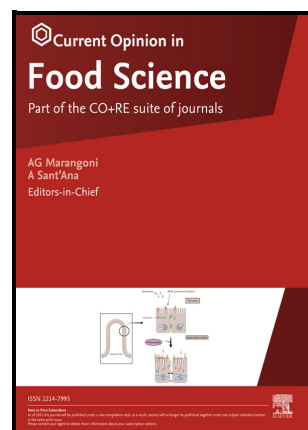


The hidden diversity of food microbiomes
Title: The hidden diversity of food microbiomes

Edoardo Pasolli, Danilo Ercolini



PII: S2214-7993(25)00048-7

DOI: <https://doi.org/10.1016/j.cofs.2025.101318>

Reference: COFS101318

To appear in: *Current Opinion in Food Science*

Received date: 17 February 2025

Revised date: 9 April 2025

Accepted date: 27 May 2025

Please cite this article as: Edoardo Pasolli and Danilo Ercolini, The hidden diversity of food microbiomes
Short Title: The hidden diversity of food microbiomes, *Current Opinion in Food Science*, (2025)
doi:<https://doi.org/10.1016/j.cofs.2025.101318>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 The Author(s). Published by Elsevier Ltd.

The hidden diversity of food microbiomes

Short Title: The hidden diversity of food microbiomes

Edoardo Pasolli ^{1,2*}, and Danilo Ercolini ^{1,2*}

¹ Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy

² Task Force on Microbiome Studies, University of Naples Federico II, Portici, Italy

edoardo.pasolli@unina.it

danilo.ercolini@unina.it

*Corresponding author at: Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy

Abstract

The microbial organisms living in food and food systems have a key role in food processing, quality, safety, and even in human health promotion. This review aims to summarize recent advancements in the characterization of novel species from metagenomics data, with a focus on their relevance for food microbiology. We discuss the generation of metagenome-assembled genomes, a crucial step that facilitates comparative genomic analyses and development of databases for taxonomic profiling and functional potential characterization. We also highlight how these approaches enable new applications in food science and examine the broader implications of these findings to better understand the interactions between food and human health. Finally, we present key challenges that need to be addressed to fully characterize the hidden diversity of food microbiomes and to exploit their potential for innovation in food science and health-related applications.

Keywords

Food microbiome; Metagenomics; Metagenomic assembly; Large-scale analysis; Unexplored microbial diversity

Introduction

In the last few decades, the food microbiome has received great attention in both research and industry due to its fundamental role in food processing and production. The study of microbial communities is crucial for example to elucidating fermentation processes, but it is also relevant for addressing issues related to food spoilage and safety [1], which emphasizes their importance in food quality but also their potential connections with human health. The characterization of the food microbiomes is not trivial due to the complexity of food matrices and is further complicated by their interactions with the environment.

The microbial communities of food have been traditionally studied through methods based on cultivation, limited by scalability and focus only on cultivable microorganisms. Advancements have been achieved thanks

to the introduction of next-generation sequencing (NGS) technologies, with both targeted and untargeted approaches widely adopted [2]. (Shotgun) metagenomics represents a state-of-the-art technology that has enabled the in-depth taxonomic investigation of microbial communities along with the definition of their potential functions. This allows the detection of uncultivated microbial species, the study of the hidden microbial diversity, and the exploration of its potential activities in a specific environment. However, advancements using this technology have been limited by the lack of appropriate computational tools for analysis of data. This is further complicated by the influence of environmental, cultural, and other hidden factors on food microbiomes, which highlights the necessity to perform integrative analyses to have a comprehensive characterization of these microbial communities.

In this review, we aim at discussing recent advancements in the characterization of food microbial communities, with emphasis on the identification and study of unexplored microbial species, a key factor to advance food microbiome research.

The food microbiome: Current understanding and gaps

The investigation of microbiomes in food systems is challenging due to the large diversity of food matrices, each characterized by specific, often unique, processing, and ecological factors. For example, fermented foods (e.g., yogurt, kimchi, sauerkraut) [3] are associated with quite dynamic communities, in which fermentation brings peculiar characteristics in terms of flavors and textures; however, the mechanisms of these dynamics are not fully explored. Also, the species present in the dairy can be linked to specific characteristics of the final product, with the influence of environmental and other processing conditions factors still less clear. The microbiome composition in fresh products like vegetables and fruits is typically modulated by agricultural practices, with implications in terms of food quality and safety still difficult to characterize [4]. Meat products are often subjected to microbial spoilage, with the proper characterization of their dynamics essential in terms of shelf life and for safety reasons [5].

Understanding the interactions among strains within food matrices is important for both food quality and safety. To this regard, several aspects such as production of secondary metabolites, horizontal gene transfer [6], and antimicrobial resistance (AMR) dynamics [7] deserve further investigations. Such characterizations are complicated by the influence of specific practices (e.g., the use of particular starters or production processes) on food microbiomes, an aspect even more relevant for traditional and artisanal products. Therefore, providing a comprehensive characterization of these microbial communities is challenging due to these specificities, an issue that becomes even more relevant when considering less characterized food types (e.g., local and novel foods), low abundance species, and non-bacterial organisms (e.g., fungi and viruses). Therefore, large-scale analyses able to incorporate sufficient microbial diversity along with standardized methodologies and proper advancements in computational tools are needed to unravel the complexities of these ecosystems.

Advances in metagenomics for food microbiome characterization

Developments in metagenomics have impacted tremendously on the food microbiome field and allowed researchers to conduct comprehensive analyses across complex food matrices. Short-read platforms are commonly used due to their accuracy and growing cost-effectiveness, although they can struggle to resolve complex microbial communities, low abundance species, and delineation of similar strains. For this reason, third generation long-read sequencing technologies are becoming more popular as a complementary tool to ease some tasks such as generation of complete genomes and improved taxonomic resolution [8].

