npj | biofilms and microbiomes

Review

Published in partnership with Nanyang Technological University

https://doi.org/10.1038/s41522-026-00985-x

Water kefir as a paradigm for multi-omics and genome-scale metabolic modelling in fermented food

Anik Khan[1,2,3] ✉, Samuel Breselge[1,2,4], A. Kate O'Mahony[1,2], Orla O'Sullivan[1,2,4,5], Paul D. Cotter[1,2,4], Sinéad N McCarthy[6], Jennifer Mahony[2,3,7] ✉ & John G. Kenny[1,2,3,4] ✉

Water Kefir is a plant-based fermented beverage, traditionally produced on a small scale by fermenting a sucrose solution with fresh or dried fruits, using water kefir grains as inoculum. The grains are relatively simple communities that consist of both eukaryotes and prokaryotes, rendering them a paradigm for studying microbial ecology and interspecies interactions. Recently, water kefir has attracted growing research and industrial interest due to its potential and perceived health benefits. Owing to its increasing popularity, there is a growing demand for controlled and standardised production on an industrial scale. However, industrial-scale production remains a challenge due to the limited knowledge of the biological interactions of the microbial consortia and the lack of defined starter cultures. This review examines the current understanding of microbial and metabolic complexity of water kefir obtained from various omics studies. It further investigates the potential of an integrated multi-omics approach to elucidate mechanisms of microbial interactions and provides a roadmap for conducting multi-omics studies on fermented foods using water kefir as an example. This review also explores the potential application of genome-scale metabolic modelling in the development of functional and defined microbial communities for food fermentation. It identifies key challenges associated with such modelling and provides perspectives to address them. Finally, this review briefly discusses the regulatory challenges associated with the use of defined communities in food systems.

Microbes exist in communities nearly everywhere on earth, where they interact with each other and with their environment to maintain the ecosystem functions and dynamics. These interactions help to maintain microbial diversity, influence metabolic phenotypes, and shape the community functionalities. Advancements in omics, particularly in meta-omics approaches have enhanced our understanding of diversity, assembly, organization, and functions of these communities in recent decades, generating vast datasets such as metagenomics, metatranscriptomics, metaproteomics, and metabolomics data[1]. Taken together, these large datasets provide a comprehensive insight into microbial community function and dynamics by revealing their composition, gene expression, protein production, and metabolite profiles[2]. The magnitude and complexity of data necessitate the development of systems biology tools to elucidate underlying mechanisms involved in complex microbial communities, such as microbial interactions, metabolic pathways, and community-level functional dynamics[3].

Genome-Scale Metabolic Modelling (GSMM) is a systems biology approach with the potential to model the metabolism of individual species as well as complex microbial communities, and host-microbiome interactions[4–6]. The modelling approach involves construction of Genome-scale metabolic model (GEM), which is a mathematical model that describes all the metabolic reactions that an organism can perform and allows quantitative prediction of its metabolic genotype-phenotype relationships[7]. More recently, individual GEMs have been combined to

[1]Department of Food Biosciences, Teagasc Food Research Centre, Cork, Ireland. [2]APC Microbiome Ireland SFI Research Centre, University College Cork, Cork, Ireland. [3]School of Microbiology, University College Cork, Cork, Ireland. [4]VistaMilk SFI Research Centre, Cork, Ireland. [5]School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland. [6]Department of Agrifood Business and Spatial Analysis, Teagasc Food Research Centre, Ashtown, Dublin, Ireland. [7]Department of Food Science, Section for Food Microbiology, Gut Health and Fermentation, University of Copenhagen, Frederiksberg C, Denmark. ✉e-mail: Anik.Khan@teagasc.ie; J.Mahony@ucc.ie; John.Kenny@teagasc.ie

model the metabolism of microbial communities, allowing the study of mechanistic connections between genotypes and community-level phenotypes[8,9]. Since the development of the first GEM of *Haemophilus influenza*[10], the technique has been successfully applied to a variety of applications including metabolic engineering[11–13], prediction of enzyme function[14], drug development[14], human diseases[15], understanding complex microbial community interactions[16–19], and more recently, in the design of synthetic microbial communities (SynComs)[20,21]. SynComs are deliberately constructed communities of multiple well-characterized microbes that are less complex than their natural counterparts but retain their key functional features[22].

