Text mining tools for extracting information about microbial biodiversity in food

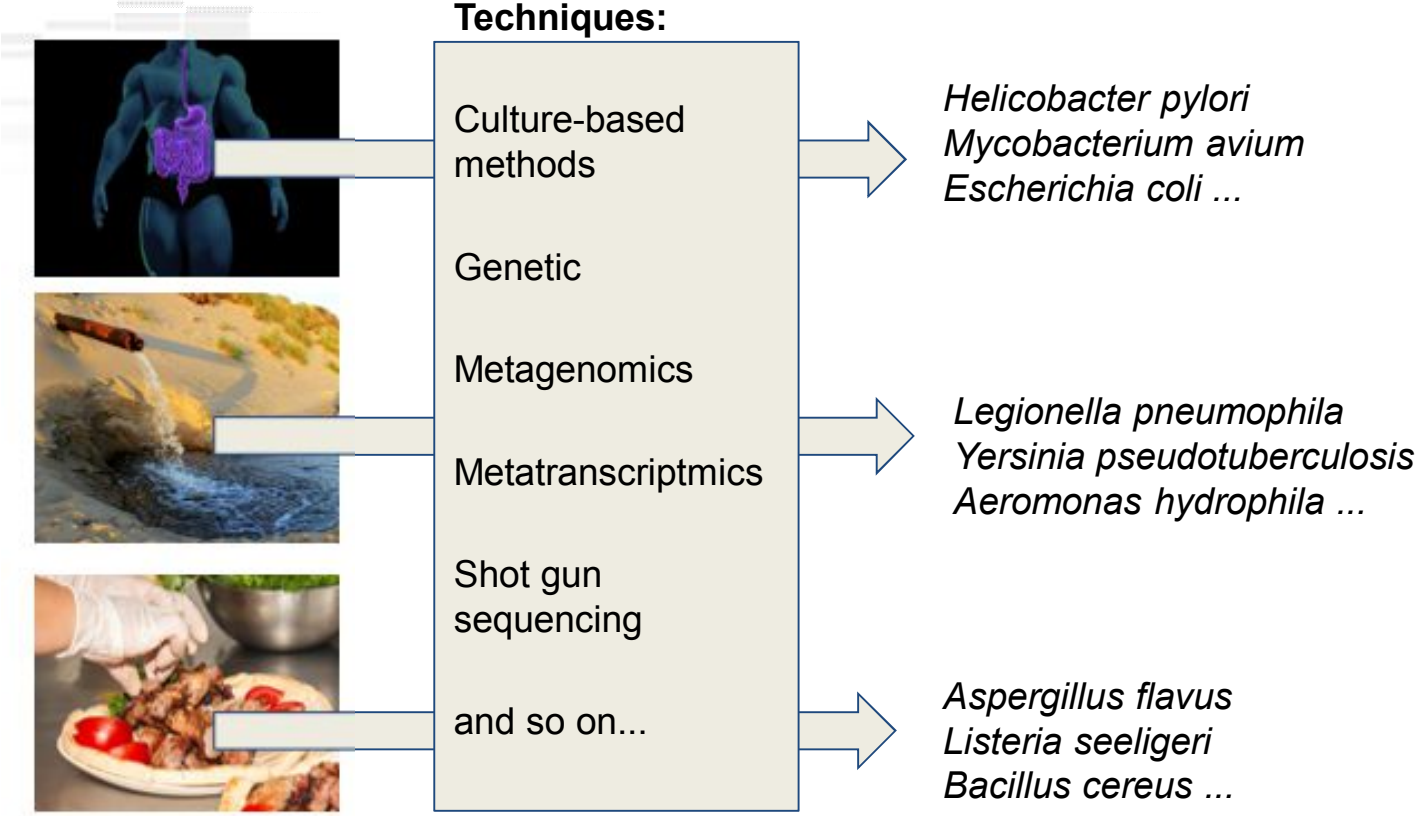
Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec

firstname.lastname@inra.fr

Bibliome Team
MaIAGE, INRA,
Université Paris-Saclay,
78350 Jouy-en-Josas,
France

Microbial ecosystems

Which microbes live in an environment?