The typical metagenomic workflow (**Figure 1**) [9] starts with experimental design, a step sometimes overlooked. Key considerations include defining in advance the number of samples and relevant characteristics to ensure robustness of the results and sufficient statistical power. This is followed by DNA extraction, which may be challenging for particular food matrices and their associated environments [10], and sequencing. While 16S rRNA sequencing is still valuable due to reduced costs and computational efforts, shotgun methodologies enable detailed taxonomic resolution and functional investigations. Two main computational approaches can be employed. Assembly-based approaches allow researchers to reconstruct metagenome-assemble genomes (MAGs) directly from raw data [11]. Their typically *de-novo* nature enables the identification and characterization of novel species, with the outcome that is however more impacted for low abundance taxa, when multiple strains coexist, and when the sample diversity is higher. Such MAGs, which can be potentially improved thanks to hybrid approaches involving also long-read sequences [12], represent key elements for the definition of reference databases, which can be integrated with genomes from isolated sources. However, such resources are far from being comprehensive due to the high variability across food types and therefore are biased toward more common food types and microbial species, a factor that is even more pronounced when dealing with functional annotation. These databases are also fundamental to enable mapping-based methodologies useful to perform sensitive species-level taxonomic and functional profiling. In this context, their completeness is crucial and limitations in reference databases can compromise the accurate characterization of novel and less-represented species, an issue that is even more problematic for non-bacterial components. Additionally, several methodologies are available to perform strain-level profiling, with initial efforts based on 16S rRNA [13] later extended to metagenomics data, enabling more accurate strain-level characterization for pangenome and strain-tracking analyses [14]. Finally, these data products are subjected to biostatistical analyses and proper interpretation of the results through a multi-disciplinary perspective aiming at linking microbial composition and functions to food system characteristics. Such data may also benefit from integration with other multi-omics data types, and in which predictive modeling based on machine learning methodologies is gaining growing interest [15].

Discovery and characterization of novel microbial species in food

Metagenomics has been fundamental in advancing our understanding of the microbial diversity in food including redefinition of the microbial taxonomy and identification of novel species. Taxonomy has been traditionally based on genomes retrieved from isolated sources, which can be suboptimal due to the underrepresentation of more rare or difficult to isolate species. Therefore, integration of reference genomes with information coming from metagenomes has been beneficial to detect new clades (e.g., genera and families) and resolve ambiguities, important steps to provide more accurate analyses in food environments.

Several metagenomic studies have been conducted across different food matrices (**Table 1**), and some of them have also led to the identification of novel microbial species by employing assembly-based computational approaches. The greatest interest was devoted to dairy products. For example, in cheese microbiomes there was the identification of 47 putative new species, potentially influencing taste and color [16]; similarly, four uncharacterized species that may affect flavor or quality were detected in fermented milk [17] and 63 novel species were discovered in a larger effort involving different dairy products [18]. Another analysis characterized dairy- as well sugar- and brine-type fermented foods and identified ten putative novel species associated potentially with health promoting attributes [19]. Also plant-based fermented products are receiving growing interest due to their possible contribution in developing novel foods, and new species were for example detected in miso [20] and in different fermented vegetables [21].

While such studies have focused on characterizing the microbial diversity of specific food matrices and types, larger investigations are also needed to have a broader understanding of the microbial composition across food types. These tasks can be accomplished through the development of curated, global databases that can ease access, re-use, and visualization of public available datasets. More efforts have been already conducted on amplicon data, and remarkably cases are represented by the FoodMicrobionet [22] and FoodMicroDB [23] resources, which aggregated thousands of samples along with metadata information. In the metagenomic context, along with broader initiatives that span multiple environments (e.g., the MGnify resource [24]), only more recently attempts have been devoted specifically to food metagenomes, such as the curatedFoodMetagenomicData (cFMD) [25] and MiFoDB [26] resources.

More specifically, the first release of cFMD integrated 2,533 food metagenomes spanning 59 independent datasets. Metadata were standardized, curated, and organized into 27 fields covering sample, food-related, and technical information. To this regard, the metadata were standardized based on previous work [27] and specifically refined by food microbiologists in the study, however the definition of more established food ontologies still deserves more attention in the field. Despite the metagenomes being acquired from 50 countries and 15 food categories, they belonged mainly (85 %) to dairy, fermented beverage, and fermented meat categories, and especially sampled from industrialized populations. This highlights the still tremendous under sampling of the great diversity of foods produced and consumed worldwide, especially in less developed countries [28], an issue that mimics what already assessed in the human microbiome field [29]. The

metagenomes collected in cFMD generated 10,899 MAGs that spanned 1,036 prokaryotic and 108 eukaryotic species-level genome bins (SGBs) [30]. Half of these prokaryotic SGBs represented unknown SGBs (uSGBs), widespread across all food categories and dozens of food types. Interestingly, the least characterized types were quite peculiar foods such as Mexican pulque, African palm wine, and Korean skate fish, which highlighted the possible reservoir of uncharacterized microbial components in local and artisan products. When looking at the taxonomy, such uSGBs mainly belonged to phyla Bacillota (175 uSGBs with 65 of them from the order Lactobacillales), Pseudomonadota (150 uSGBs), and Actinomycetota (131 uSGBs with main contributions from genera *Brevibacterium*, *Garicola*, and *Ruania*). The eukaryotic SGBs also exhibited a non-negligible contribution from uncharacterized species (28% of the 108 SGBs), mainly from dairy and fermented beverages but also from eight additional food categories. These SGBs were assigned taxonomically to fungal clades and especially to families *Aspergillaceae* and *Mucoraceae*. Therefore, such results highlighted the need to perform more in-depth investigations of both prokaryotic and non-prokaryotic communities present in food, with the role and impact of fungi [31], in addition to virus [32], on these ecosystems still largely underexplored.

Applications and implications for food science

The discovery and investigation of uncharacterized microbial species can have strong implications in food science and many applications such as food production and fermentation, food preservation, food safety, and development of functional and novel ingredients. However, the identification of new species represents only a first step, and more in-depth characterizations and studies are necessary to identify functional capabilities and assess possible positive and negative impacts in food production and even short- or long-term effects on human health.

