

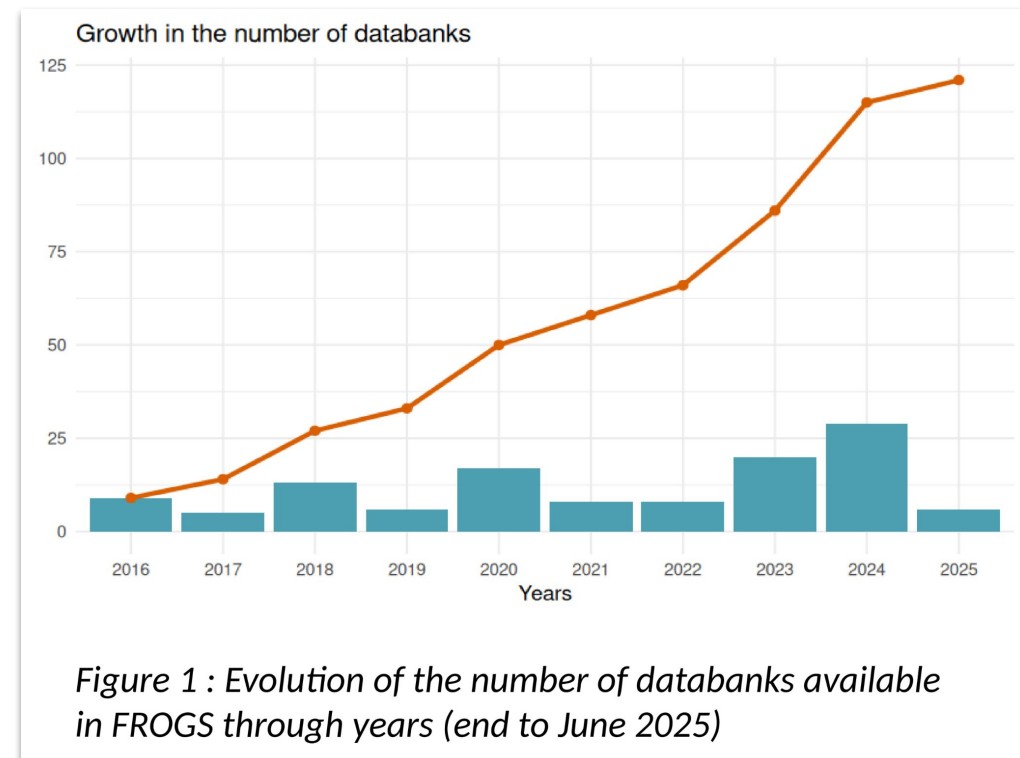


LEAP - *L*earn and *E*valuate *A*ffiliation databanks on an online *P*latform

Olivier Rué, Maria Bernard, Lucas Auer, Gabryelle Agoutin, Maëlle Pomiès,
Géraldine PASCAL

Context

- FROGS offers access to >120 databanks for taxonomic affiliation
 - ❑ Generic or specialized
 - ❑ Several versions (SILVA v. 128, 132, 138, 138.1 ...)



One main question

Which databank is the most relevant for my data?

?



