



HAL
open science

Florilege, a database gathering microbial habitats, phenotypes and uses

Sandra Derozier, Louise Deleger, Estelle Chaix, Reda Mekdad, Mouhamadou Ba, Delphine Sicard, Valentin Loux, H el ene Falentin, Claire N edellec

► To cite this version:

Sandra Derozier, Louise Deleger, Estelle Chaix, Reda Mekdad, Mouhamadou Ba, et al.. Florilege, a database gathering microbial habitats, phenotypes and uses. JOBIM 2020, Jun 2020, Montpellier, France. hal-02904156

HAL Id: hal-02904156

<https://hal.inrae.fr/hal-02904156>

Submitted on 21 Jul 2020

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 destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



INRAE



➤ Florilege, a database gathering microbial habitats, phenotypes and uses

Dérozier S.^{1,2}, Deléger L.¹, Chaix E.¹, Mekdad R.¹, Ba M.^{1,2}, Bossy R.¹, Sicard D.⁴, Loux V.^{1,2}, Falentin H.⁵ and Nédellec C.¹

¹ Université Paris-Saclay, INRAE, MaiAGE, 78350, Jouy-en-Josas, France

² Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

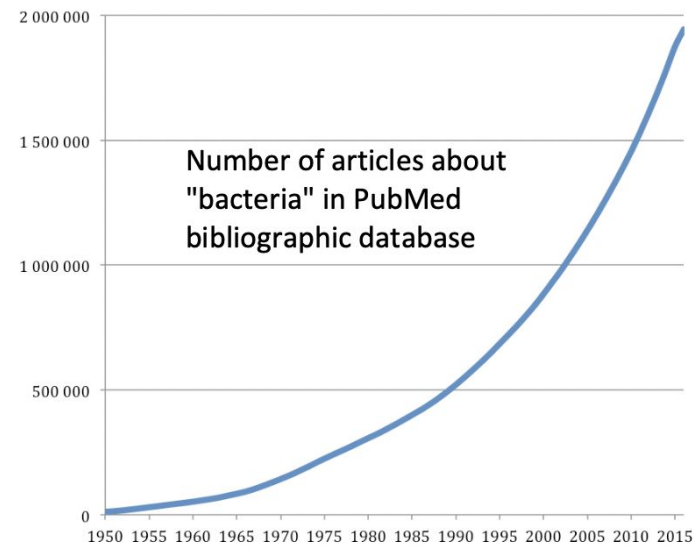
³ INRAE, UMR SPO, Montpellier, France

⁴ INRAE, UMR STLO, Rennes, France

Context

Food microbiology research has led to an exponential growth of experimental data and publications. It is now crucial for researchers to have bioinformatics applications that offer unified access to both data and related scientific articles.

Florilege uses an Information Extraction workflow to populate its database.



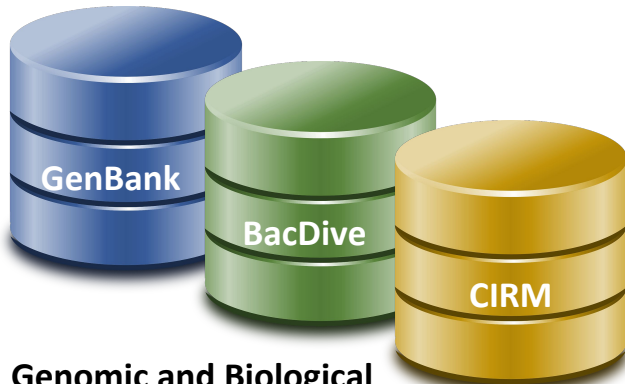
The workflow is designed to (1) extract microorganism taxa, their habitats, their phenotypes and their uses and (2) categorize the extracted information by means of taxa from the NCBI taxonomy and concepts from the OntoBiotope ontology⁵. The Florilege application combines information from other databases with knowledge from the literature (PubMed) on microbial biodiversity, to support their comparison for further analysis.

⁵ Chaix E. et al. Text mining tools for extracting information about microbial biodiversity in food Food Microbiology, 2018.

Relationship types



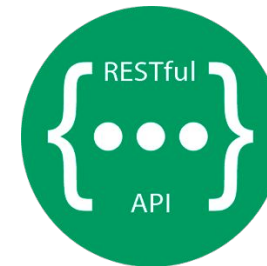
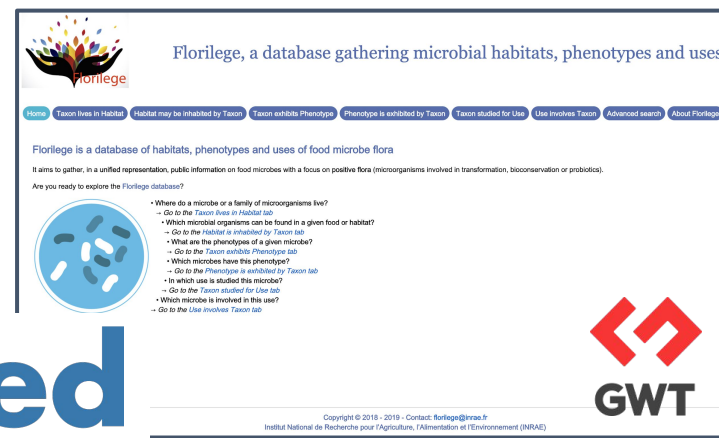
Florilege database



