



Finanziato dall'Unione europea NextGenerationsEU

Ministero dell'Università e della Ricerca

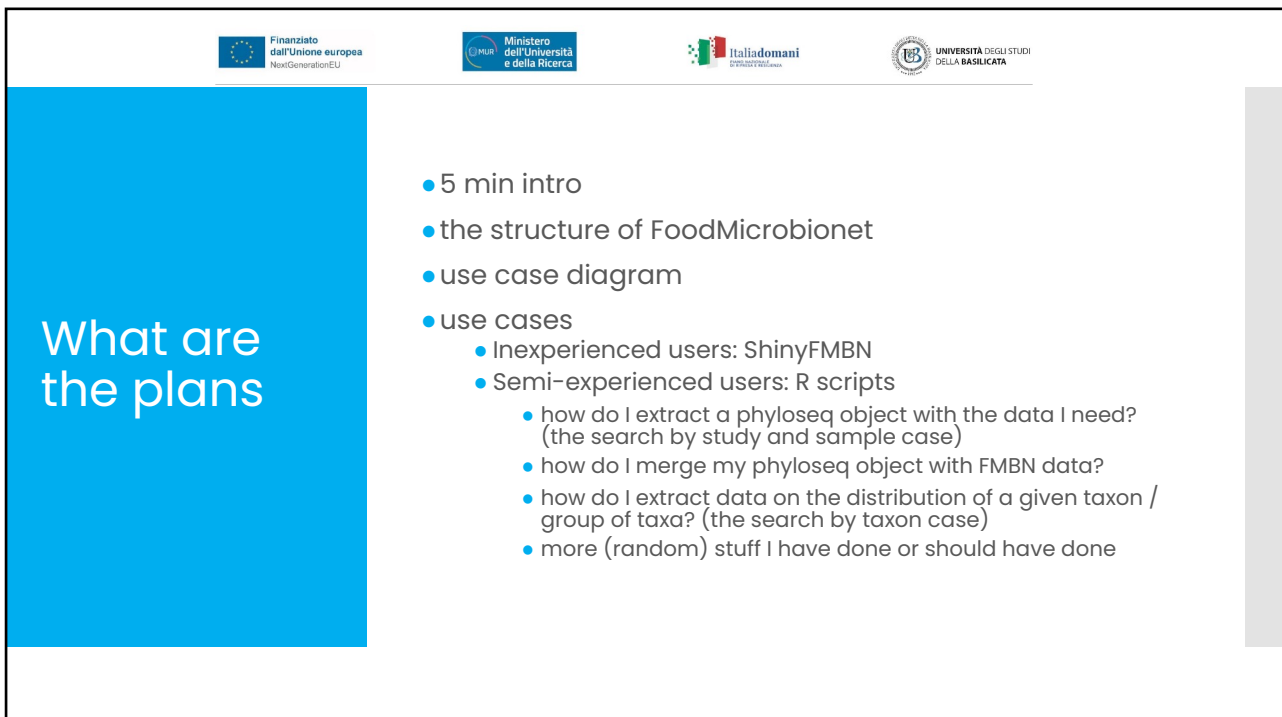
Italiadomani

UNIVERSITÀ DEGLI STUDI DELLA BASILICATA

Making the most of FoodMicrobionet 5.0

Prof. Eugenio Parente
ovvero: come usare FMBN senza troppa fatica

1



Finanziato dall'Unione europea NextGenerationsEU

Ministero dell'Università e della Ricerca

Italiadomani

UNIVERSITÀ DEGLI STUDI DELLA BASILICATA

What are the plans

- 5 min intro
- the structure of FoodMicrobionet
- use case diagram
- use cases
 - Inexperienced users: ShinyFMBN
 - Semi-experienced users: R scripts
 - how do I extract a phyloseq object with the data I need? (the search by study and sample case)
 - how do I merge my phyloseq object with FMBN data?
 - how do I extract data on the distribution of a given taxon / group of taxa? (the search by taxon case)
 - more (random) stuff I have done or should have done

