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, Hélène Falentin, Claire Nédellec

**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-02904156v1**

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.

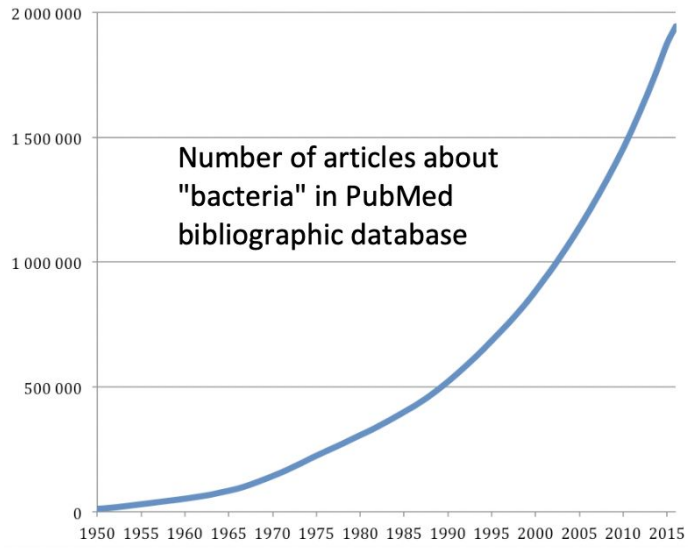
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 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

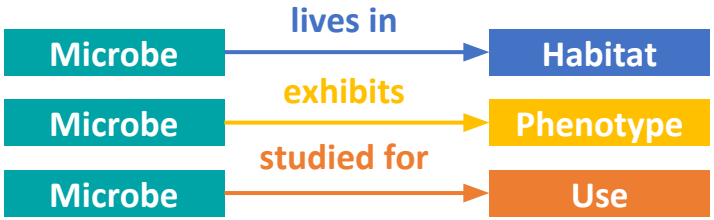



Food microbiology research has led to an exponential growth of experimental data and publications. It is now crucial for researchers to have bioinformatics applications that offer unified access to both data and related scientific articles.

Florilege uses an Information Extraction workflow to populate its database.

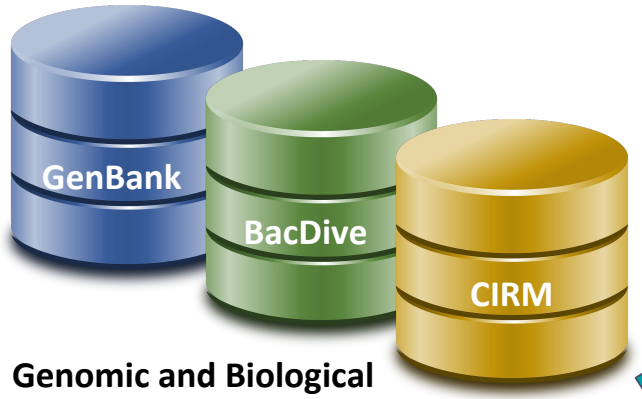
The workflow is designed to (1) extract microorganism taxa, their habitats, their phenotypes and their uses and (2) categorize the extracted information by means of taxa from the NCBI taxonomy and concepts from the OntoBiotope ontology[5]. The Florilege application combines information from other databases with knowledge from the literature (PubMed) on microbial biodiversity, to support their comparison for further analysis.

**Relationship types**




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

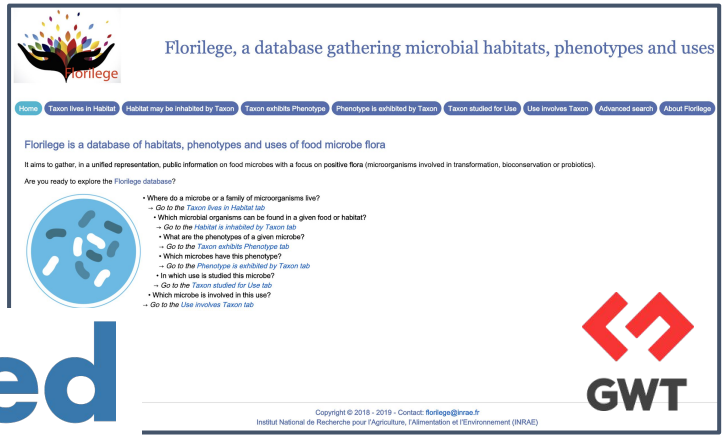
# Florilege database

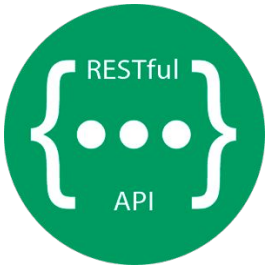



Genomic and Biological Resource Centers databases

PubMed

Abstracts of scientific publications




PostgreSQL

## Relations extracted by text-mining

Lactobacillus_rhamnosus_HN001 —lives in→ fermented goat milk

NCBI taxID 486408 *Lactobacillus rhamnosus* str. HN001

OBT:002065 goat milk

Lactobacillus_rhamnosus —exhibits→ gram - positive

NCBI taxID 47715 *Lactobacillus rhamnosus*

OBT:000649 gram-positive

Lactobacillus_rhamnosus —studied for→ probiotic capacity

NCBI taxID 47715 *Lactobacillus rhamnosus*

EC:0000015 probiotic capacity

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 deriva
          - animal based juice
        - egg and egg product
        - honey and apiculture product
        - meat and meat product
        - milk and milk product
          - butter
          - buttermilk
        - cheese

1

Search relations by habitat: cheese | TSV Download | Filter Selection 2

190 relations for the habitat "cheese

Source: PubMed, GenBank, CIRM, DSMZ | Taxon: Lactobacillus 3 | QPS only | Apply 4

| 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 | ✓ | PubMed |
| 17357571, 17582095, 21264685 | | may be inhabited by | | ✓ | PubMed |
| 27112363, 24020254 | | may be inhabited by | | ✓ | PubMed |
| 21943729, 19751954, 22720913 | | may be inhabited by | Lactobacillus acidophilus | ✓ | PubMed |

5 6 7 8

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

Taxon: Lactobacillus acidophilus
Appears in the text as: L._acidophilus, Lactobacillus_acidophilus

Copyright © 2 ... e.fr
Institut National de Recherche p ... nvironnement (INRAE)

# 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 BacDive[6])
- ➔ **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.

[6] Söhngen C. et al. (2015). Bac Dive - The Bacterial Diversity Metadatabase in 2016. Nucleic acids research, 44 (D1), D581-D585.