Methodology



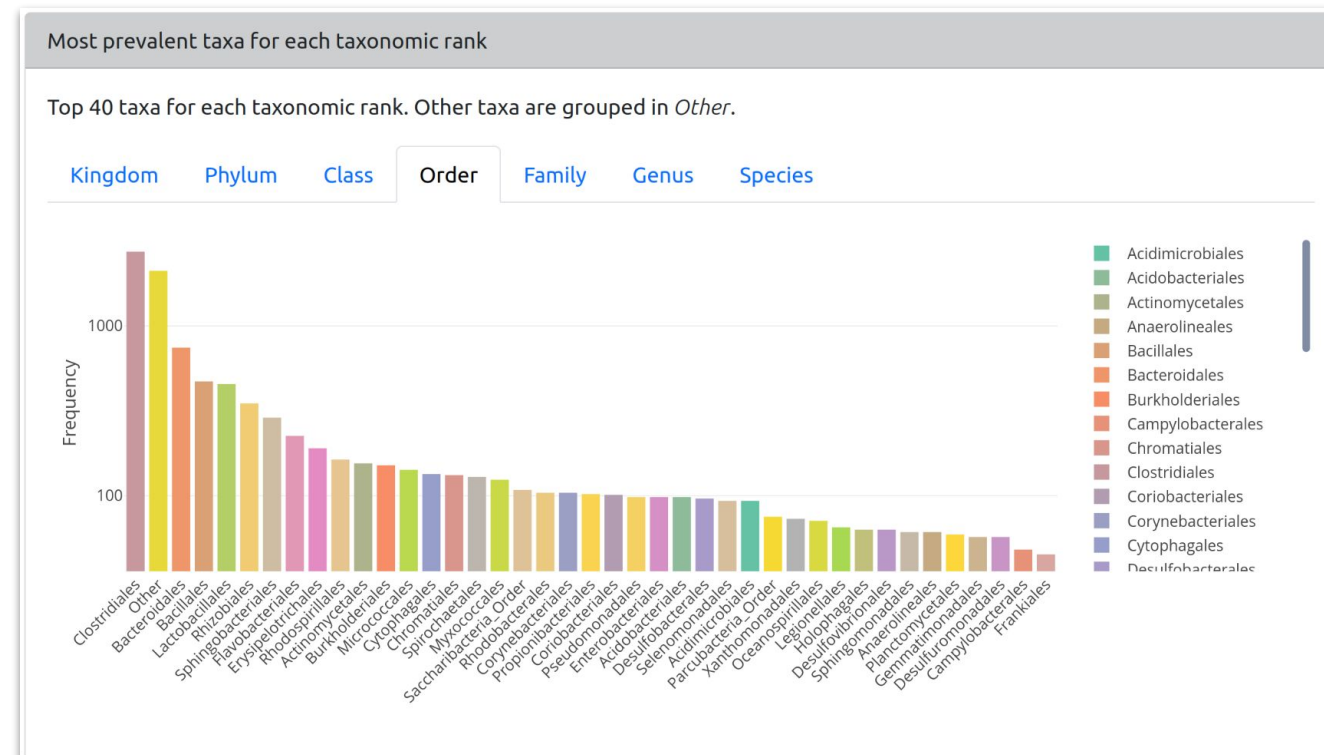
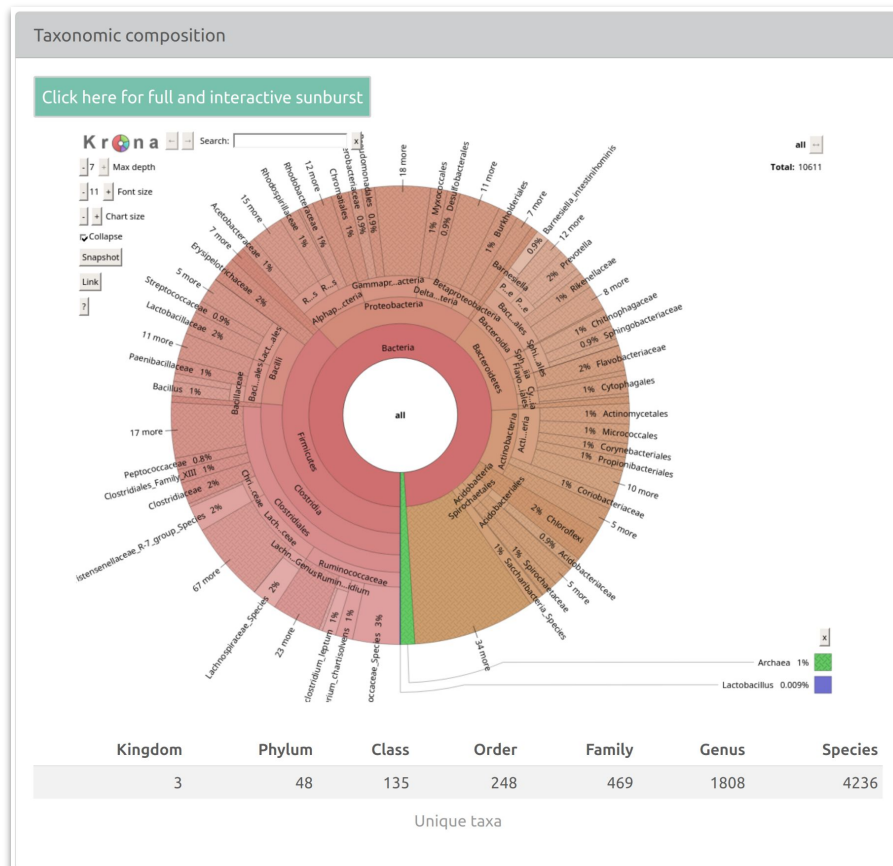
- Web-based interactive platform using Quarto Dashboards to help users explore and understand the content of each databank
- Hosted on Forge INRAE, the site is backed by a CI/CD deployment pipeline, ensuring that dashboards are automatically updated and published as soon as a new databank becomes available. This approach guarantees reproducibility, traceability, and continuous access to the latest data.

