



HAL
open science

Text mining tools for extracting information about microbial biodiversity in food

Estelle Chaix, Louise Deleger, Robert Bossy, Claire Nédellec

► **To cite this version:**

Estelle Chaix, Louise Deleger, Robert Bossy, Claire Nédellec. Text mining tools for extracting information about microbial biodiversity in food. Microbial Spoilers in Food 2017, Association pour le Développement de la Recherche Appliquée aux Industries Agricoles et Alimentaires (ADRIA). FRA., Jun 2017, Quimper, France. pp.95. hal-01602552

HAL Id: hal-01602552

<https://hal.science/hal-01602552>

Submitted on 2 Jun 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.

open
MIN
TED



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

Text mining tools for extracting information about microbial biodiversity in food

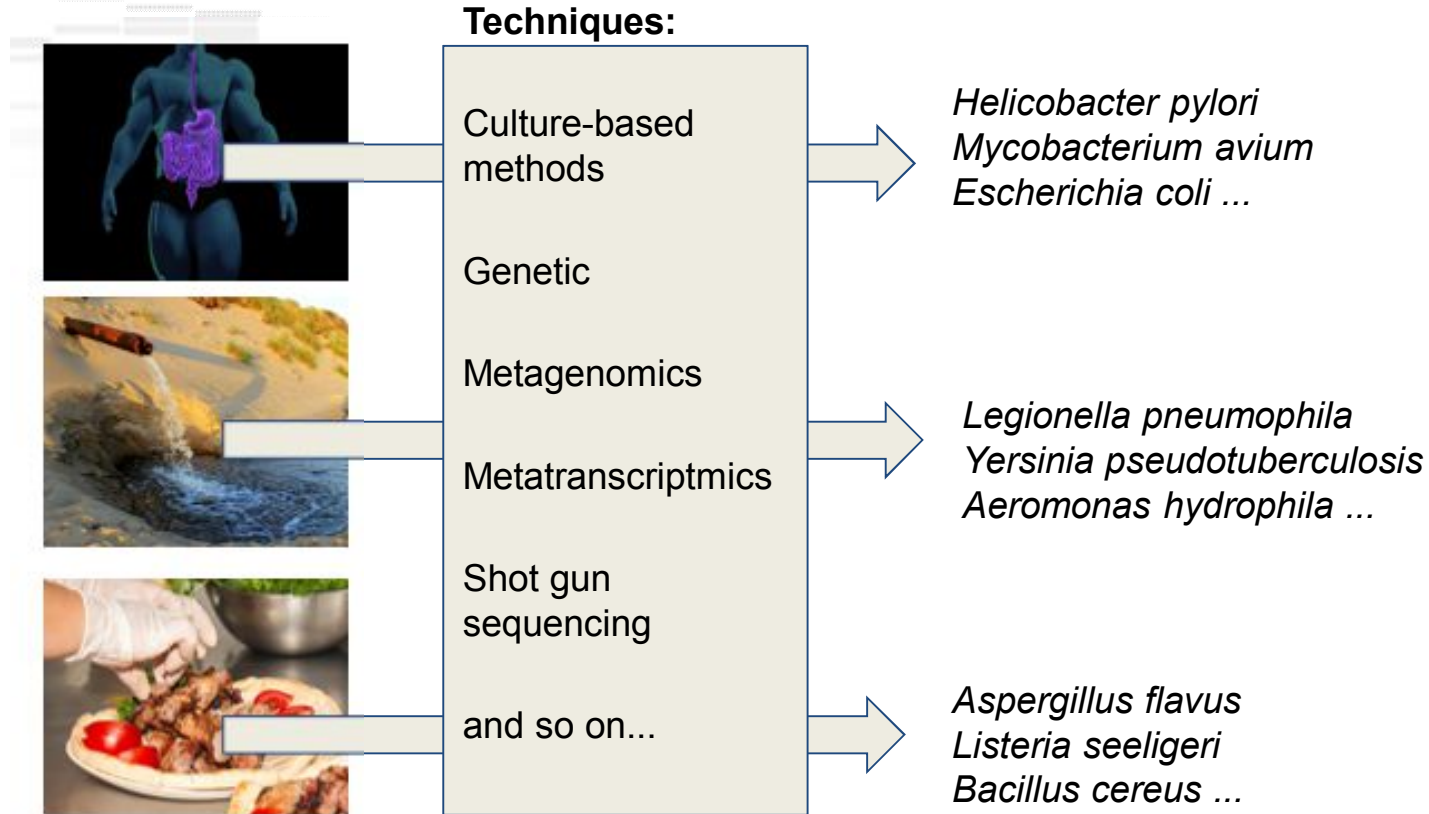
Estelle Chaix, Louise Deléger, Robert Bossy
and Claire Nédellec

firstname.lastname@inra.fr



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”