Properties of environment? Microbial Interaction?

Crossing between microbial species and habitats

Difficulty : Highly variable forms in text or genomics database (GOLD, SRA, GenBank)

RESEARCH LETTER

Biogenic amine-forming microbial communities in cheese

Radka Burdychova & Tomas Komprda

Department of Food Technology, Mendel University of Agriculture and Forestry in Brno, Brno, Czech Republic

"Bacteria of the genera *Enterococcus* and *Lactobacillus* and coliform bacteria were isolated from Dutch-type semi-hard cheese"

ELSEVIER

International Journal of Food Microbiology 63 (2001) 91–98

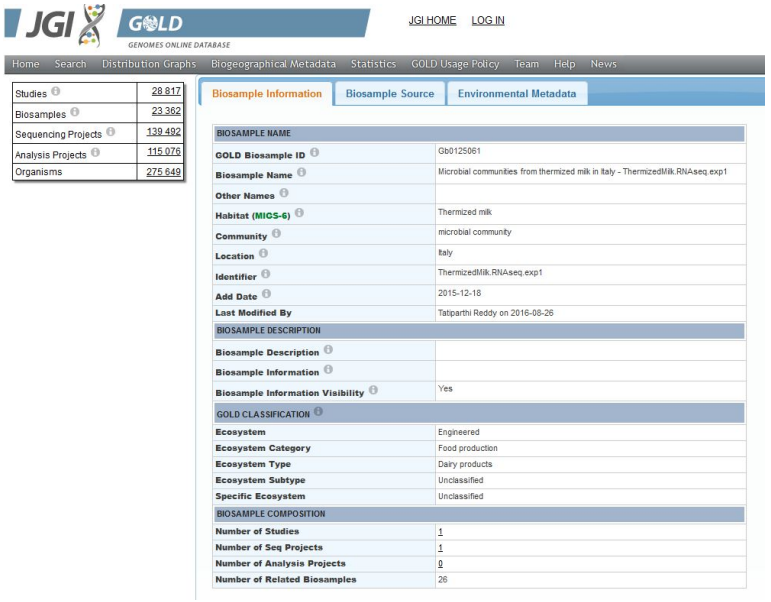
INTERNATIONAL JOURNAL OF Food Microbiology

www.elsevier.nl/locate/ijfoodmicro

High incidence of *Listeria monocytogenes* in European red smear cheese

Melanie Rudolf, Siegfried Scherer*

"Out of European red-smear cheese samples of various types [...] 1.2% of the samples were contaminated with *L. seeligeri*"



e.g.

- Artisanal cheeses from Tucuman
- Dairy cheese
- Caciocavallo cheese in Italy

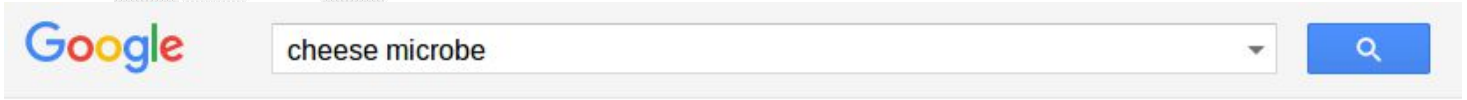
Habitat information is neither queryable nor comparable

Described at different levels of accuracy
and not standardized

What is the cheese microflora?

"Geotrichum candidum strains isolated from a traditional Spanish goats' milk cheese."
"Escherichia coli O157:H7 isolated from raw beef, soft cheese and vegetables in Lima"
"Microbial ecology of Gorgonzola rinds and occurrence of different biotypes of *Listeria monocytogenes*."