2

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani

UNIVERSITÀ DEGLI STUDI DELLA BASILICATA

Why FoodMicrobionet

3

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani

UNIVERSITÀ DEGLI STUDI DELLA BASILICATA


The exponential growth of metataxonomic studies


The growth of metataxonomic studies on food microbial communities


The chart illustrates the exponential growth of metataxonomic studies on food microbial communities from 2009 to 2021. The y-axis represents the cumulative number of papers, ranging from 0 to 600. The x-axis represents the year, from 2009 to 2021. The studies are categorized by food group: cereal (light blue), dairy (dark blue), fruit (light green), meat (dark green), other (pink), seafood (red), and vegetable (orange). The total number of studies shows a significant upward trend, reaching over 600 by 2021.


Year	vegetable	seafood	other	meat	fruit	dairy	cereal
2009	0	0	0	0	0	0	0
2011	10	5	5	5	5	5	5
2013	20	10	10	10	10	10	10
2015	30	20	20	20	20	20	20
2017	40	30	30	30	30	30	30
2019	50	40	40	40	40	40	40
2021	60	50	50	50	50	50	50

4









The search for meta-taxonomic data: mines or dumpsters (1)

- **data availability:** raw sequences have been deposited in publicly available repositories only for a small percentage of studies (<30%)
- **heterogeneity:** there is a wide variety of experimental approaches in the wet- and dry-laboratory approaches, and in the bioinformatic approaches used to process data
- **size:** the number of samples is typically low (dairy foods: median number of samples 24, range 1-1674), but steadily increasing; sequence length highly variable
- **design:** most studies are purely descriptive, even for longitudinal studies there are generally severe issues with experimental design (lack of replication and randomization, issues with sampling...)
- **annotation:** lack of detailed annotation for food studies (improving with the enforcement of MIMARKS standards)

5











The search for meta-taxonomic data: mines or dumpsters (2)






6

Existing repositories


Repo	Advantages	Disadvantages	
IMNGS	The Integrated Microbial NGS platform automatically retrieve and processes 16S amplicons from public databases; can be queried using sequences or taxonomic queries (down to family); <u>standardized, automated pipelines used for processing</u>	Uses OTUs and RDP as a database, metadata not optimized for food classification	
MGnify	Large (5004 public studies, 597736 public samples); available on line; data obtained from ENA or submitted by users; <u>standardized, automated pipelines used for processing</u>	2,919 samples on food biomes; metadata not optimized for food classification; limited number of formats for export of analysis results	


7


Existing repositories


Repo	Advantages	Disadvantages	
QIITA	Large (782 public studies, 393,388 public samples); new studies added as in MGnify, but users can carry out their own analyses and meta studies; raw data can be reprocessed as improved <u>pipelines</u> become available; <u>standardized, automated pipelines used for processing</u>	Very few studies/samples on food biomes available publicly, metadata not optimized for food classification	
FMBN	Specialized (food bacterial and fungal biomes, 251 studies, 14,035 samples with data on bacteria and 1,114 samples with data on fungi), sample metadata structure optimized for foods, can export data in a several formats, ready for further analysis; can combine data from different studies; connected to several other databases (NCBI SRA, LPSN, NCBI Taxonomy, Omnicrobe)	"pragmatic" approach with a single pipeline used to analyse most data in current version; only available offline; long time between updates; limited to foods	

8




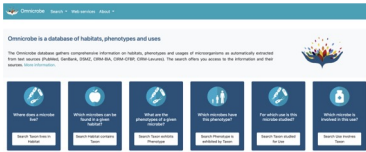





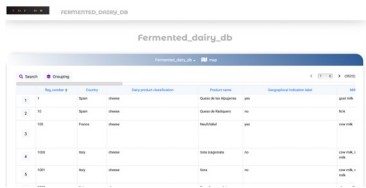


Database based on text mining or curated searches

Omnicrobe

Fermented Dairy Products Database

9









More stuff

dbBact




Microbiome atlas




10

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani

UNIVERSITÀ DEGLI STUDI DELLA BASILICATA

Understanding FMBN

- As a researcher dealing with or looking for information on metataxonomic studies, what sort of info I need?
 - study metadata**
 - sample metadata**
 - taxa metadata**
 - taxa abundance in samples**
 - any other info which may help in searches and inclusion/exclusion criteria**