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*”



The screenshot shows the GOLD (Genomes Online Database) interface. At the top, there are navigation links: Home, Search, Distribution, Graphs, Biogeographical Metadata, Statistics, GOLD Usage Policy, Team, Help, News. A summary table on the left shows: Studies (28,817), Biosamples (23,362), Sequencing Projects (139,492), Analysis Projects (115,076), and Organisms (275,649). The main content area is titled 'Biosample Information' and shows details for a specific biosample:

BIOSAMPLE NAME	
GOLD Biosample ID	G00125061
Biosample Name	Microbial communities from thermized milk in Italy - ThermizedMik.RNAseq.exp1
Other Names	
Habitat (MICS-6)	Thermized milk
Community	microbial community
Location	Italy
Identifier	ThermizedMik.RNAseq.exp1
Add Date	2015-12-18
Last Modified By	Tatparthi Reddy on 2016-08-26

Below this, there are sections for 'BIOSAMPLE DESCRIPTION', 'BIOSAMPLE INFORMATION', 'BIOSAMPLE VIABILITY', 'GOLD CLASSIFICATION', and 'BIOSAMPLE COMPOSITION'. The 'BIOSAMPLE COMPOSITION' table shows: Number of Studies (1), Number of Seq Projects (1), Number of Analysis Projects (9), and Number of Related Biosamples (26).

e.g.

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



Text mining tools for extracting information about microbial biodiversity in food

Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec

Microbial spoilers in food 2017 - 28th - 30th June - Quimper



Habitat information is neither queryable nor comparable

Described at different levels of accuracy
and not standardized

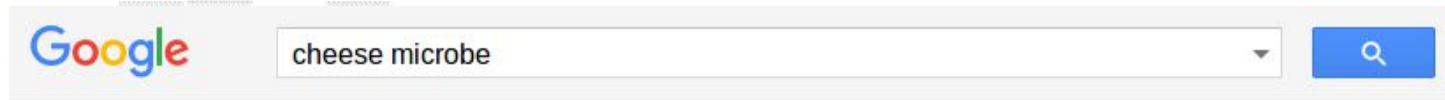
What is the cheese microflora?

“*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 10² to 10⁷ CFU g⁻¹ of food, which are equivalent to 10³–10⁸ 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.



Semantic search engine of microbial habitat in food

Interpretation of the query

Aspergillus :

Aspergillus (taxon)

▲ Synonyms (3)

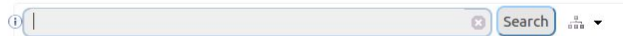
Aspergillus
Aspergilli
Petromyces

▲ 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

The screenshot shows the Alvis Search Engine interface with the query 'aspergillus cheese'. The results are displayed in a list of three articles, each with a title, journal information, and an abstract. The interface includes navigation controls, a search bar, and a sidebar with filters for Microorganisms, Habitats, and Journals.

Microorganisms

facet value	freq.	doc.
Aspergillus	25	19
Fungi	34	19
Penicillium	26	17
Aspergillus niger	30	12
Penicillium roqueforti	17	11
Aspergillus versicolor	12	10
Saccharomyces cerevisiae	24	10
Aspergillus flavus	19	10
Bacteria	18	7
Penicillium commune	6	6

Habitats

facet value	freq.	doc.
cheese	189	52
living organism	33	19
food	32	17
experimental medium	29	13
cell	12	9
fungi and related product	13	9
milk	15	7
dairy industry	13	6
yeast	7	6
fruit and primary derivati	8	6

Journals

facet value	freq.	doc.
International Journal Of F	8	8
Journal Of Dairy Sci	7	7
Journal Of Food Protect	3	3
Applied Microbiology And	3	3
Zeitschrift Fur Lebensmit	2	2
The Journal Of Dairy Rese	2	2
Die Nahrung	2	2

Search Results:

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 κ 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 α s1-CN and β -CN. Three of the identified peptides (P(1), P(2), P(3) and P(4)) are known to be 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.