Classic search engine query



Scholar — Environ 25 800 résultats (**0,06** s)

POST-PASTEURIAN CULTURES: The Microbiopolitics of Raw-Milk **Cheese** in the United States
H Paxson - Cultural Anthropology, 2008 - Wiley Online Library
... War II servicemen, the FDA now directs its sternest warning about **cheese**-residing **microbes** at pregnant **...** 8 The fear is Listeria monocytogenes, the **microbe** behind listeriosis, which has been linked to **...** that the category "soft" is neither self-evident nor used in the **cheese** world but **...**
Cité 166 fois Autres articles Les 15 versions Citer Enregistrer

The perils and promises of microbial abundance: Novel natures and model ecosystems, from artisanal **cheese** to alien seas
H Paxson, S Helmreich - Social Studies of Science, 2014 - journals.sagepub.com
Cité 43 fois Autres articles Les 7 versions Citer Enregistrer

Molecular approaches to analysing the microbial composition of raw milk and raw milk **cheese**
L Quigley, O O'Sullivan, TP Beresford, RP Ross... - International journal of ..., 2011 - Elsevier
... milk/**cheese**; Microbial composition; Culture-independent microbiology. 1. Introduction. Raw milk is known to harbour a complex microbial community. Indeed the high nutritional value of this food, its high water content and near neutral pH allows the growth of many **microbes** (**...**
Cité 102 fois Autres articles Les 8 versions Citer Enregistrer

Food commensal **microbes** as a potentially important avenue in transmitting antibiotic resistance genes
HH Wang, M Manuzon, M Lehman... - FEMS ..., 2006 - femsle.oxfordjournals.org
... 1. Twenty out of the 23 **cheese** samples analyzed contained Tet r and/or Em r **microbes** ranging from 10 2 to 10 7 CFU g −1 of food, which are equivalent to 10 3 –10 8 CFU ART **microbes** per slice of **cheese** (about 20 g). In general, the number of Tet r **microbes** was greater in **...**
Cité 126 fois Autres articles Les 8 versions Citer Enregistrer

[LIVRE] **Cheese** and **microbes**
CW Donnelly - 2014 - books.google.com
A scientific overview of the association of **microbes** with **cheese**, through the lens of select **cheese** varieties that result due to surface mold ripening, internal mold ripening, rind washing, cave aging, or surface smear rind development. Over the past decade, there has
Cité 11 fois Autres articles Les 8 versions Citer Enregistrer

The query matches
"cheese" and "microbe"
but not
"Camembert" , "Roquefort" or
"*Listeria monocytogenes*"

We propose a semantic search engine dedicated to microbial biodiversity in food.

Alvis Search Engine

Semantic search engine of microbial habitat in food

Interpretation of the query

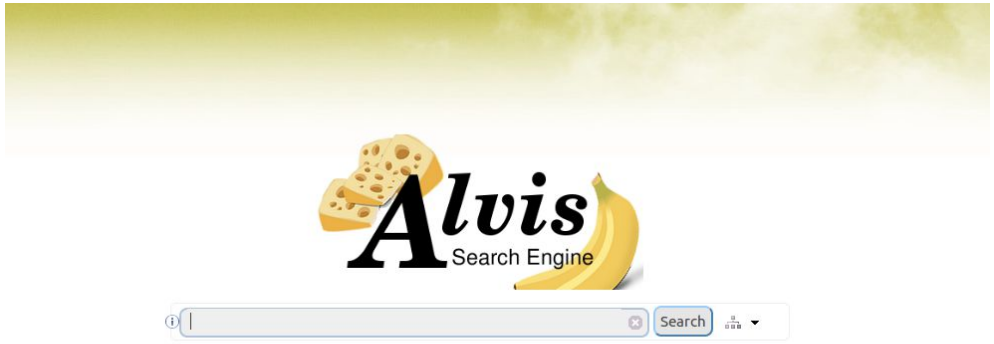
Aspergillus :

Aspergillus (taxon)

- Synonyms (3)
  - Aspergillus
  - Aspergilli
  - Petromyces
- Sub-concepts (50)
  - Aspergillus aculeatus
  - Aspergillus amstelodami
  - Aspergillus clavatus
  - Aspergillus ficuum
  - ...

