Florilege, a database gathering microbial habitats, phenotypes and uses

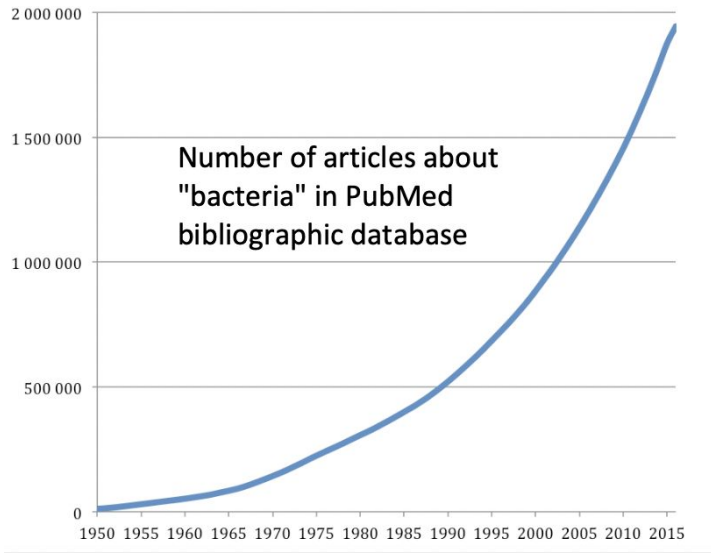
Sandra Derozier, Louise Deleger, Estelle Chaix, Reda Mekdad, Mouhamadou Ba, Delphine Sicard, Valentin Loux, Hélène Falentin, Claire Nédellec

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

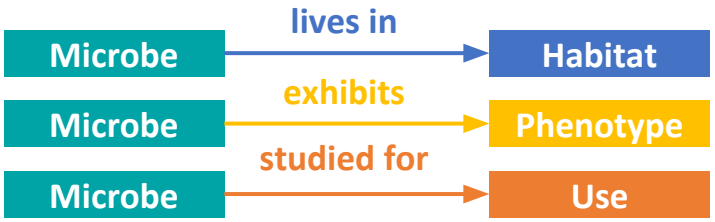


Food microbiology research has led to an exponential growth of experimental data and publications. It is now crucial for researchers to have bioinformatics applications that offer unified access to both data and related scientific articles.

Florilege uses an Information Extraction workflow to populate its database.

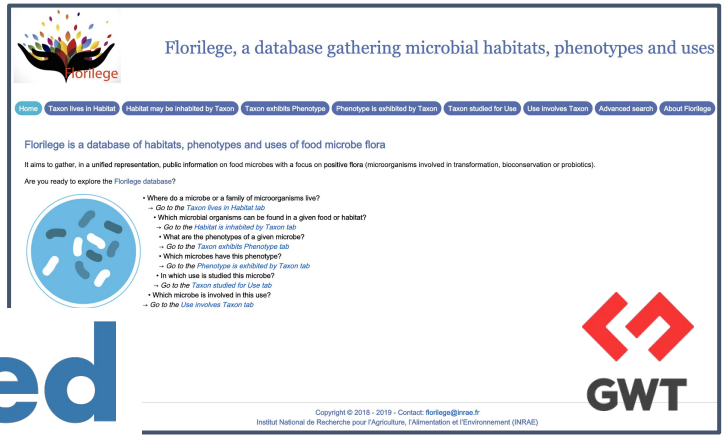
The workflow is designed to (1) extract microorganism taxa, their habitats, their phenotypes and their uses and (2) categorize the extracted information by means of taxa from the NCBI taxonomy and concepts from the OntoBiotope ontology[5]. The Florilege application combines information from other databases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