The most immediate application is their use in the production and fermentation of traditional and novel foods. The characteristics in terms of texture, flavor, and nutritional content of dairy and also other fermented foods depend in part on their microbial communities [33]. Linking such microbial compositions with food characteristics is still a challenging task, and even more difficult for newly identified species. Therefore, strain- and genome-resolved metagenomic analyses can bring relevant information to understand functional aspects of such species and have subsequently a key role in understanding the fermentation dynamics, optimizing the production process, and potentially improving the beneficial activities of the developed foods. In these cases, the strain-level taxonomic resolution of the metagenomics studies is of utmost importance, as the activities mentioned are strain-specific traits.

Another important use is related to food safety and food quality [34]. Uncharacterized species detected in food and related processing environments may represent a source of possible pathogens. For example, certain strains of *Cronobacter sakazakii*, found in powdered infant formula, were initially unrecognized but later associated with severe infections in neonates [35]. On the other hand, they could be associated with

protective factors that may have instead a favorable role through inhibition of spoilage and pathogens and production of bioactive metabolites and antimicrobial compounds. Such investigations are relevant for advancing the field forward toward developing microbiome-based solutions, such as protective cultures, which can enhance food safety through natural conservation and relying less on chemical additives and preservatives [36].

Beyond food processing, the investigation of uncharacterized microbes can have wider implications for having a better understanding of the interactions between food, human microbiome, and human health [37]. More specifically, unknown species detected in foods could have probiotic potential, with important consequences for human wellness through gut microbiota interaction. In this regard, some previously uncharacterized species found in food are now recognized for their probiotic potential; for example, *Weissella cibaria* (identified in fermented vegetables and now explored for its oral health benefits [38]) and *Lactobacillus sakei* (commonly found in meat and fish and now investigated for its probiotic properties [39]). This requires advancing the field by elucidating genomic and metabolic characteristics and identifying strains that could have beneficial properties such as short-chain fatty acid production. This is an important research line that can lead to the development of next-generation probiotic products with the final aim of improving gut and overall human health.

Along with scientific advances, regulations and public policies have also an essential role in translating the discovery of uncharacterized species into safe and authorized applications in the food science industry [40]. More specifically, newly proposed strains require thorough safety evaluations for their use in food in line with international guidelines. Such authorizations have often complex processes, and harmonization of regulations between scientists, policymakers, and stakeholders is crucial to maximize the exploitation of these new microbial resources.

Challenges in detection and characterization of unexplored species

Despite advances in sequencing and bioinformatics technologies [41], identification and characterization of unexplored microbial members from food metagenomes have challenges that cover technical, methodological, and interpretative aspects. These go from limitations in taxonomic resolution and functional potential investigation to impacts affecting practical applications such as food production, monitoring of food safety, and possible links with human microbiomes and health (**Table 2**).

While metagenomics enables the rapid identification of not yet cultured microbes, even at large-scale, further experiments including cultivation are often necessary to perform functional characterization of these microorganisms, which is essential to assess their role in food's environments. Functional prediction, however, does not imply actual microbial activity, and its characterization is further complicated by the incompleteness of reference databases [42]. Computational advancements have made metagenomic assembly and MAG reconstruction in complex ecosystems, including foods, quite straightforward. However, the quality of the

generated MAGs depends on several factors such as sample complexity, sequencing depth, and species abundance [11]. Therefore, the generation of high-quality genomes rely on deep sequencing and proper bioinformatics pipelines, which can also involve non-trivial manual curation. Moreover, most of the studies are performed with short-read sequencing technologies which struggle in assembling similar strains and low-abundance species, which could still have a significant role in the community. Long-reads can improve genome assembly, but this comes with added costs. Strengths of both technologies can be combined using hybrid approaches, which are still costly and require methodological computational advancements. There is also the opportunity to perform metagenomic-driven isolation of unknown species [43]; however, biases in culture media and growth environment may still exclude specific microbes. Existing tools enable high-resolution (species- and strain-level) taxonomic profiling of known and unknown microbes in foods. Besides required resources, low-abundance species can still be missed in their identification, and characterization at strain-level is often not simple to interpret [44].

The characterization of interactions among known and unknown species is important to elucidate their role in foods' quality and safety. This relies on complex network analyses, and experimental validations must be also performed. Food microbial composition is influenced by food properties and processing, but such relationships are also confounded by environmental, cultural, and other hidden factors. Analytical and experimental protocols should be implemented to guarantee reproducibility, and standardization procedures should be developed to minimize biases across studies. Additionally, it is important to improve protocols for low-biomass samples, which are often encountered in food-related studies [10]. Metagenomic studies have mainly targeted the bacterial component, while the fungal and viral ones remain less characterized, despite their key role in food processing, quality, and safety. However, some methodological barriers exist such as non-optimal extraction protocols and computational challenges including uncomplete reference databases. Integration with multi-omics data is also important to provide more functional characterization [45]. As an example, metagenomics can enable the identification of new species involved in fermentation processes, which however need to be functionally confirmed through transcriptomic and metabolic analyses. This approach, despite being appealing, cannot still be adopted routinely due to high cost and complexity. Similarly, identification of unrecognized pathogens, AMR genes, and microbial toxins is important for food security aspects [46], but again levels of expression and actual danger must be validated through follow-up experiments. Metagenomics can be also useful to identify candidates for potential probiotics. However, this is only a very preliminary step that requires extensive demonstrations, involving also long and expensive clinical trials, to assess beneficial effects. The identification of known and unknown microbial markers linked to food quality, food properties (e.g., sensory characteristics [47]), and for food traceability and authentication [48] is also appealing. This is however particularly challenging due to the high diversity in food types and food processing characteristics, which imply complex validations on large datasets. Finally, the intricate interaction between food, human microbiome, and human health have become a hot topic in the

field. The evaluation on how the gut microbiome can be affected through microbes in foods is complicated by the difficulty in distinguish between transient and long-term colonizers. Moreover, interventional and long-term studies are required to identify causal mechanisms and assess the impact of food and diet on gut microbiome composition and host health [49].