Cheese :

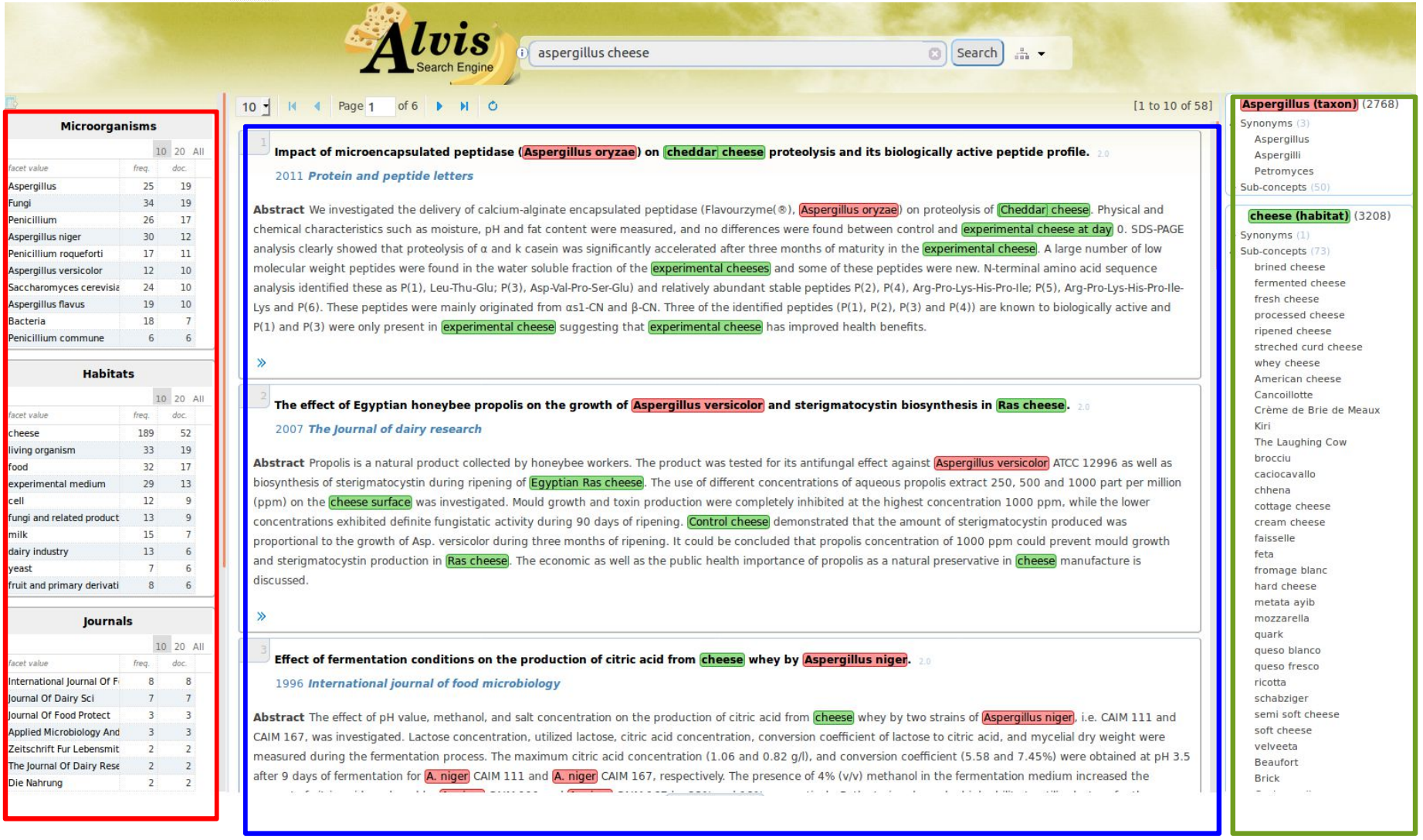
- American cheese
- Cancoillotte
- Crème de Brie de Meaux
- Kiri
- The Laughing Cow
- brocciu
- caciocavallo
- chhena
- cottage cheese
- cream cheese
- ...



http://bibliome.jouy.inra.fr/demo/food/alvisir/webapi/search
Mini-link: https://frama.link/AlvisFood

Has *Aspergillus* been isolated in cheese?