6 different questions

1. What is the taxonomic composition of my databank at each rank?

6 use cases

1. What is the taxonomic composition of my databank at each rank?

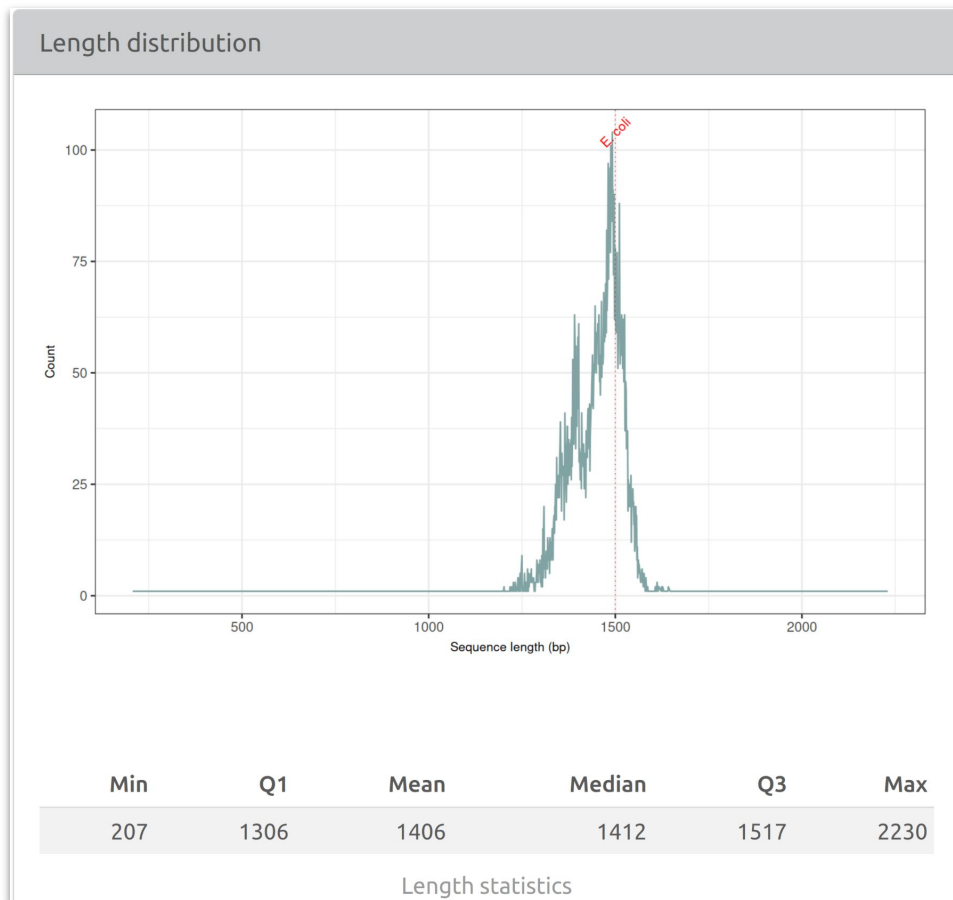


6 use cases

2. What is the distribution of sequence lengths?

6 use cases

2. What is the distribution of sequence lengths?

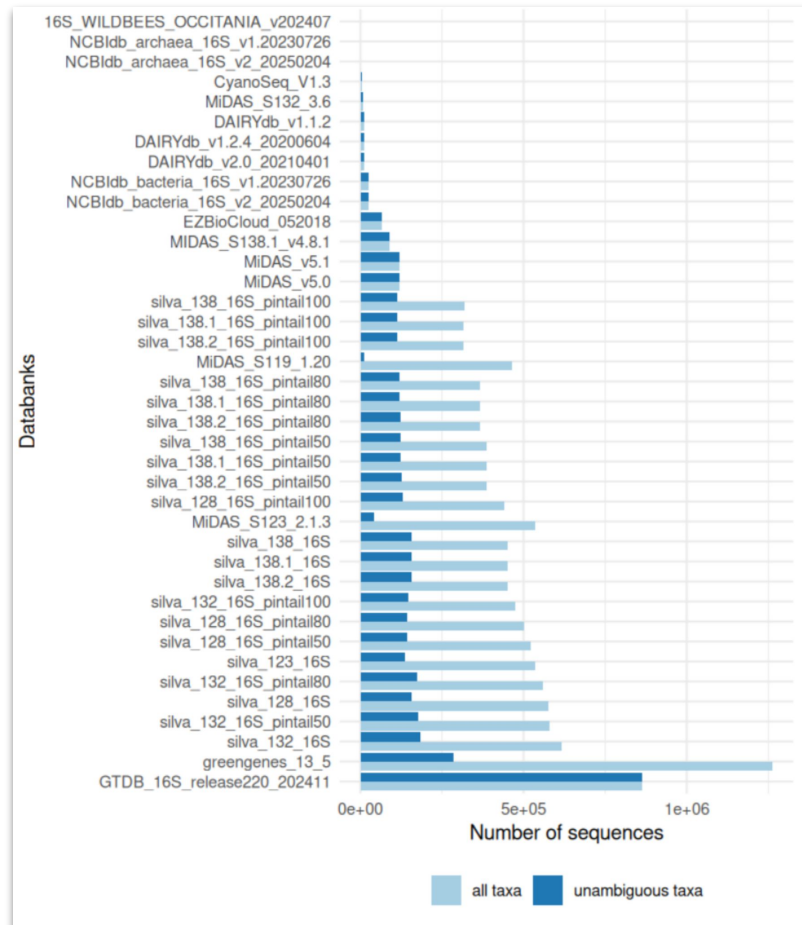


6 use cases

3. What proportion of taxa are unknown, uncertain, or ambiguous?

6 use cases

3. What proportion of taxa are unknown, uncertain, or ambiguous?



Ambiguous taxa

keyword	count
metagenome	70
unknown	14468
unspecified	0
unidentified	0
unclassified	0
incertae	6
NA	0
None	0

