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?

Techniques:

- Culture-based methods
- Genetic
- Metagenomics
- Metatranscriptomics
- Shot gun sequencing
- and so on...

Helicobacter pylori
Mycobacterium avium
Escherichia coli ...

Legionella pneumophila
Yersinia pseudotuberculosis
Aeromonas hydrophila ...

Aspergillus flavus
Listeria seeligeri
Bacillus cereus ...

Properties of environment? Microbial Interaction?
Crossing between microbial species and habitats

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

“Bacteria of the genera Enterococcus and Lactobacillus and coliform bacteria were isolated from Dutch-type semi-hard cheese.”

“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

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
- brocchi
- 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

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

Abstract: We investigated the delivery of calcium-alginate encapsulated peptidase (Flavourzyme 8M). 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 β caseins 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 cheese, and some of those peptides were new. N-terminal amino acid sequence analysis identified these as P11, Leu-Thr-Glu; P13, Asp-Val-Pro-Ser-Glu) and relatively abundant stable peptides P12, P14, Arg-Pro-Lys-His-Pro-Lys; P15, Arg-Pro-Lys-His-Pro-Ile-Lys and P16. These peptides were mainly originated from α1-CN and β-CN. Three of the identified peptides (P11, P12, P13) and P14 are known to biologically active and P11 and P12 were only present in experimental cheese suggesting that experimental cheese has improved health benefits.

The effect of Egyptian honeybee propolis on the growth of Aspergillus versicolor and sterigmatocystin biosynthesis in Kas cheese.

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 Kas cheese. The use of different concentrations of aqueous propolis extract 250; 500 and 1000 part per million (ppm) on the cheese surface was investigated. Mold 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 mold growth and sterigmatocystin production in Kas cheese. The economic as well as the public health importance of propolis as a natural preservative in cheese manufacture is discussed.

Effect of fermentation conditions on the production of citric acid from Cheddar whey by Aspergillus niger.

Abstract: The effect of pH value, methanol, and salt concentration on the production of citric acid from Cheddar 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
Impact of microencapsulated peptidase (Aspergillus oryzae) on cheddar cheese proteolysis and its biologically active peptide profile.

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-Thr-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.

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

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

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

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 fungicidal 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.

Effect of Zataria multiflora Boiss. essential oil on growth and aflatoxin formation by Aspergillus flavus in culture media and cheese.

Abstract: The effect of Zataria multiflora Boiss. essential oil (EO) against growth, spore production and aflatoxin formation by Aspergillus flavus ATCC 15540 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.

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
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**
      - **Rhizophyctidales**
      - **Rhizophycales**
      - **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
Text mining tools for extracting information about microbial biodiversity in food

Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec

firstname.lastname@inra.fr

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