Results of the query: *aspergillus* cheese



Result of the query: *aspergillus* cheese

1 **Impact of microencapsulated peptidase (Aspergillus oryzae) on cheddar cheese proteolysis and its biologically active peptide profile.** 2.0

2011 ***Protein and peptide letters***

Abstract We investigated the delivery of calcium-alginate encapsulated peptidase (Flavourzyme(®), Aspergillus oryzae) on proteolysis of Cheddar cheese. Physical and chemical characteristics such as moisture, pH and fat content were measured, and no differences were found between control and experimental cheese at day 0. SDS-PAGE analysis clearly showed that proteolysis of α and k casein was significantly accelerated after three months of maturity in the experimental cheese. A large number of low molecular weight peptides were found in the water soluble fraction of the experimental cheeses and some of these peptides were new. N-terminal amino acid sequence analysis identified these as P(1), Leu-Thu-Glu; P(3), Asp-Val-Pro-Ser-Glu) and relatively abundant stable peptides P(2), P(4), Arg-Pro-Lys-His-Pro-Ile; P(5), Arg-Pro-Lys-His-Pro-Ile-Lys and P(6). These peptides were mainly originated from αs1-CN and β-CN. Three of the identified peptides (P(1), P(2), P(3) and P(4)) are known to biologically active and P(1) and P(3) were only present in experimental cheese suggesting that experimental cheese has improved health benefits.

»

2 **The effect of Egyptian honeybee propolis on the growth of Aspergillus versicolor and sterigmatocystin biosynthesis in Ras cheese.** 2.0

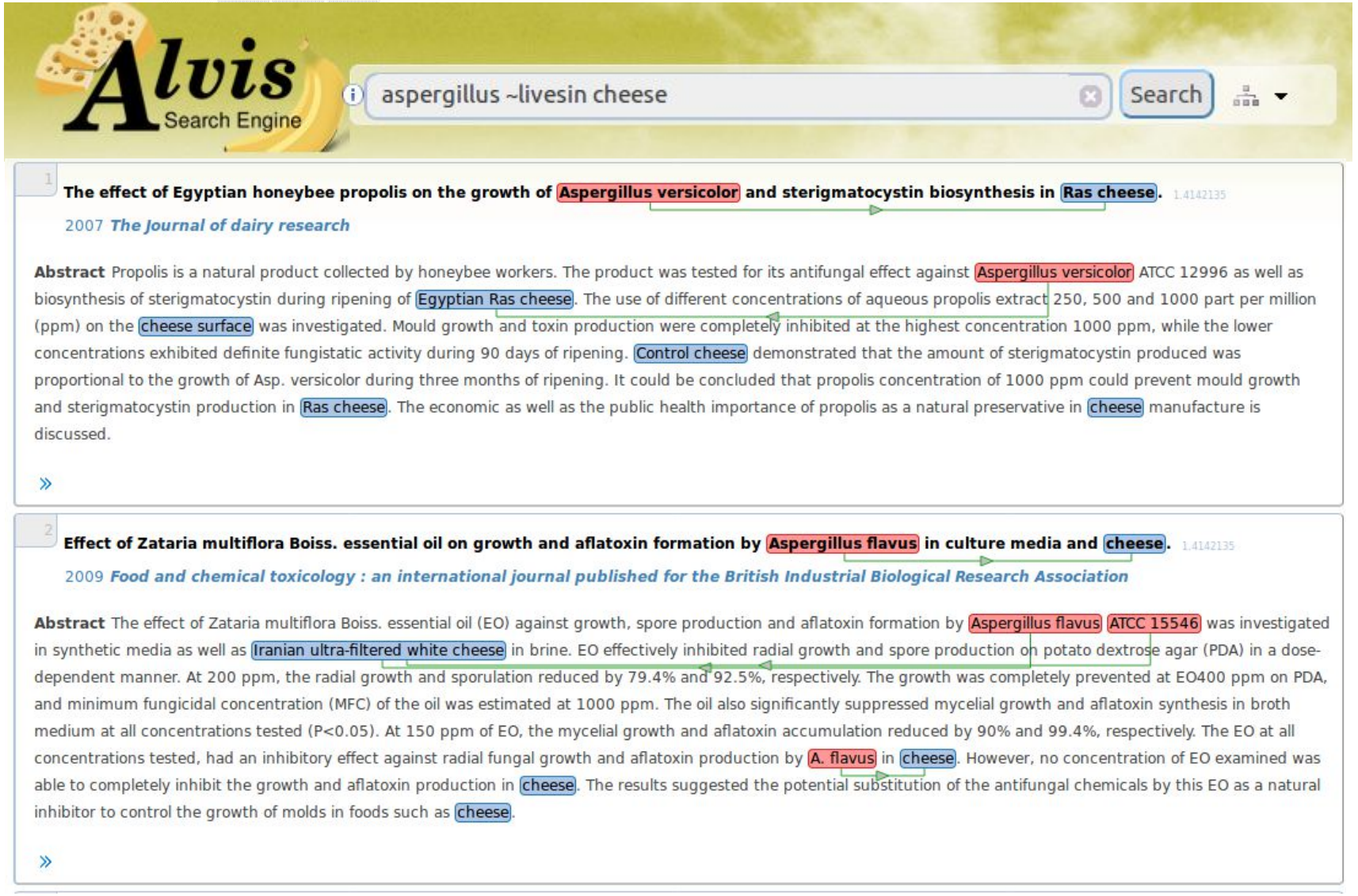
2007 ***The Journal of dairy research***