Conclusion and future perspectives

Experimental and computational developments in metagenomics have allowed to advance the understanding of microbial communities associated with food systems including the identification and characterization of novel species that have potential for application in food science. However, still many barriers must be overcome to have a proper characterization of food microbiomes, their dynamics, and their roles in food production, safety, and innovation. Continuous improvements in integrating short- with long-read sequencing technologies are necessary to have a better understanding of unknown members, especially when low abundant and when similar strains co-exist. Advancements in computational tools need to be focused on improving high-resolution (i.e., strain-level) taxonomic profiling, with a better characterization also of non-prokaryotic communities. Development of more comprehensive databases is also important to enable proper functional characterization, which can also benefit from integration with multi-omics information, a fundamental step to assess potential beneficial or harmful effects of the newly identified species and strains. Environmental, cultural, and other factors introduce complexity layers in the characterization of these communities that must be resolved through large, multidisciplinary, integrative analyses. Moreover, diversity in foods is enormous and under-described, especially in less developed countries and for traditional/artisan products. These analyses can be effective only through curation of protocols, accessibility to data to guarantee reproducibility (e.g., FAIR principles [50]), and development of standardization procedures and ontologies to reduce biases across studies and enable global-scale investigations, with the final aim of linking microbes living in foods and food systems with other environments including human microbiomes and health.

Table 1: Overview of shotgun metagenomic studies investigating food microbiomes. Only studies with accessible raw data are provided.

Study name	Food type	#samples	DOI
AlexaEA_2024	Blue-veined cheese	133	10.1186/s40168-024-01790-4
AlmeidaOGG_2021	Cocoa bean	10	10.1128/AEM.00584-21
ArikanM_2020	Kombucha	12	10.1111/1750-3841.14992
BarcenillaC_2024	Meat	85	10.1186/s40168-024-01856-3
BertuzziAS_2018	Cheddar cheese	42	10.1128/msystems.00211-17
BreselgeS_2025	Water kefir	356	10.1038/s42003-025-07808-3

CaffreyEB_2024	Fermented food	90	10.1101/2024.03.29.587370
CarlinoN_2024	Multiple	>1000	10.1016/j.cell.2024.07.039
ChaconVargasK_2020	Pulque	5	10.1038/s41598-020-71864-4
Corral-JaraKF_2024	Fish	12	10.1016/j.heliyon.2024.e33866
CrovadoreJ_2017	Honey	5	10.1128/genomea.00425-17
DecadtH_2024	Gouda cheese	19	10.1016/j.fm.2020.103608
DeFilippisF_2024	Cheese	>500	10.1038/s41522-024-00541-5
DeRoosJ_2020	Beer	6	10.3389/fmicb.2020.01692
DuR_2020	Chinese liquor	3	10.1128/aem.00456-20
DuruIC_2018	Maasdam cheese	6	10.1016/j.ijfoodmicro.2018.05.017
EinsonEJ_2018	Sauerkraut	5	10.1128/aem.01680-18
EscobarZepedaA_2016	Cotija cheese	1	10.1016/j.fm.2016.02.004
FalardeauJ_2023	Cheese	56	10.1128/aem.02004-22
FerrocinoI_2018	Sausage	16	10.1128/aem.02120-17
FontanaF_2023	Raw milk cheese	128	10.1128/msystems.01068-22
Franciosal_2021	Sausage	15	10.1016/j.foodres.2021.110379
Goes-NetoA_2021	Kombucha	9	10.1111/1462-2920.15405
González-OrozcoBD_2023	Kefir	2	10.1111/1750-3841.16614
HeilCS_2018	Beer	1	10.1002/yea.3280
HellmannSL_2020	Sausage	12	10.1007/s00217-019-03404-y
Hernandez-OaxacaD_2021	Pozol	4	10.3389/fmicb.2021.683050
HuY_2021	Suansun	18	10.3390/pr9091669
JiangL_2021	Paocai	39	10.3390/foods11010062
JiangL_2025	Horse bean-chili paste	8	10.1016/j.foodchem.2024.141076
KastmanEK_2016	Cheese	18	10.1128/mbio.01157-16
KawaiT_2012	Fish	2	10.1093/cid/cir1040
KharnaiorP_2023	Soybean	4	10.1016/j.ijfoodmicro.2023.110417
KotheCI_2024	Miso	6	10.1016/j.fm.2023.104372
KumarJ_2019	Kinema	4	10.3389/fmicb.2019.01744
LandisEA_2021	Fermented grain	40	10.7554/elife.61644
LandisEA_2022	Kombucha	23	10.1128/msystems.00157-22
LeechJ_2020	Fermented food	58	10.1128/msystems.00522-20
LeonardSR_2016	Spinach	12	10.1371/journal.pone.0167870
LimaCOC_2021	Cocoa bean	14	10.1016/j.fm.2020.103608
LiuD_2021	Potherb mustard	12	10.1016/j.foodres.2021.110622
LiZ_2018	Pu-erh tea	6	10.1016/j.fm.2018.07.001
LiZ_2019	Red pepper sauce	8	10.1016/j.fm.2019.103295
LordanR_2019	Yogurt	5	10.1016/j.foodchem.2020.127384
MagliuloR_2024	Mozzarella cheese	49	10.1016/j.foodres.2024.114798
ManaraS_2023	Injera	2	10.1016/j.cub.2023.04.011
McHughAJ_2020	Milk	16	10.1128/msystems.00226-20