Recent advances in meta-omics technologies, have significantly advanced our understanding of the genetic, functional, and metabolic characteristics of traditional fermented food microbial communities[23]. However, reproducible industrial scale production of these traditional fermented foods remains a challenge. The lack of defined starter cultures and optimised fermentation parameters often leads to inconsistent production across fermentation cycles in the industrial setting[24]. To meet the growing consumer demand for healthy and sustainable fermented foods, industrial scale production and the development of defined starter cultures have become increasingly attractive[25]. GSMM has emerged as a promising approach to developing defined starter cultures for traditional fermented foods. Among these fermented foods, water kefir (WK) serves as an ideal paradigm for GSMM-guided SynCom development and analysis, due to its relatively simple community and substrates which nevertheless offer rich ecological interactions and recently established extensive multi-omics insights on WK microbial community.

In this review, the state-of-the-art understanding of WK microbial communities obtained through various omics studies is summarized. The potential of WK as a model system for microbial ecology research is highlighted and a roadmap for an integrated meta-omics study is provided. The likely challenges of such a multi-omics study are identified, and possible circumvention strategies are discussed. The potential application of GSMM is explored in the formulation of robust, functional microbial communities for novel fermented food development using WK as an example. Finally, the regulatory challenges associated with the use of these newly formulated microbial communities in the food system are briefly discussed, with a focus on safety and consumer acceptance.

Introduction to water kefir

Water kefir is a traditional, plant-based fermented beverage produced through the metabolic activity of a symbiotic microbial community residing within water kefir grains (WKGs). These grains are microbially produced polysaccharide matrices, primarily composed of the exopolysaccharide dextran[26,27]. WKGs host a diverse consortium of bacteria and yeasts, including lactic acid bacteria (LAB), acetic acid bacteria (AAB), various yeasts, and often *Bifidobacterium* species and *Zymomonas mobilis*[28]. Traditionally, WK fermentation is initiated by inoculating 5–10% (w/v) sugar water with 5–10% (w/v) WKGs and adding dried or fresh fruits. Over the course of one to several days, WK fermentation yields a mildly acidic, effervescent, and slightly alcoholic beverage containing organic acids, ethanol, carbon dioxide, and volatile aroma compounds[24]. The microbial community structure varies significantly across different grains and fermentation practices. However, a recent pan multi-omics study of global WKGs that enabled the identification of core genera and functional redundancies within these ecosystems[29].

Earlier studies primarily focused on taxonomic composition using culture-based methods, amplicon sequencing, and targeted metabolite analysis of compounds such as organic acids, sugars, and ethanol[30–36]. More recent research has employed multi-omics approaches to gain novel insights into WK ecology.

Insights from omics studies

Omics-based investigations have greatly enhanced our understanding of WK ecology. The first shotgun metagenomics study of WK was conducted in 2019, which analysed microbial composition in both the liquid and grain fractions at 24-h and 72-h timepoints of fermentation and provided key insights into WK ecology[37]. Since then, several studies used shotgun metagenomics to study WK microbial communities, enabling detailed characterisation of microbial communities, including comparative analyses of bacterial and fungal composition, temporal investigation of community dynamics during the fermentation, and the identification of novel species, some of which were subsequently isolated through targeted cultivation[29,38–40]. Additionally, analysis of the metagenomic data suggested that most WK-associated species are auxotrophic for at least one essential amino acid, indicating a high potential for metabolic cross-feeding and symbiotic interactions[36]. Functional analysis of metagenomic datasets linked the production of metabolites with different microbial species. For example, in WK the conversion of fructose into mannitol was particularly linked to *Liquorilactobacillus hilgardii, Oenococcus aquikefiri*, and *Bifidobacterium aquikefiri*, and the production of glycerol was linked to *Saccharomyces cerevisiae*. Mannitol and glycerol are known to contribute to the sweetness and texture of the resulting WK beverage, thereby improving the product quality. It also revealed the evidence of metabolic cross-feeding between bacteria and yeast due to the bacterial dependence on yeast or substrate-derived amino acids and cofactors[37]. Existing metagenomic datasets can be further utilised to predict specific cross-feeding and other symbiotic interactions, which can provide insights into key species identification and microbial interactions during the fermentation, guiding the design and optimisation of SynComs.