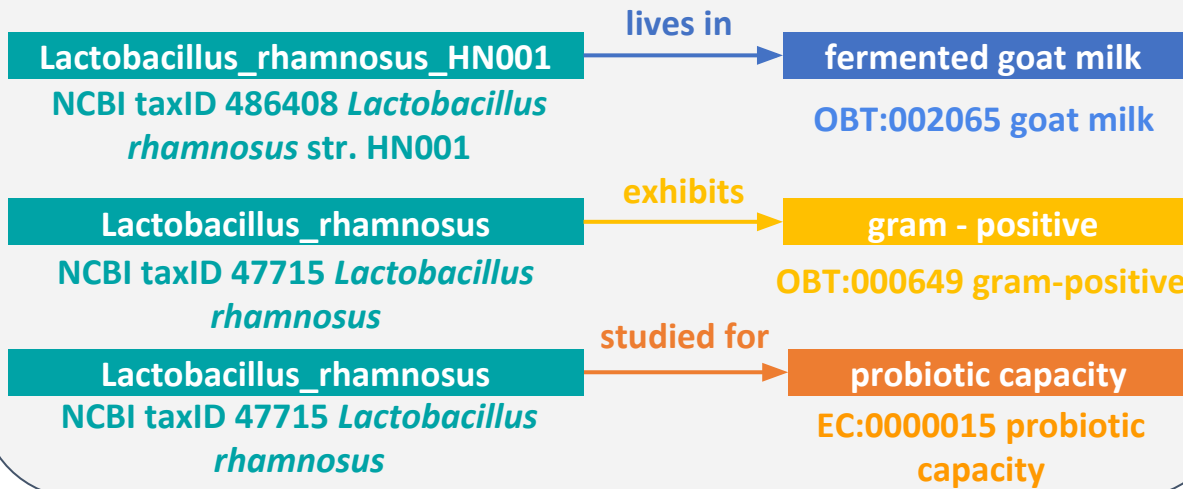
Genomic and Biological Resource Centers databases



Abstracts of scientific publications



Relations extracted by text-mining



Snakemake

Florilege web interface



Florilege, a database gathering microbial habitats, phenotypes and uses

- Home
- Taxon lives in Habitat
- Habitat may be inhabited by Taxon
- Taxon exhibits Phenotype
- Phenotype is exhibited by Taxon
- Taxon studied for Use
- Use involves Taxon
- Advanced search
- About Florilege

- microbial habitat
 - animal habitat
 - animal husbandry and agricultural habitat
 - aquaculture habitat
 - artificial environment
 - experimental medium
- food
 - animal feed
 - food for human
 - commodity and primary derivative th
 - additive
 - animal product and primary derive
 - animal based juice
 - egg and egg product
 - honey and apiculture product
 - meat and meat product
 - milk and milk product
 - butter
 - buttermilk
 - cheese

Search relations by habitat

TSV Download

Filter Selection

190 relations for the habitat "cheese"

Source

- PubMed
- GenBank
- CIRM
- DSMZ

Taxon

QPS only

Apply

SOURCE TEXT	HABITAT	RELATION TYPE	TAXON	QPS	SOURCE
6605	cheese	may be inhabited by	Lactobacillus acidipiscis		DSMZ
11211271, 20538362, 21742864	cheese	may be inhabited by	Lactobacillus acidipiscis		PubMed
22574688	semi soft cheese	may be inhabited by	Lactobacillus acidophilus	<input checked="" type="checkbox"/>	PubMed
17357571, 17582095, 21264685	Habitat: cheese Appears in the text as: ovine cheese, Pecorino cheese, petit - suisse cheese, experimental cheese, Minas Frescal cheese, Brazilian goat semi - hard cheese, Egyptian home - made cheese, Fresco cheese environment, probiotic cheese, Scamorza cheese, creamy goat cheese, regular cheese, cheese	may be inhabited by	Taxon: Lactobacillus acidophilus Appears in the text as: L._acidophilus, Lactobacillus_acidophilus	<input checked="" type="checkbox"/>	PubMed
27112363, 24020254		may be inhabited by	Lactobacillus acidophilus	<input checked="" type="checkbox"/>	PubMed
21943729, 19751954, 22720913		may be inhabited by	Lactobacillus acidophilus	<input checked="" type="checkbox"/>	PubMed

Copyright © Institut National de Recherche p

e.fr vironnement (INRAE)



Florilege - <http://migale.iouy.inrae.fr/florilege/>
30/06/2020 - 03/07/2020 JOBIM 2020 florilege@inrae.fr

Conclusion

Florilege is integrating an increasing volume of textual and non-textual information from relevant biological databases:

- **659 508** Taxa - Habitat relations (575 822 PubMed, 63 534 GenBank, 639 INRAE CIRM BIA, 19 513 DSMZ through BacDive⁶)
- **43 742** Taxa - Phenotype relations (PubMed)
- **10 408** Taxa - Use relations (PubMed)

Florilege offers a powerful semantic search engine that enables ontology-based query to support Information Retrieval.

Access to Florilege:

- a web application displays a unique set of structured information on food microbiota, publicly accessible at <http://migale.jouy.inra.fr/florilege/>.
- an API (Application Programming Interface) that allows one to automatically integrate microbe biodiversity in external information systems.
API Documentation: <http://migale.jouy.inra.fr/florilege-api/api-doc/>.

In a recent study conducted at STLO, Florilege was used for the selection of species fermenting soy milk.