Abstract Propolis is a natural product collected by honeybee workers. The product was tested for its antifungal effect against Aspergillus versicolor ATCC 12996 as well as biosynthesis of sterigmatocystin during ripening of Egyptian Ras cheese. The use of different concentrations of aqueous propolis extract 250, 500 and 1000 part per million (ppm) on the cheese surface was investigated. Mould growth and toxin production were completely inhibited at the highest concentration 1000 ppm, while the lower concentrations exhibited definite fungistatic activity during 90 days of ripening. Control cheese demonstrated that the amount of sterigmatocystin produced was proportional to the growth of Asp. versicolor during three months of ripening. It could be concluded that propolis concentration of 1000 ppm could prevent mould growth and sterigmatocystin production in Ras cheese. The economic as well as the public health importance of propolis as a natural preservative in cheese manufacture is discussed.

»

Does *Aspergillus* lives in cheese ?

Result of the query: aspergillus ~livesin cheese



Behind the AlvisFood Search Engine

- Our approach is to extract from text
  - →"Microbe" and "Habitat" concepts
  - →Links between them

- We use
  - →AlvisNLP: Methods and tools for automatic extraction and analysis of biological text (*i.e.* Text Mining and Natural Language Processing)
  - →Machine learning methods trained with examples from microbiological and food domain experts
  - →Internal and external resources

- AlvisFood Search Engine: > 100,000 references from PubMed
  - →Selected by MeSH terms