Detail of ambiguous key-words

6 use cases

4. How many species are associated with a genus of interest?

6 use cases

4. How many species are associated with a genus of interest?

Taxonomies							
Taxonomic information of each rank. Only unique taxonomies and their counts are displayed.							
CSV export		Search...					
Kingdom	Phylum	Class	Order	Family	Genus	Species	Count
Eukaryota	Phragmoplastophyta	unknown class	unknown order	unknown family	unknown genus	Pinus taeda (loblolly pi...	462
Eukaryota	Phragmoplastophyta	unknown class	unknown order	unknown family	unknown genus	Ipomoea nil (Japanese...	436
Eukaryota	Phragmoplastophyta	unknown class	unknown order	unknown family	unknown genus	Cucumis sativus (cucu...	328
Eukaryota	Phragmoplastophyta	Embryophyta	Liliopsida	unknown family	Zea	Zea mays	224
Eukaryota	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	Fusarium	Fusarium graminearum	179
Eukaryota	Phragmoplastophyta	unknown class	unknown order	unknown family	unknown genus	Gossypium hirsutum (...)	162
Eukaryota	Phragmoplastophyta	unknown class	unknown order	unknown family	unknown genus	Gossypium barbadens...	159
Eukaryota	Phragmoplastophyta	unknown class	unknown order	unknown family	unknown genus	Triticum aestivum (bre...	153
Eukaryota	Phragmoplastophyta	unknown class	unknown order	unknown family	unknown genus	Arabidopsis thaliana (t...	144
Eukaryota	Phragmoplastophyta	Embryophyta	Solanales	unknown family	Capsicum	Capsicum annuum	114
Eukaryota	Phragmoplastophyta	Embryophyta	Sapindales	unknown family	Citrus	Citrus maxima	111
Eukaryota	Arthropoda	Insecta	unknown order	unknown family	unknown genus	Drosophila melanogas...	101
Eukaryota	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycopsidaceae	Saccharomycopsis	Saccharomycopsis fibu...	101
Eukaryota	Phragmoplastophyta	Embryophyta	Malvales	unknown family	Gossypium	Gossypium arboreum	101
Eukaryota	Phragmoplastophyta	unknown class	unknown order	unknown family	unknown genus	Brassica napus (rape)	95
Eukaryota	Ascomycota	Dothideomycetes	Pleosporales	Leptosphaeriaceae	Leptosphaeria	Leptosphaeria biglobo...	84