11

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani

UNIVERSITÀ DEGLI STUDI DELLA BASILICATA

Understanding FMBN


Table structure version 5


```


    graph TD
      ET["Edge table (B & F)  
• TaxonId  
• SampleId  
• Weight (Abundance)"]
      ST["Sample table (B & F)  
• SampleId  
• FoodEx2 FoodId  
• StudyId  
• BioSample  
• other sample metadata"]
      SUT["Study table  
• StudyId  
• BioProject  
• SRA Study Accn  
• DOI  
• Primers, Region  
• Version  
• other study metadata"]
      AT["Abstracts table  
• StudyId  
• DOI link  
• Abstract"]
      TT["Taxa table  
• TaxonId  
• Lineage (Domain - Species)  
• External links"]
      TD["Taxonomic databases"]
      EFC["EFSA FoodEx2 classification"]
      NSRA["NCBI Sequence Read Archive"]
      PO["Phyloseq objects"]
      SL["Scientific literature"]
      PT["Primers table"]
      VT["Version table"]


      ET --> TT
      ET --> TD
      ST --> TT
      ST --> EFC
      ST --> NSRA
      SUT --> TT
      SUT --> PO
      SUT --> SL
      SUT --> PT
      SUT --> VT
      AT --> TT
      AT --> SL
      AT --> VT
  
```

12















FMBN resources


Resource	Used for	QR code
GitHub repository	Storing and sharing data, technical documents and scripts, collaborating on scripts, flagging and discussing issues... (use https://download-directory.github.io/ to download)	
Mendeley data	Shiny app and manual (but also FMBN); citable using DOI	
Zenodo	Mindata files (R lists with the ASVs, divided by study); citable using DOI	

13

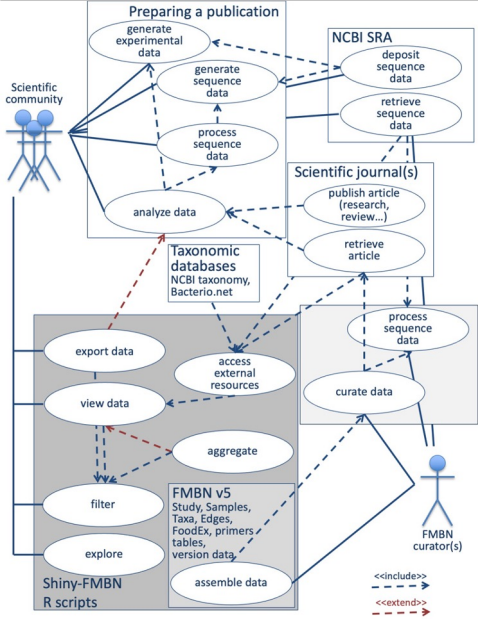












FMBN Use case diagram







14

Application scenarios

- 1. I want to have a look at the microbiology of a given set of food products and therefore I need to extract data from FMBN
 - perform searches and optional aggregation using the ShinyFMBN app
 - perform more complex searches and create a phyloseq object for further processing using extract_studies_v2.R (can be automated)
- 2. I want to analyze the data sets I have extracted in 1. above
 - use phyloseq objects and any of the many R packages which do the job
 - use the agg.Rdata file obtained from ShinyFMBN with FMBNanalyzer
- 3. I want to look at the distribution of a specific taxon in foods or food environments
 - use one of the many scripts available here: <https://github.com/ep142/FoodMicrobionet/tree/master/WIMB>

15

Finanziamenti e ringraziamenti

- updating of data and software related to FoodMicrobionet is currently funded by P20229JMMH «**Mining the biodiversity of non conventional yeasts as bioresources for innovative fermented beverages through a genomics and bioinformatics approach – NCYdiversity**» funded within the PRIN 2022 PNRR program, Research Unit Prof. Eugenio Parente, PI Prof. Nicola Vitulo
- some collateral and specialized activities are funded on other measures

16

Copyright / Copyleft /License

- for the code and FMBN documents see the the relevant script or document (usually code under MIT license)
- this presentation CC BY 4.0