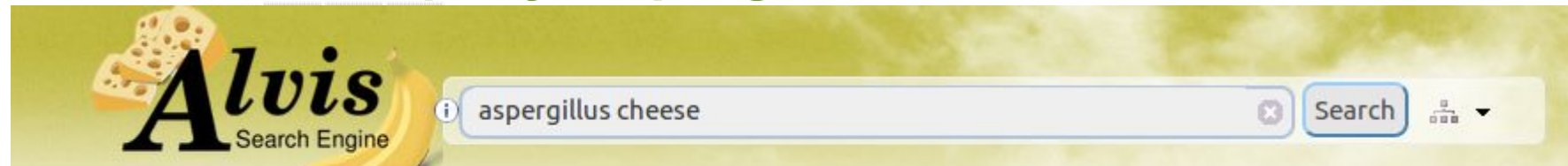
3 **Effect of fermentation conditions on the production of citric acid from cheese whey by *Aspergillus niger*.** 2.0
1996 *International journal of food microbiology*

Abstract The effect of pH value, methanol, and salt concentration on the production of citric acid from cheese whey by two strains of *Aspergillus niger*, i.e. CAIM 111 and CAIM 167, was investigated. Lactose concentration, utilized lactose, citric acid concentration, conversion coefficient of lactose to citric acid, and mycelial dry weight were measured during the fermentation process. The maximum citric acid concentration (1.06 and 0.82 g/l), and conversion coefficient (5.58 and 7.45%) were obtained at pH 3.5 after 9 days of fermentation for *A. niger* CAIM 111 and *A. niger* CAIM 167, respectively. The presence of 4% (v/v) methanol in the fermentation medium increased the

Aspergillus (taxon) (2768)
Synonyms (3)
Aspergillus
Aspergilli
Petromyces
Sub-concepts (50)

cheese (habitat) (3208)
Synonyms (1)
Sub-concepts (73)
brined cheese
fermented cheese
fresh cheese
processed cheese
ripened cheese
stretched curd cheese
whey cheese
American cheese
Cancoillotte
Crème de Brie de Meaux
Kiri
The Laughing Cow
brocciu
cacioavallu
chhena
cottage cheese
cream cheese
faiselle
feta
fromage blanc
hard cheese
metata ayib
mozzarella
quark
queso blanco
queso fresco
ricotta
schabziger
semi soft cheese
soft cheese
velveeta
Beaufort
Brick

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 κ 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 α s1-CN and β -CN. Three of the identified peptides (P(1), P(2), P(3) and P(4)) are known to be 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 ?



aspergillus ~livesin cheese

Search



Result of the query: aspergillus ~livesin cheese



Alvis Search Engine

aspergillus ~livesin cheese Search

1 **The effect of Egyptian honeybee propolis on the growth of *Aspergillus versicolor* and sterigmatocystin biosynthesis in Ras cheese.** 1.4142135
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.

2 **Effect of *Zataria multiflora* Boiss. essential oil on growth and aflatoxin formation by *Aspergillus flavus* in culture media and cheese.** 1.4142135
2009 *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*

Abstract The effect of *Zataria multiflora* Boiss. essential oil (EO) against growth, spore production and aflatoxin formation by *Aspergillus flavus* ATCC 15546 was investigated in synthetic media as well as *Iranian ultra-filtered white cheese* in brine. EO effectively inhibited radial growth and spore production on potato dextrose agar (PDA) in a dose-dependent manner. At 200 ppm, the radial growth and sporulation reduced by 79.4% and 92.5%, respectively. The growth was completely prevented at EO400 ppm on PDA, and minimum fungicidal concentration (MFC) of the oil was estimated at 1000 ppm. The oil also significantly suppressed mycelial growth and aflatoxin synthesis in broth medium at all concentrations tested (P<0.05). At 150 ppm of EO, the mycelial growth and aflatoxin accumulation reduced by 90% and 99.4%, respectively. The EO at all concentrations tested, had an inhibitory effect against radial fungal growth and aflatoxin production by *A. flavus* in *cheese*. However, no concentration of EO examined was able to completely inhibit the growth and aflatoxin production in *cheese*. The results suggested the potential substitution of the antifungal chemicals by this EO as a natural inhibitor to control the growth of molds in foods such as *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