Metabolomics and volatilomics along with metagenomics have provided broader insights into the metabolic dynamics of WK fermentations, revealing links between fermentation substrates, microbial community composition, and metabolic outputs[29,40–44]. Finally, metatranscriptomics has provided a valuable perspective on gene expression activity during WK fermentation[45]. Omics-based studies of WK microbial communities along with their key findings are summarized in Table 1.

Water kefir as a model system for microbial ecology and fermentation research

WK presents a valuable model system for studying microbial ecology, interspecies interactions, and food fermentation processes. Furthermore, it is an efficient system to validate metabolic models. WK communities are relatively simple, yet sufficiently rich in ecological interactions, with most communities consisting of 5–20 dominant species, which account for the majority of the microbial population during fermentation[29]. This compositional structure allows researchers to examine a range of ecological relationships, including cross-feeding, competition, syntrophy, and niche partitioning, within a tractable and controlled environment.

The short fermentation cycle and well-characterised metabolic outputs (e.g., ethanol, organic acids, exopolysaccharides, and volatile compounds) provide a direct and measurable link between microbial composition and function. Unlike many natural or industrial microbiomes, WK fermentation can be carried out under aseptic conditions using sterile media[46], enabling reproducible experiments and perturbation studies. Although the fermentation medium is currently only semi-defined, efforts have been made toward the development of a more defined medium[47].

Initial SynComs for WK have had variable success. In one study, two-species SynComs consisting of one LAB and one yeast were developed for WK[47]. Among six tested combinations, all of them showed slower acidification and higher residual sugar levels compared to the fermentations using WKGs. Another study also described nine SynComs for WK, each comprising three to four species[48]. These SynComs were based on *Zymomonas mobilis* in combination with *Liquorilactobacillus hordei*, *Saccharomyces bayanus*, *Bifidobacterium aquikefiri*, and/or *Brettanomyces bruxellensis*. Here, a SynCom lacking *L. hordei* exhibited markedly increased gluconic acid production, showing the ability to alter organic acid profiles to modify taste profiles. Another study of co-culture experiments involving one yeast (*Zygotorulaspora florentina* or *S. cerevisiae*) and one LAB (*L. hordei* or *Liquorilactobacillus nagelii*) demonstrated mutualistic interactions