MilaniC_2019	Parmesan cheese	4	10.1038/s41467-019-09303-w
Mota-GutierrezJ_2021	Cocoa bean	20	10.1128/AEM.00425-21
PapadimitriouK_2024	Staka	1	10.3390/foods13071129
PasolliE_2020	Commercial probiotic, Yogurt	16	10.1038/s41467-020-16438-8
PatroJN_2016	Commercial probiotic	10	10.1128/msphere.00057-16
PfeferT_2016	Cheese	36	PRJNA430402 (BioProject)
PorcellatoD_2016	Dutch-type cheese	12	10.1016/j.idairyj.2016.05.005
PothakosV_2020	Coffee	6	10.1016/j.crbiot.2020.02.001
QuigleyL_2016	Continental-type cheese	10	10.1128/msystems.00023-16
QuijadaNM_2022	Cheese	2	10.3389/fmicb.2022.813480
QuT_2024	Yogurt	22	10.3168/jds.2023-23983
RippF_2014	Sausage	1	10.1186/1471-2164-15-639
SaakCC_2023	Cheese	18	10.1128/msystems.00701-22
SabaterC_2024	Honey	63	10.1016/j.ijfoodmicro.2024.110789
SalvetiE_2016	Grape	2	10.3389/fmicb.2016.00937
Santamarina-GarciaG_2024	Idiazabal cheese	46	10.1186/s40168-024-01980-0
SequinoG_2024	Anchovy	79	10.1016/j.foodres.2023.113788
SequinoG_2024b	Beef	70	10.1016/j.foodres.2024.114318
SequinoG_2025	Poultry	80	10.1016/j.fm.2025.104734
ShangpliangHNJ_2023	Laal dahi	2	10.1016/j.ijfoodmicro.2023.110300
ShangpliangHNJ_2023b	Jalebi	2	10.1016/j.foodres.2023.113130
SomervilleV_2019	Whey starter	1	10.1186/s12866-019-1500-0
SonetsIV_2024	Beer, cider	12	10.1016/j.fm.2024.104520
SongQ_2021	Suancai	3	10.1016/j.foodres.2020.109924
SternesPR_2017	Grape	20	10.1093/gigascience/gix040
SuhreT_2025	Kombucha	16	10.1016/j.fbio.2025.106122
SulaimanJ_2014	Soybean	7	10.3389/fmicb.2014.00556
TamangJP_2022	Soybean	4	10.1016/j.foodres.2021.110904
TangG_2025	Douchi	37	10.1016/j.fbio.2025.106128
Tenorio-SalgadoS_2021	Milk kefir	2	10.1002/mbo3.1183
TomarS_2023	Keem	1	10.1007/s12033-023-00839-3
TsougouN_2024	Sfela cheese	4	10.3390/foods13071023
ValentinoV_2022	Vegetables	10	10.1016/j.foodres.2022.112202
ValentinoV_2023	Ice cream	10	10.1128/aem.00765-23
VerceM_2019	Water kefir	4	10.3389/fmicb.2019.00479
WalshAM_2016	Kefir	18	10.1128/msystems.00052-16
WalshAM_2017	Nunu	10	10.1128/AEM.01144-17
WalshAM_2020	Cheese	77	10.1038/s43016-020-0129-3
WalshLH_2023	Milk kefir	266	10.1016/j.isci.2023.108004
WolfeBE_2014	Cheese	24	10.1016/j.cell.2014.05.041
XieM_2019	Soybean	1	10.1016/j.foodres.2018.10.076

YangC_2021	Cheese	8	10.1016/j.lwt.2021.112043
YangY_2021	Daqu	8	10.1016/j.foodres.2020.110037
YaoG_2017	Koumiss	30	10.3389/fmicb.2017.00165
YapM_2020	Milk	18	10.1038/s41598-020-78773-6
YapM_2024	Milk	241	10.1128/msystems.01290-23
YasirM_2020	Laban	2	10.1016/j.foodres.2020.109639
YasirM_2022	Pickle	18	10.1016/j.foodres.2022.111190
YouL_2022	Koumiss	23	10.1016/j.lwt.2021.112945
YulandA_2020	Soybean	2	10.1186/s13104-020-05406-6
YuY_2022	Mustard	9	10.3389/fmicb.2022.824644
ZhangJ_2021	Xuecai	20	10.1016/j.foodres.2020.109986
ZhaoCC_2020	Hongeo	5	10.1016/j.foodres.2020.109030

Table 2: Advantages and challenges in characterizing the hidden diversity in food metagenomes.

Aspect	Advantages	Challenges
Detection of uncultured microbes in food [25]	Identification of food-associated microbes without cultivation	Validations are required for their functional characterization
Detection of putative novel species [16, 25]	Identification of unknown species contributing to food microbiome	Species may lack close genomic references, complicating their characterization
Reconstruction of genomes from food metagenomes [11]	Computational-based automatic extraction of metagenome-assembled genomes (MAGs)	Quality of MAGs is affected by multiple factors (e.g., sample complexity, sequencing depth, species abundance)
Long - vs short- read sequencing [8, 12]	Long-read sequencing can improve assembly results	Long-read sequencing has still higher error rates and costs
Generation of taxonomic profiles from food metagenomes [9]	Current methods can distinguish closely related species and strains	Bioinformatics pipelines can be time-consuming; low-abundance species are difficult to detect; strain-level differentiation is difficult to interpret
Characterization of functional potential of microbes in food [42]	Estimation of metabolic characteristics in food microbes	Functional prediction may not reflect actual microbial activity; it is also limited by the incompleteness of functional annotation databases
Detection of low-abundance species [40]	Characterization of low-abundance or rare species, still potentially functional relevant	Computational approaches struggle in detecting and characterizing low-abundance microbes
Evaluation of food characteristics and processing on microbiome composition [16, 19, 21]	Identification of how food characteristics and processing affect microbial diversity	Results can be affected by biases due to environmental, cultural, and other hidden factors
Metagenomic-driven cultivation of uncharacterized species [43]	Combination of sequencing technologies with culturomics to characterize novel species	Biases in culturing may still exclude certain microbes
Standardization of protocols for food metagenomics [10]	Identification of unknown species is facilitated by more standard and reproducible workflows	Protocols differ across studies and food types, making comparisons more challenging
Identification of novel fungi and viruses in foods [31, 32]	Characterization of fungal and viral components that may affect food processing or safety	Results are limited by methodological challenges and incompleteness in reference databases