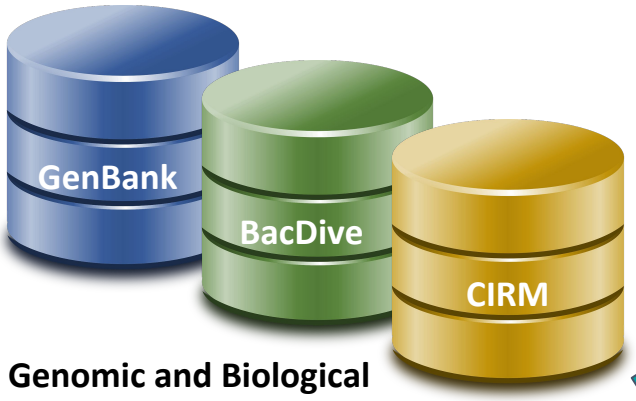
Relationship types



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

Florilege database

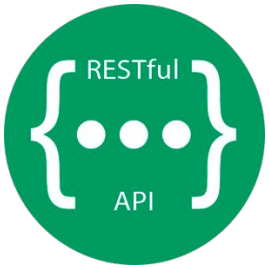




Genomic and Biological Resource Centers databases

PubMed

Abstracts of scientific publications



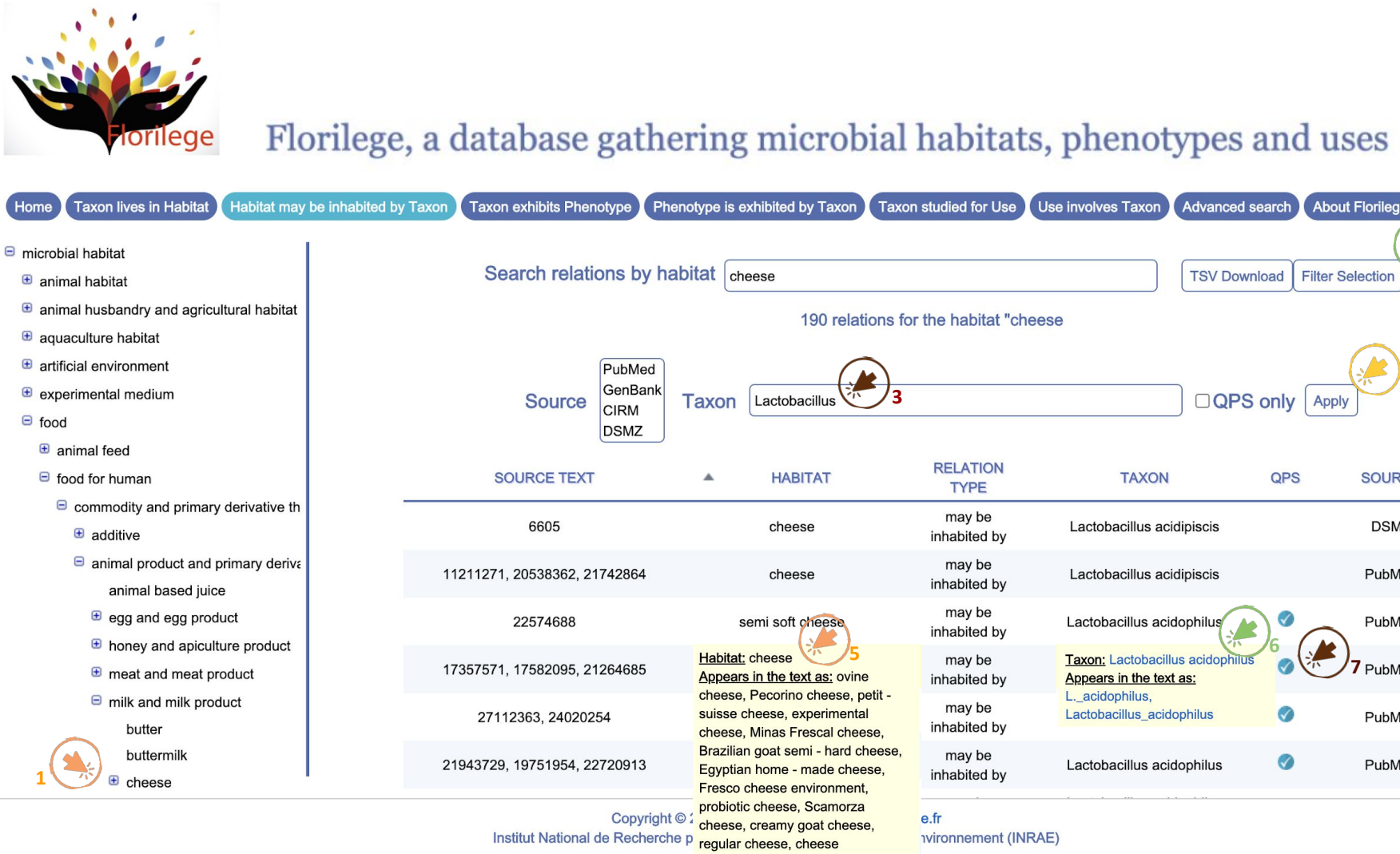
Relations extracted by text-mining

Lactobacillus_rhamnosus_HN001 NCBI taxID 486408 *Lactobacillus rhamnosus* str. HN001	lives in	fermented goat milk OBT:002065 goat milk
Lactobacillus_rhamnosus NCBI taxID 47715 *Lactobacillus rhamnosus*	exhibits	gram - positive OBT:000649 gram-positive
Lactobacillus_rhamnosus NCBI taxID 47715 *Lactobacillus rhamnosus*	studied for	probiotic capacity EC:0000015 probiotic capacity

INRAE

Snakemake

Florilege web interface



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.