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

\textsuperscript{1} Université Paris-Saclay, INRAE, MaiAGE, 78350, Jouy-en-Josas, France
\textsuperscript{2} Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France
\textsuperscript{3} INRAE, UMR SPO, Montpellier, France
\textsuperscript{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.

\(^5\) Chaix E. et al. Text mining tools for extracting information about microbial biodiversity in food Food Microbiology, 2018.
Florilege database

GenBank  BacDive  CIRM

Abstracts of scientific publications

Genomic and Biological Resource Centers databases

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
Florilege - http://migale.jouy.inrae.fr/florilege/
30/06/2020 - 03/07/2020  JOBIM 2020  florilege@inrae.fr
Florilege web interface

Florilege, a database gathering microbial habitats, phenotypes and uses

Search relations by habitat: cheese

190 relations for the habitat "cheese"

<table>
<thead>
<tr>
<th>Source</th>
<th>Taxon</th>
<th>Relation Type</th>
<th>Habitats</th>
<th>Type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>6605</td>
<td>Lactobacillus</td>
<td>may be inhabited by</td>
<td>cheese</td>
<td>Lactobacillus acidipiscis</td>
<td>DSMZ</td>
</tr>
<tr>
<td>11211271, 20538362, 21742864</td>
<td>Lactobacillus acidipiscis</td>
<td>may be inhabited by</td>
<td>cheese</td>
<td>DSMZ</td>
<td></td>
</tr>
<tr>
<td>22574688</td>
<td>Lactobacillus acidophilus</td>
<td>may be inhabited by</td>
<td>semi soft cheese</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>17357571, 17582096, 21264685</td>
<td>Lactobacillus acidophilus</td>
<td>may be inhabited by</td>
<td>cheese</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>27112363, 24020254</td>
<td>Lactobacillus acidophilus</td>
<td>may be inhabited by</td>
<td>cheese</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>21943729, 19751954, 22720913</td>
<td>Lactobacillus acidophilus</td>
<td>may be inhabited by</td>
<td>cheese</td>
<td>PubMed</td>
<td></td>
</tr>
</tbody>
</table>

Copyright © Institut National de Recherche pour l'Environnement (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)
- **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:

- an API (Application Programming Interface) that allows one to automatically integrate microbe biodiversity in external information systems.

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

---