Table 1 | Overview of omics studies in water kefir and key insights obtained

Approach	Insights obtained	References
Metagenomics	This study used metagenomics to analyse microbial composition in both the liquid and grain fractions at 24-h and 72-h timepoints of fermentation and provided key insights into WK ecology	37
Metagenomics	This study applied shotgun metagenomics sequencing to a variety of fermented foods, including WK and milk kefir samples. Their analysis revealed clear taxonomic differences between milk kefir and WK communities.	39
Metagenomics, Metabolomics, and Volatilomics	This study applied shotgun metagenomics to sequence a WK fermentation time series, identifying correlations between microbial species, metabolites, and volatile compounds. This work provided valuable insights into potential producers of key metabolites and volatiles and led to the detection of two potentially novel *Curvibacter* species.	40
Metagenomics, Metabolomics, and Volatilomics	This study performed a large-scale study of 69 WK communities using shotgun metagenomics, metabolomics, and volatilomics. Their analysis revealed 18 putatively novel species, including two new *Bifidobacterium* spp. In total, 96 different species were detected across samples, supported by either metagenome-assembled genomes (MAGs) or at least 35% genome breadth. Consequently, a core microbiome was proposed for WK, comprising 13 genera, including yeasts, LAB, AAB, *Bifidobacterium*, and *Zymomonas*. Correlation analyses suggested functional redundancy for major WK metabolites (e.g., lactic acid, acetic acid, sugars, and ethanol), whereas volatile compound production appeared to be more species-specific.	29
Metagenomics and Metabolomics	This study used WKG microbial community to produce a novel beverage using soy whey as a substrate. Untargeted metabolomics revealed substantial changes in the metabolite profile of the fermented soy whey compared to the unfermented control. Notably, fermentation led to an increase in angiotensin-converting enzyme inhibitory activity, suggesting potential health-promoting properties.	42
Metagenetics and Metabolomics	This study conducted a time-series analysis of WK fermentation using metabolomics and amplicon sequencing. Correlation analysis enabled the identification of microbial taxa positively or negatively associated with various metabolites, including potentially bioactive compounds.	41
Volatilomics	This study investigated the volatilome of WK fermentations using either white sugar or dried figs as substrates. Volatilomics revealed differences in volatile compound production depending on the carbon source, providing insights into how substrate selection shapes flavour development in WK.	43
Metagenetics and Volatilomics	This study analysed eight distinct WK fermentations using volatilomics and amplicon sequencing and identified 41 volatile compounds and uncovered correlations between specific volatiles and microbial taxa potentially involved in their production.	44
Metatranscriptomics	This study employed metatranscriptomics to investigate transcriptional activity in WK and milk kefir fermentations. Despite notable differences in microbial community composition between WK and milk kefir, functional analysis revealed that major biological processes were conserved. Specifically, 70% of KEGG Orthology entries for yeast genes and 36% for bacterial genes were shared between the two systems.	45

between WK-associated yeasts and LABs[36]. These beneficial interactions were further supported by proteomic analyses[26,49].

Together, these features position WK as a versatile platform for investigating microbial succession, functional redundancy, ecological stability, and metabolic modelling. WK is particularly well-suited for integrated multi-omics and genome-scale metabolic modelling approaches, offering direct relevance to food biotechnology, synthetic ecology, and microbiome engineering. As omics-driven insights continue to grow, WK emerges as a compelling model system for studying fermented food microbiomes, microbial community assembly, and the development of functional foods.

Revolutionizing fermented food research with multi-omics: opportunities and challenges

Limitations of single-omics approaches and the need for multi-omics integration

While single meta-omics studies can provide valuable insights into WK microbial communities, they cannot capture the full complexity of microbial interactions and functional dynamics in isolation. For example, metagenetics (also known as metabarcoding, metataxonomics, amplicon sequencing, and sometimes amplicon metagenomics) amplifies and sequences taxonomic marker genes (e.g., 16S ribosomal RNA [16S rRNA] for bacteria and Internal Transcribed Spacer [ITS] for yeast) to identify microbial community composition[50]. This approach is widely used to monitor the taxonomic composition of WK liquid and grains at different time points during fermentation[41,51,52]. As it normally uses short-read sequencing, metagenetics typically reveals genus-level taxonomic composition but lacks species-level resolution and functional insights. In contrast, shotgun metagenomics sequences the total DNA from a sample to provide comprehensive taxonomic and functional gene profiling. This approach can reveal species-level composition, including the detection of rare and novel taxa[29,37]. It can also reveal key functional genes in the community, such as genes involved in carbohydrate metabolism, exopolysaccharide (EPS) synthesis, stress resistance, and probiotic traits in WK, helping to infer the role of each microbe in the WK community[37]. However, both approaches lack information relating to cell viability, active gene expression, and metabolic activity, which limits the understanding of microbial functional activity. Furthermore, shotgun metagenomics of foods can be subject to issues such as inherent compositionality of the sequencing, presence of contaminating non-microbial reads originating from the food matrix, and unavailability of adequate reference databases, specially where the WK species are relatively novel.