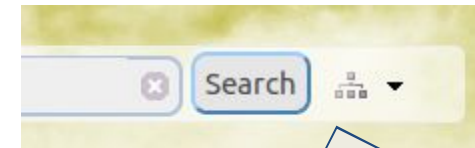
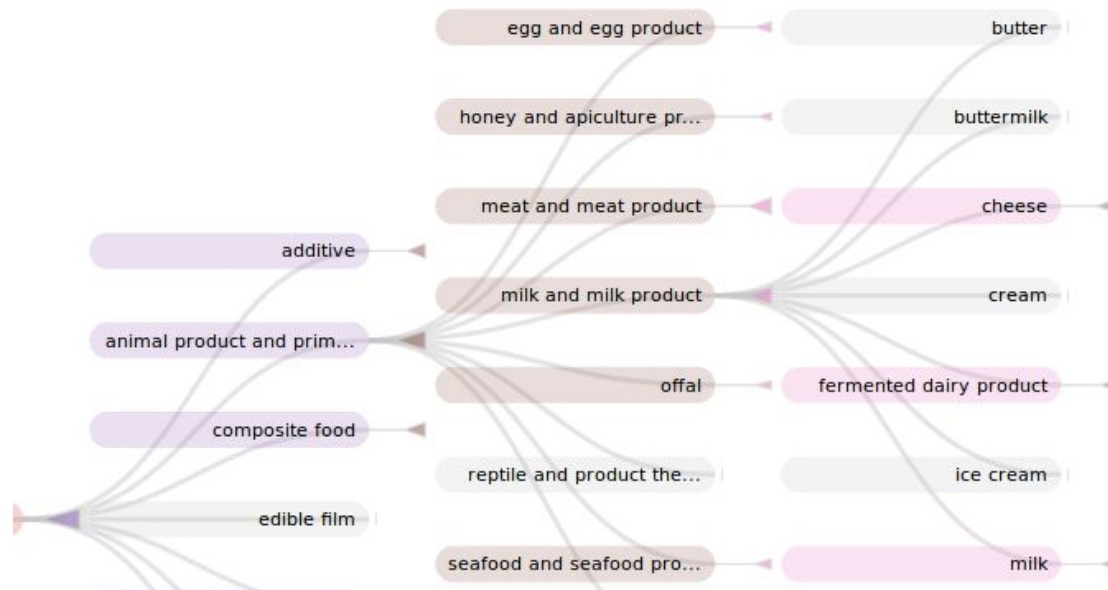
Microbial entity detection