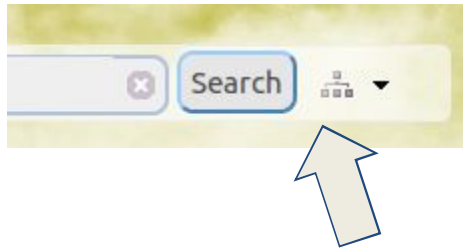
Microbial entity detection

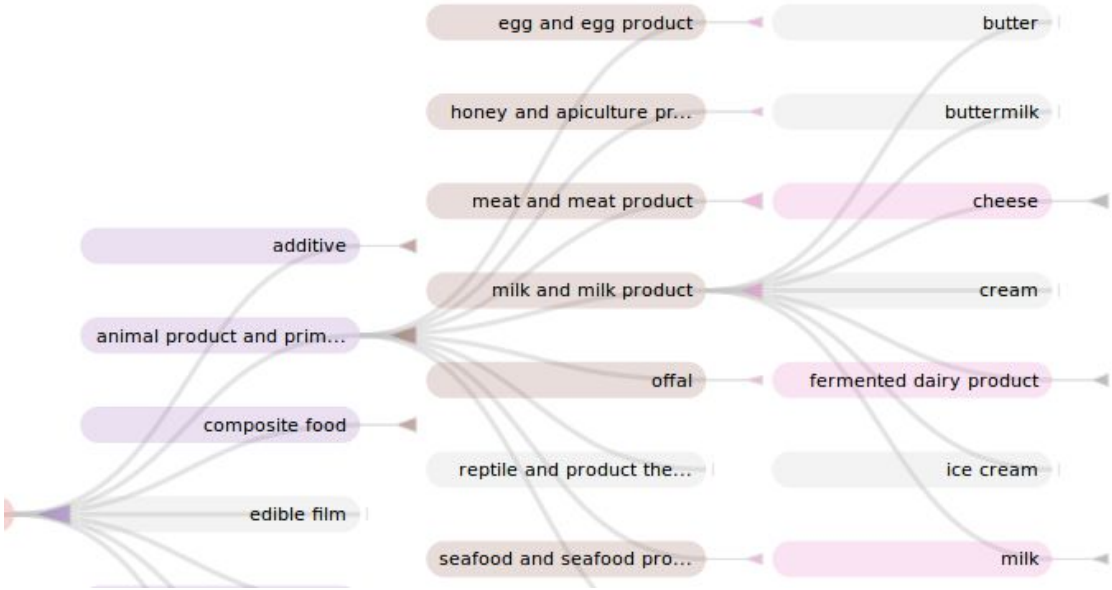
NCBI taxonomy

- Fungi (fungi) *Click on organism name to get more information.*
  - Blastocladiomycota
    - Blastocladiomycetes
      - Blastocladiales
      - Blastocladiales incertae sedis
    - environmental samples
      - uncultured Blastocladiomycota
  - Chytridiomycota
    - Chytridiomycetes
      - Chytridiales
      - Cladochytriales
      - Gromochytriales
      - Lobulomycetales
      - Mesochytriales
      - Polychytriales
      - Rhizophlyctidales
      - Rhizophydiales
      - Spizellomycetales
      - unclassified Chytridiomycetes
      - Chytridiomycetes incertae sedis
      - environmental samples
    - Monoblepharidomycetes
      - Monoblepharidales
      - unclassified Monoblepharidomycetes

Habitat entity detection

- Detection in text of nominal or adjectival groups
- Categorization of these groups with the *Ontobiotope ontology*
  - →Formal and structured representation of microbial habitats
  - →Partially reused in AlvisFoodSE

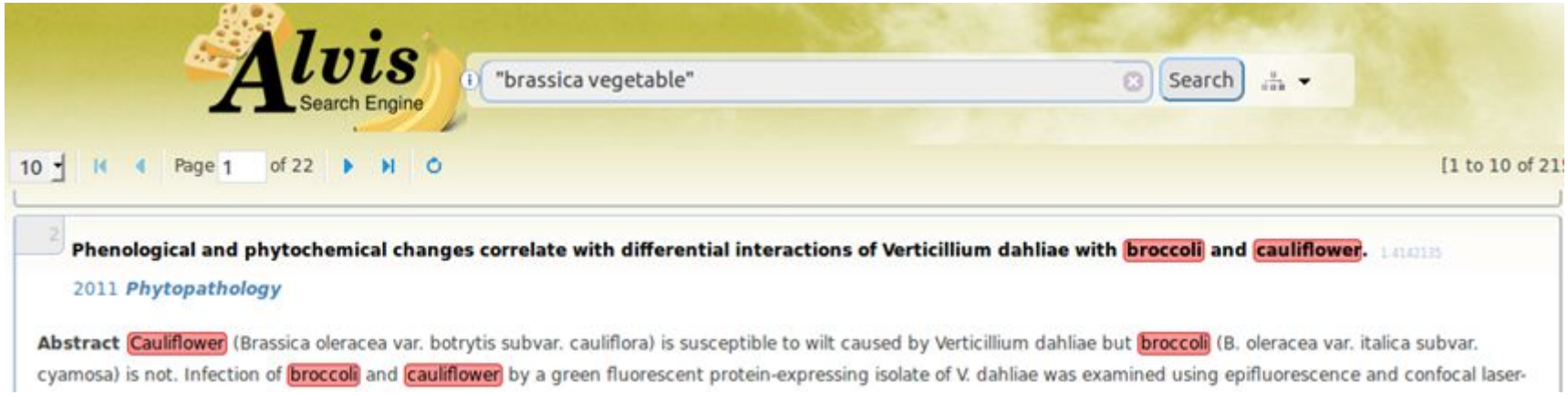




Food sub-categories of Ontobiotope ontology

- From the EFSA classification
- Enrichment by microbial and food domains experts