Integration with multi-omics data [45, 48]	Better functional characterization of the unknown microbes	Acquisition and analysis of multi-omics data is costly and time-consuming
Monitoring of food microbiomes spatially and longitudinally [14, 23]	Tracking of microbial composition in space and/or time	Extensive sampling in space and/or time is often required for meaningful results
Insights into fermentation processes [26, 33]	Detection of new species contributing to fermentation	Validations are required to assess their role in these processes
Discovery of probiotic strains [41]	Detection of potential probiotics with beneficial effects	Isolation of strains is required; Validations and regulatory approval are costly and time-consuming
Detection of novel antimicrobial resistance (AMR) genes or toxins [7, 12, 46]	Identification of AMR genes and toxin-producing microbes	Validations are required to evaluate expression levels and risks in food consumption
Monitoring food safety [1, 7, 34]	Detection of previously unrecognized pathogens	Isolation of strains and additional tests are required to assess pathogenicity and other risk factors
Development of microbial indicators for food quality control [5, 16, 33]	Identification of microbial markers, including novel species, for food quality assessment	Variability across food types makes universal markers difficult to define
Linking microbial characteristics with food sensory properties [47]	Identification of known and unknown species in food sensory properties	Validations are required to confirm associations between microbial and food characteristics
Authenticity and traceability of foods [48]	Identification of species characteristic of traditional or local products	Results are often not reproducible, partly due to biases and contamination issues
Characterization of food-to-gut microbial transmission [18, 25]	Identification of how known and unknown species contribute to the (human) gut microbiome composition	Validations are required to distinguish transient members from long-term colonizers
Characterization of the food-microbiome-human axis [37, 49]	Exploration of the role of food microbes and diet on human microbiomes and health	Long-term studies are required to validate causal mechanisms and possible microbiome-based interventions

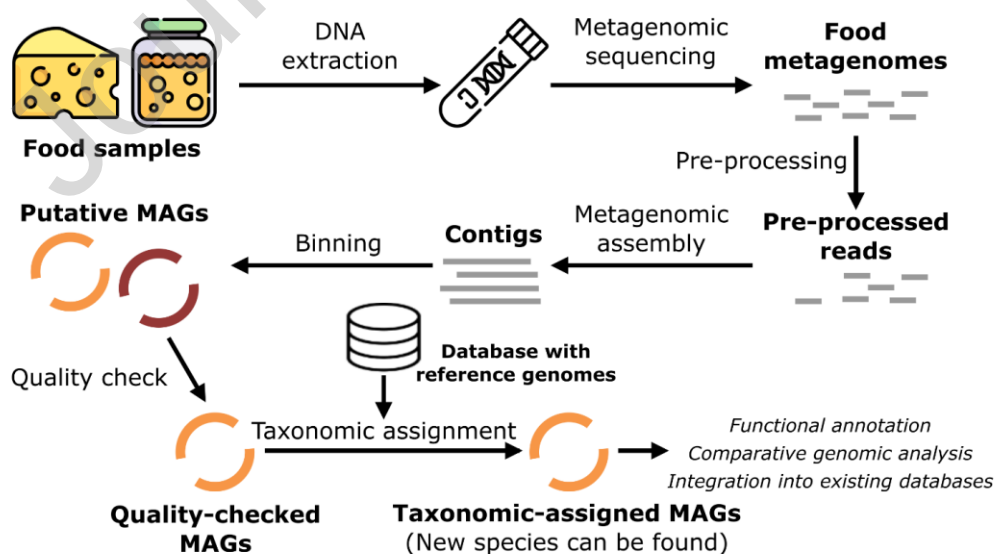


Figure 1: Workflow for characterizing novel species in the food microbiome. The figure summarizes the step-by-step process of species characterization, starting from raw metagenomics data to functional analysis. MAGs: Metagenome-Assembled Genomes.

Data availability

No data were used for the research described in the article.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no 818368 (MASTER project). The project was also funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 – Call for Tender No. 341 of 15 March 2022 of the Italian Ministry of University and Research funded by the European Union – NextGenerationEU; Project code PE00000003, Concession Decree No. 1550 of 11 October 2022 adopted by the Italian Ministry of University and Research, Project title “ON Foods – Research and Innovation Network on Food and Nutrition Sustainability, Safety and Security – Working ON Foods.”

Declaration of competing interest

The authors declare no conflicts of interest.

References

1. Beck KL, Haiminen N, Chambliss D, Edlund S, Kunitomi M, Huang BC, Kong N, Ganesan B, Baker R, Markwell P, et al.: **Monitoring the microbiome for food safety and quality using deep shotgun sequencing.** *Npj Sci Food* 2021, **5**:3.
2. Cao Y, Fanning S, Proos S, Jordan K, Srikumar S: **A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies.** *Front Microbiol* 2017, **8**:1829.
3. Walsh AM, Leech J, Huttenhower C, Delhomme-Nguyen H, Crispie F, Chervaux C, Cotter PD: **Integrated molecular approaches for fermented food microbiome research.** *FEMS Microbiol Rev* 2023, **47**.
4. Brennan FP, Alsanius BW, Allende A, Burgess CM, Moreira H, Johannessen GS, Castro PML, Uyttendaele M, Truchado P, Holden NJ: **Harnessing agricultural microbiomes for human pathogen control.** *ISME Commun* 2022, **2**:44.
5. Pellissery AJ, Vinayamohan PG, Amalaradjou MAR, Venkitanarayanan K: **Spoilage bacteria and meat quality.** In *Meat Quality Analysis*. Elsevier; 2020:307–334.