6 use cases

5. Which taxa are shared or different between databanks or versions?

6 use cases

5. Which taxa are shared or different between databanks or versions?

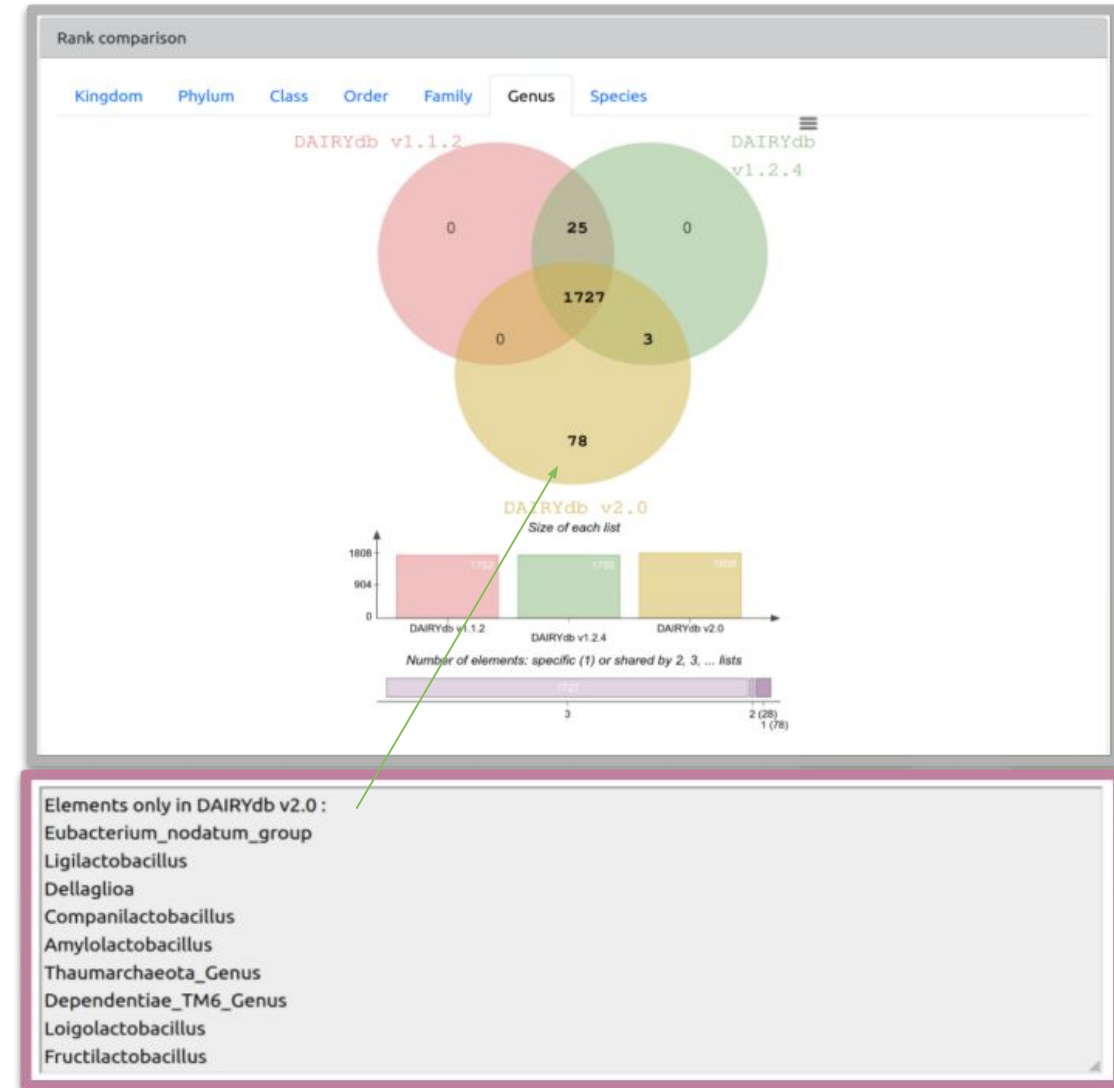


6 use cases

6. Which taxa are unique to a specific databank or version?

6 use cases

6. Which taxa are unique to a specific databank or version?



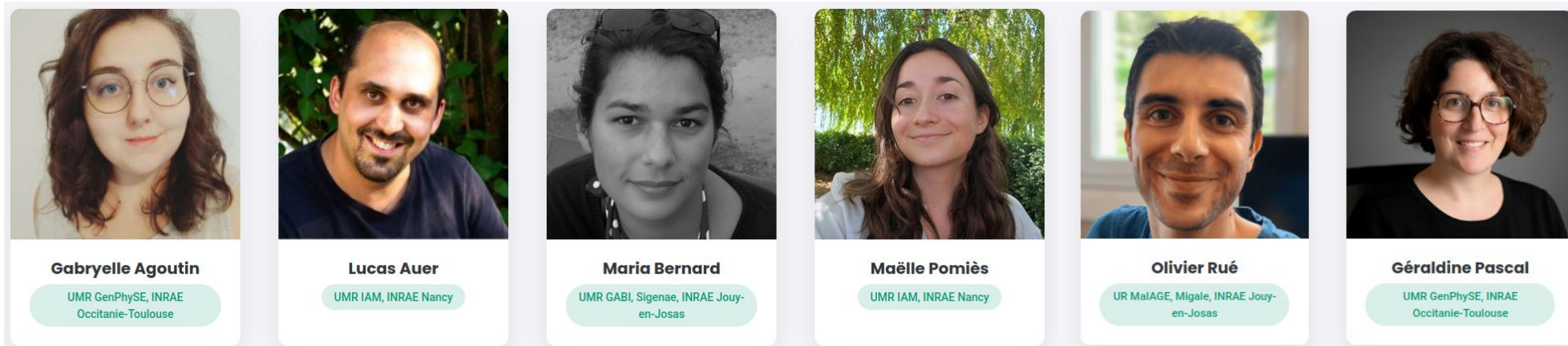
Access

<https://leap.frogs.toulouse.inrae.fr/>

Perspectives

- ❑ Enhanced exploration
 - Enable searches across multiple or all databanks simultaneously
 - Offer on-demand databank comparisons (currently pre-defined)
 - Use sequence content to detect inconsistent taxonomic assignments
- ❑ Interoperability & Integration
 - Create links with the Affiliation Explorer application:
https://shiny.migale.inrae.fr/app_direct/affiliationexplorer/
- ❑ Automation & Distribution
 - Automate dashboard generation during databank formatting
 - Distribute a generic, standalone version via a Conda package

Thank you for your attention



INRAE

GenOToul Bioinfo

SIGENAE

GenPhySE
Génétique, Physiologie et Systèmes d'Élevage

MaIAGE

GABI

migale

IAM
INTERACTIONS
ARBRES-MICROORGANISMES