- Formal indication that “Roquefort” is a “Cheese”
  → allows semantic search
- Our automatic AlvisNLP tools link groups of words from the text to an Ontobiotope category
  → achieve normalisation

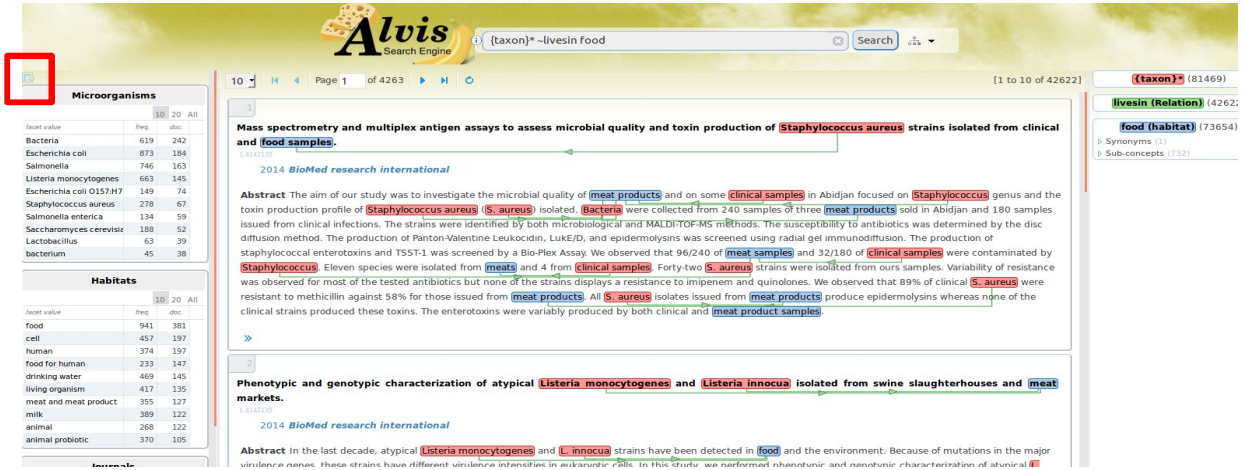


Relationship between Microbe and Habitat

- Extraction of ~livesin relationship
- Hard problems in automatic language processing and artificial intelligence
- Achieved by machine learning methods trained with annotated examples

What are the taxa living in food?
A query : {taxon}* ~livesin food

Results downloadable as table with occurrence counts.
Displayed as facets



To conclude

http://bibliome.jouy.inra.fr/demo/food/alvisir/webapi/search
Mini-link: https://frama.link/AlvisFood

- Our tools are pioneers in the field of text-mining for microbial biodiversity

- Bibliome is a research team so:
  - →If you use AlvisFoodSE for your research, please cite us
  - →If you see an error, please send us an email, this will help us to improve our tools

On going work

- Ambiguous cases for automatic tools

"Byssochlamys fulva and *Neosartorya fischeri* are heat-resistant fungi which are a concern to food industries"

- Automatic detection of microbial phenotypes

i.e. halophile, thermophile, phototroph ...

Acknowledgments

INRA Ontobiotope and Florilège working groups
Food Microbiome project

Thank you for your attention

Text mining tools for extracting information about microbial biodiversity in food

Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec

firstname.lastname@inra.fr

Bibliome Team
MaIAGE, INRA,
Université Paris-Saclay,
78350 Jouy-en-Josas, France

http://bibliome.jouy.inra.fr/demo/food/alvisir/webapi/search
Mini-link: https://frama.link/AlvisFood