6. Vinayamohan PG, Pellissery AJ, Venkitanarayanan K: **Role of horizontal gene transfer in the dissemination of antimicrobial resistance in food animal production.** *Curr Opin Food Sci* 2022, **47**:100882.
7. Pennone V, Cobo-Díaz JF, Prieto M, Alvarez-Ordóñez A: **Application of genomics and metagenomics to improve food safety based on an enhanced characterisation of antimicrobial resistance.** *Curr Opin Food Sci* 2022, **43**:183–188.
8. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q: **Opportunities and challenges in long-read sequencing data analysis.** *Genome Biol* 2020, **21**:30.
9. Liu S, Rodriguez JS, Munteanu V, Ronkowski C, Sharma NK, Alser M, Andreade F, Blekhman R, Błaszczyk D, Chikhi R, et al.: **Analysis of metagenomic data.** *Nat Rev Methods Primers* 2025, **5**. •• This review explores the methodological approaches and current practices that characterize the analysis of metagenomic data.
10. Barcenilla C, Cobo-Díaz JF, De Filippis F, Valentino V, Cabrera Rubio R, O’Neil D, Mahler de Sanchez L, Armanini F, Carlino N, Blanco-Míguez A, et al.: **Improved sampling and DNA extraction procedures for microbiome analysis in food-processing environments.** *Nat Protoc* 2024, **19**:1291–1310. •• This paper presents a protocol based on an optimized DNA purification method to generate high-quality sequencing data from samples collected in food production and processing environments.
11. Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu A, Bian Z, Zhang L: **A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data.** *Comput Struct Biotechnol J* 2021, **19**:6301–6314.
12. Brown CL, Keenum IM, Dai D, Zhang L, Vikesland PJ, Pruden A: **Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes.** *Sci Rep* 2021, **11**:3753.
13. Stellato G, Utter DR, Voorhis A, De Angelis M, Eren AM, Ercolini D: **A few *Pseudomonas* oligotypes dominate in the meat and dairy processing environment.** *Front Microbiol* 2017, **8**:264.
14. Pham VD, Simpson DJ, Gänzle MG: **Strain-level identification of bacteria persisting in food and in food processing facilities: when do two isolates represent the same strain and which tools identify a strain?** *Curr Opin Food Sci* 2025, **61**:101245. • This review discusses the concept of ‘microbial strains’ to perform the tracking of strains in food and food processing environments.
15. Asnicar F, Thomas AM, Passerini A, Waldron L, Segata N: **Machine learning for microbiologists.** *Nat Rev Microbiol* 2024, **22**:191–205.
16. Walsh AM, Macori G, Kilcawley KN, Cotter PD: **Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality.** *Nature Food* 2020, **1**:500–510. •• This paper presents a meta-analysis of cheese microbiomes with the identification of several putative new species that could influence taste or color.
17. You L, Yang C, Jin H, Kwok L-Y, Sun Z, Zhang H: **Metagenomic features of traditional fermented milk products.** *Lebenson Wiss Technol* 2022, **155**:112945.
18. Pasolli E, De Filippis F, Mauriello IE, Cumbo F, Walsh AM, Leech J, Cotter PD, Segata N, Ercolini D: **Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome.** *Nat Commun* 2020, **11**:2610.
19. Leech J, Cabrera-Rubio R, Walsh AM, Macori G, Walsh CJ, Barton W, Finnegan L, Crispie F, O’Sullivan O, Claesson MJ, et al.: **Fermented-food metagenomics reveals substrate-associated differences in**

taxonomy and health-associated and antibiotic resistance determinants. *mSystems* 2020, **5**. • This paper expands the knowledge of fermented food microbiomes by analyzing 58 artisanal products from various countries.

20. Kothe CI, Rasmussen JA, Mak SST, Gilbert MTP, Evans J: **Exploring the microbial diversity of novel misos with metagenomics.** *Food Microbiol* 2024, **117**:104372.
21. Liu Y, Yu L, Tian F, Chen W, Zhai Q: **Meta-analysis of microbiomes reveals metagenomic features of fermented vegetables.** *Food Res Int* 2023, **173**:113248.
22. Parente E, Ricciardi A: **A comprehensive view of food microbiota: Introducing FoodMicrobionet v5.** *Foods* 2024, **13**:1689.
23. Li Y, Lyu H, Yang H, Ju Z, Ma C, Hou H, Wang Y, Zhou Y, Gao Y, Yang J, et al.: **FoodMicroDB: A microbiome database for composition and time-series research in food.** *iMetaOmics* 2024, **1**.
24. Richardson L, Allen B, Baldi G, Beracochea M, Bileschi ML, Burdett T, Burgin J, Caballero-Pérez J, Cochrane G, Colwell LJ, et al.: **MGnify: the microbiome sequence data analysis resource in 2023.** *Nucleic Acids Res* 2023, **51**:D753–D759.
25. Carlino N, Blanco-Míguez A, Punčochář M, Mengoni C, Pinto F, Tatti A, Manghi P, Armanini F, Avagliano M, Barcenilla C, et al.: **Unexplored microbial diversity from 2,500 food metagenomes and links with the human microbiome.** *Cell* 2024, **187**:5775-5795.e15. •• This resource paper presents an integrated analysis of >2,500 food metagenomes, including the generation of prokaryotic and eukaryotic MAGs, and explores their links with the human microbiome.
26. Caffrey EB, Olm MR, Kothe CI, Evans J, Sonnenburg JL: **MiFoDB, a workflow for microbial food metagenomic characterization, enables high-resolution analysis of fermented food microbial dynamics.** *bioRxiv* 2024, doi:10.1101/2024.03.29.587370.
27. Gänzle M: **The periodic table of fermented foods: limitations and opportunities.** *Appl Microbiol Biotechnol* 2022, **106**:2815–2826.
28. Obafemi YD, Oranusi SU, Ajanaku KO, Akinduti PA, Leech J, Cotter PD: **African fermented foods: overview, emerging benefits, and novel approaches to microbiome profiling.** *Npj Sci Food* 2022, **6**:15.
29. Abdill RJ, Adamowicz EM, Blekhman R: **Public human microbiome data are dominated by highly developed countries.** *PLoS Biol* 2022, **20**:e3001536. • This paper analyzes the country of origin of human microbiome samples available in public repositories and highlights the need for greater global representation in microbiome studies.
30. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, et al.: **Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle.** *Cell* 2019, **176**:649-662.e20.
31. Case NT, Gurr SJ, Fisher MC, Blehert DS, Boone C, Casadevall A, Chowdhary A, Cuomo CA, Currie CR, Denning DW, et al.: **Fungal impacts on Earth's ecosystems.** *Nature* 2025, **638**:49–57.
32. Mahony J, van Sinderen D: **Virome studies of food production systems: time for “farm to fork” analyses.** *Curr Opin Biotechnol* 2022, **73**:22–27.
33. Sharma R, Garg P, Kumar P, Bhatia SK, Kulshrestha S: **Microbial fermentation and its role in quality improvement of fermented foods.** *Fermentation* 2020, **6**:106.

