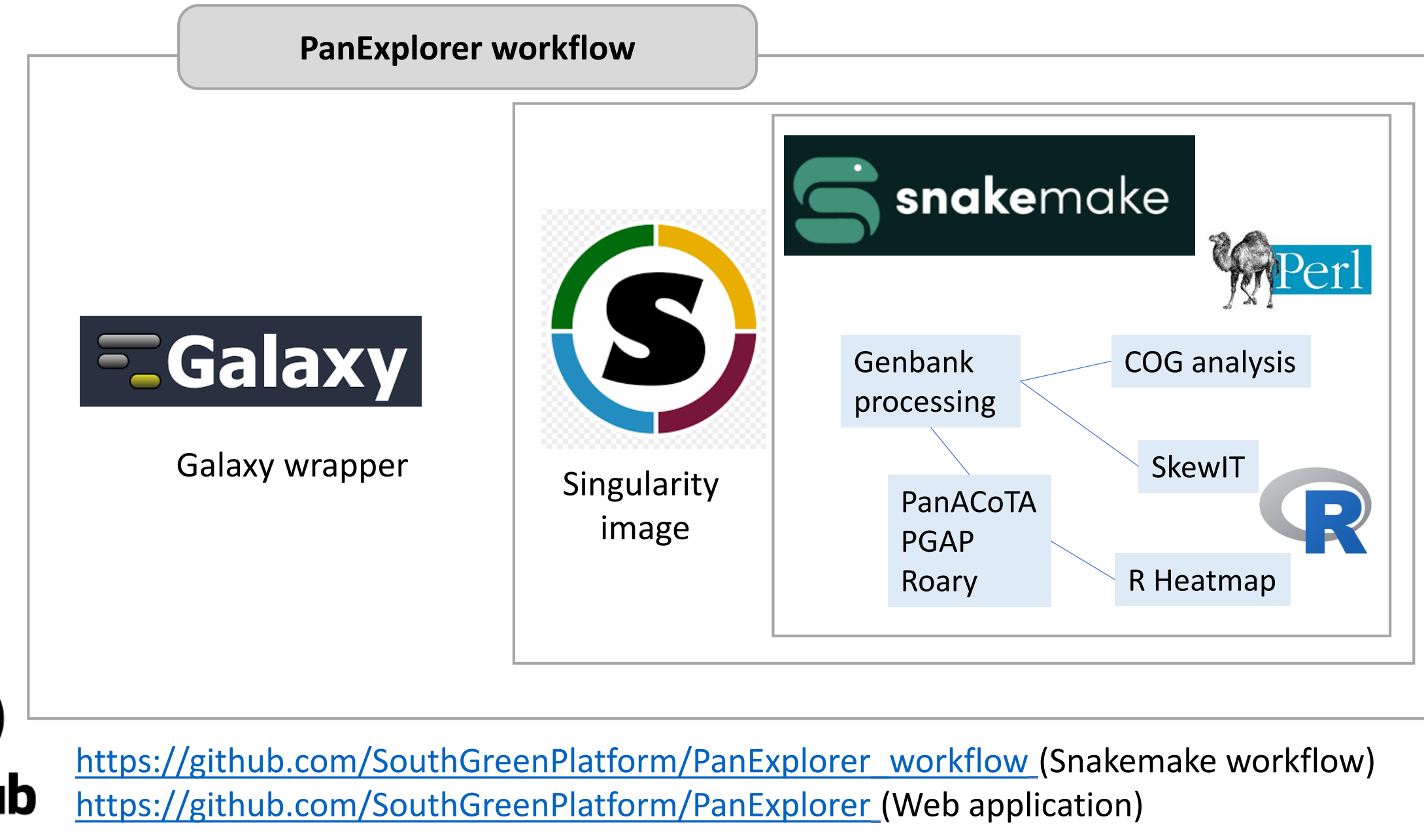
- ✓ a list of **Genbank identifiers** (public genomes)
- ✓ a collection of **Genbank files** (private genomes)



## 2) Process the PanExplorer workflow

The workflow consists of a Snakemake workflow that executes successively:

- ✓ Retrieval of the corresponding gene sequences
- ✓ Gene clustering and pan-genome analysis using PGAP (1), Roary (2) or PanAcota (3),
- ✓ Attribution of COG functional categories
- ✓ GC content calculation using SkewIT (4)
- ✓ Hierarchical clustering, Heatmap, Upset diagram using R



[https://github.com/SouthGreenPlatform/PanExplorer\\_workflow](https://github.com/SouthGreenPlatform/PanExplorer_workflow) (Snakemake workflow)  
<https://github.com/SouthGreenPlatform/PanExplorer> (Web application)

## OUTPUTS/VISUALIZATIONS

**Overview (HivePlot)**

**Macro-synteny**

**Zoomable (Mauve viewer (4))**

**Micro-synteny: focus on a specific region (Clinker)**

**Core-genes-based synteny**

**Investigation of a specific cluster**

**5) Visual inspection of a specific cluster.**  
Sequence alignment, haplotypes, NJ tree, SNPs, network

**Advanced search of clusters respecting presence/absence pattern**

**6) Towards pan-GWAS (association studies)**  
A dedicated module allows to search for clusters that respect particular pattern of gene presence among strains (phenotype)

**Pan-genome overview**

**Presence/Absence Variation (PAV) matrix**

**3) Pan-genome overview:**

- ✓ Distribution pie chart of **core-genes**, strain-specific genes, other accessory genes
- ✓ Visualization as a presence/absence heatmap
- ✓ Representation of accessory genes as an UpSet diagram

**Upset diagram**

**4) Physical map of core-genes and strain-specific genes** can be displayed as a circular genomic representation (**Circos**), for each genome taken independently

**The web application allows interactive data exploration at different levels**

**Phylogeny (PhyloTree.js)**

## References

- (1) Zhao Y. et al. (2012) PGAP: pan-genomes analysis pipeline. *Bioinformatics*, 28, 416–418.
- (2) Page A.J. et al. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31, 3691–3693.
- (3) Perrin A. et al. (2021) PanAcota: a modular tool for massive microbial comparative genomics. *NAR Genomics Bioinform.*, 3, lqaa106.
- (4) Lu J. et al. (2020) SkewIT: the skew index test for large-scale GC skew analysis of bacterial genomes. *PLoS Comput. Biol.*, 16, e1008439.
- (5) Darling A.C. et al. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.*, 14, 1394-1403.



Take a picture to download the full paper