NCBI taxonomy

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

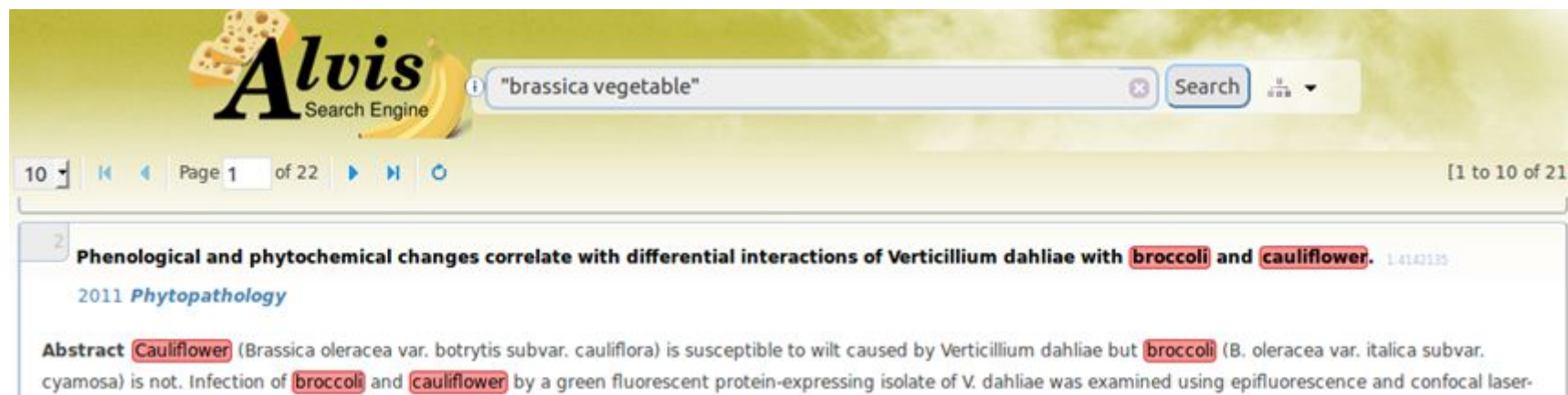
Habitat entity detection

- 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



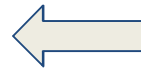
The screenshot shows the Alvis Search Engine interface. The search bar contains the text "brassica vegetable". Below the search bar, there are navigation controls including "Page 1 of 22" and "1 to 10 of 21". The search results display a snippet from a 2011 *Phytopathology* article. The snippet title is "Phenological and phytochemical changes correlate with differential interactions of *Verticillium dahliae* with broccoli and cauliflower." The abstract text is: "Abstract Cauliflower (*Brassica oleracea* var. botrytis subvar. cauliflora) is susceptible to wilt caused by *Verticillium dahliae* but broccoli (*B. oleracea* var. italica subvar. *cyamosa*) is not. Infection of broccoli and cauliflower by a green fluorescent protein-expressing isolate of *V. dahliae* was examined using epifluorescence and confocal laser-

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



Microorganisms

facet value	freq	doc
Bacteria	619	242
Escherichia coli	873	184
Salmonella	746	163
Listeria monocytogenes	663	145
Escherichia coli O157:H7	149	74
Staphylococcus aureus	278	67
Salmonella enterica	134	59
Saccharomyces cerevisiae	188	52
Lactobacillus	63	39
bacterium	45	38

Habitats

facet value	freq	doc
food	941	381
cell	457	197
human	374	197
food for human	233	147
drinking water	469	145
living organism	417	135
meat and meat product	355	127
milk	389	122
animal	269	122
animal probiotic	370	105

Alvis Search Engine {taxon}* ~livesin food Search

Page 1 of 4263

Mass spectrometry and multiplex antigen assays to assess microbial quality and toxin production of *Staphylococcus aureus* strains isolated from clinical and food samples.

2014 *BioMed research international*

Abstract The aim of our study was to investigate the microbial quality of **meat products** and on some **clinical samples** in Abidjan focused on **Staphylococcus** genus and the toxin production profile of **Staphylococcus aureus** (**S. aureus**) isolated. **Bacteria** were collected from 240 samples of three **meat products** sold in Abidjan and 180 samples issued from clinical infections. The strains were identified by both microbiological and MALDI-TOF-MS methods. The susceptibility to antibiotics was determined by the disc diffusion method. The production of Panton-Valentine Leukocidin, LukE/D, and epidermolytins was screened using radial gel immunodiffusion. The production of staphylococcal enterotoxins and TSST1 was screened by a Bio-Plex Assay. We observed that 96/240 of **meat samples** and 32/180 of **clinical samples** were contaminated by **Staphylococcus**. Eleven species were isolated from **meats** and 4 from **clinical samples**. Forty-two **S. aureus** strains were isolated from our samples. Variability of resistance was observed for most of the tested antibiotics but none of the strains displays a resistance to imipenem and quinolones. We observed that 89% of clinical **S. aureus** were resistant to methicillin against 58% for those issued from **meat products**. All **S. aureus** isolates issued from **meat products** produce epidermolytins whereas none of the clinical strains produced these toxins. The enterotoxins were variably produced by both clinical and **meat product samples**.

Phenotypic and genotypic characterization of atypical *Listeria monocytogenes* and *Listeria innocua* isolated from swine slaughterhouses and meat markets.

2014 *BioMed research international*

Abstract In the last decade, atypical **Listeria monocytogenes** and **L. innocua** strains have been detected in **food** and the environment. Because of mutations in the major virulence genes, these strains have different virulence intensities in eukaryotic cells. In this study, we performed phenotypic and genotypic characterization of atypical

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

open
MIN
TED

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

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

firstname.lastname@inra.fr

<http://bibliome.jouy.inra.fr/demo/food/alvisir/webapi/search>

Mini-link: <https://frama.link/AlvisFood>