34. Billington C, Kingsbury JM, Rivas L: **Metagenomics approaches for improving food safety: A review.** *J Food Prot* 2022, **85**:448–464.
35. Chauhan R, Tall BD, Gopinath G, Azmi W, Goel G: **Environmental risk factors associated with the survival, persistence, and thermal tolerance of *Cronobacter sakazakii* during the manufacture of powdered infant formula.** *Crit Rev Food Sci Nutr* 2023, **63**:12224–12239.
36. Verma DK, Thakur M, Singh S, Tripathy S, Gupta AK, Baranwal D, Patel AR, Shah N, Utama GL, Niamah AK, et al.: **Bacteriocins as antimicrobial and preservative agents in food: Biosynthesis, separation and application.** *Food Biosci* 2022, **46**:101594.
37. Ross FC, Patangia D, Grimaud G, Lavelle A, Dempsey EM, Ross RP, Stanton C: **The interplay between diet and the gut microbiome: implications for health and disease.** *Nat Rev Microbiol* 2024, **22**:671–686. •• This review examines the impact of geographical location on the gut microbiome and how various diets influence their composition and function.
38. Kang M-S, Lee D-S, Lee S-A, Kim M-S, Nam S-H: **Effects of probiotic bacterium *Weissella cibaria* CMU on periodontal health and microbiota: a randomised, double-blind, placebo-controlled trial.** *BMC Oral Health* 2020, **20**:243.
39. Lim S, Moon JH, Shin CM, Jeong D, Kim B: **Effect of *Lactobacillus sakei*, a probiotic derived from Kimchi, on body fat in Koreans with obesity: A randomized controlled study.** *Endocrinol Metab* 2020, **35**:425–434.
40. Fortin ND: **Food regulation: law, science, policy, and practice.** *John Wiley & Sons* 2022.
41. Chelliah R, Banan-MwineDaliri E, Khan I, Wei S, Elahi F, Yeon S-J, Selvakumar V, Ofosu FK, Rubab M, Ju HH, et al.: **A review on the application of bioinformatics tools in food microbiome studies.** *Brief Bioinform* 2022, **23**.
42. Rodríguez Del Río Á, Giner-Lamia J, Cantalapiedra CP, Botas J, Deng Z, Hernández-Plaza A, Munar-Palmer M, Santamaría-Hernando S, Rodríguez-Herva JJ, Ruscheweyh H-J, et al.: **Functional and evolutionary significance of unknown genes from uncultivated taxa.** *Nature* 2024, **626**:377–384.
43. Kothe CI, Renault P: **Metagenomic driven isolation of poorly culturable species in food.** *Food Microbiol* 2025.
44. Van Rossum T, Ferretti P, Maistrenko OM, Bork P: **Diversity within species: interpreting strains in microbiomes.** *Nat Rev Microbiol* 2020, **18**:491–506. • This review illustrates the advances in studying within-species variation using metagenomics data and highlights the challenges posed by the overwhelming numbers of methods and terminology.
45. Ferrocino I, Rantsiou K, McClure R, Kostic T, de Souza RSC, Lange L, FitzGerald J, Kriaa A, Cotter P, Maguin E, et al.: **The need for an integrated multi-OMICs approach in microbiome science in the food system.** *Compr Rev Food Sci Food Saf* 2023, **22**:1082–1103. • This review discusses the need to perform multi-omics approaches within food systems to enhance our understanding of microbial dynamics and improve food quality and safety.
46. Lee AW-T, Ng IC-F, Wong EY-K, Wong IT-F, Sze RP-P, Chan K-Y, So T-Y, Zhang Z, Ka-Yee Fung S, Choi-Ying Wong S, et al.: **Comprehensive identification of pathogenic microbes and antimicrobial resistance genes in food products using nanopore sequencing-based metagenomics.** *Food Microbiol* 2024, **121**:104493.

47. dos Santos Rocha C, Magnani M, de Paiva Anciêns Ramos GL, Bezerril FF, Freitas MQ, Cruz AG, Pimentel TC: **Emerging technologies in food processing: impacts on sensory characteristics and consumer perception.** *Curr Opin Food Sci* 2022, **47**:100892.
48. Varunjikar MS, Moreno-Ibarguen C, Andrade-Martinez JS, Tung H-S, Belghit I, Palmblad M, Olsvik PA, Reyes A, Rasinger JD, Lie KK: **Comparing novel shotgun DNA sequencing and state-of-the-art proteomics approaches for authentication of fish species in mixed samples.** *Food Control* 2022, **131**:108417.
49. Leeming ER, Johnson AJ, Spector TD, Le Roy CI: **Effect of diet on the gut Microbiota: Rethinking intervention duration.** *Nutrients* 2019, **11**:2862.
50. Lange L, Berg G, Cernava T, Champomier-Vergès M-C, Charles T, Cocolin L, Cotter P, D'Hondt K, Kostic T, Maguin E, et al.: **Microbiome ethics, guiding principles for microbiome research, use and knowledge management.** *Environ Microbiome* 2022, **17**:50.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

Declaration of Competing Interest

The authors declare no conflicts of interest.