Metatranscriptomics sequences the RNA of a sample to track all the actively expressed genes and reveal real-time functional activity of microbial community during the fermentation. This approach helps to identify active taxa and key metabolic functions in the community[53]. For example, metatranscriptomics analysis of WK fermentation at 0-h and 24-h timepoints revealed highly expressed genes related to the utilisation of sucrose, glucose, fructose, and vitamin B biosynthesis, with the most abundant transcripts from *Lactobacillus*, *Bifidbacterium*, and *Saccharomyces* species at 24-h timepoint[45]. As transcript-mapping is aided by sample-specific databases, metatranscriptomics and either genome-resolved metagenomic or whole-genome sequencing data of strains from the same samples should ideally be paired to accurately interpret the transcriptional activity[54], while metaproteomics and metabolomics may be further applied to confirm whether transcripts are translated into functional proteins or lead to the production of metabolites.

Metaproteomics is based on analysis of the entire protein content of the community and provides insights into which microbes are metabolically active. This is achieved through the extraction and identification of proteins using mass spectrometry, which can also be used to study protein-protein

interactions. However, metaproteomics relies on an adequately complete protein database for taxonomic and functional assignment and is unable to capture the flux of gene expression and metabolite production, thus providing an incomplete understanding of microbial community function[55]. Therefore, metaproteomics should be paired with metagenomics to identify the taxa producing these proteins. Applying this approach to traditional fermented foods such as WK can provide comprehensive information regarding the enzyme activity, functions, and microbial interactions during the fermentation[55].

Metabolomics, including volatilomics, identifies and quantifies biochemical compounds using techniques, including High-Performance Liquid Chromatography (HPLC), Gas Chromatography-Mass Spectrometry (GC-MS), Liquid Chromatography-Mass Spectrometry (LC-MS) and Nuclear Magnetic Resonance Spectroscopy (NMR). These techniques allow both targeted and untargeted detection of key metabolites, including volatile compounds and potential health-promoting metabolites and enable detection of compounds responsible for the flavour, aroma, and bioactivity of fermented foods[29,41]. Metabolomics has successfully identified and quantified major metabolites in WK, including lactic and acetic acids, ethanol, acetoin, mannitol, flavour compounds, vitamins, EPS, and other bioactive molecules, including novel compounds[29,41,56]. However, this approach alone cannot reveal which microbes produce specific compounds or how their production varies as the community shifts during fermentation. Hence, metabolomics data should be accompanied by metagenomics from the same samples to link the metabolites with their associated taxa and ideally with metatranscriptomics data to elucidate how metabolite production varies with community activity over time.

These limitations of individual meta-omics approaches necessitate an integrated meta-omics study to fully decipher the genetic potential, gene expression patterns, protein functions, and metabolite profiles of fermented food microbial communities[23,57]. Although, to our knowledge, a multi-omics analysis combining all meta-omics approaches has not yet been applied to WK, such a study has the potential to significantly improve the current understanding of the molecular complexity of WK communities.

A roadmap to integrated multi-omics study of fermented foods using water kefir as a model system

Before conducting an integrated multi-omics study of WK microbial communities, some considerations should be taken into account. It is crucial to select appropriate representative samples of WK from different timepoints of fermentation and include an adequate sample size to ensure that the data provides meaningful biological interpretation. The microbial community composition between the WKGs can vary based on their geographical origin or fermentation substrates. It is also demonstrated that grain-to-grain variation in microbial composition shape the succession dynamics and metabolic output (e.g., organic acid profile, ethanol levels and volatile profiles), even under similar fermentation conditions[29]. Hence, it is important to analyse multiple WKGs across the fermentation timepoints to capture representative community dynamics and metabolic outputs. To estimate the adequate sample size, power analysis can be performed using tools like MultiPower, which is an R package that performs joint power analysis across multiple omics layer by taking the variability, effect sizes, and omics-specific differences into considerations and optimise the sample size and experimental design for an multi-omics analysis[58]. Additionally, standardised sample handling techniques should be used, tailored to each meta-omics technique. In multi-omics studies, samples are often split into multiple aliquots for different omics experiments, as different biomolecules require different extraction procedures and analytical instruments, which may cause data heterogeneity. Therefore, a consistent sample handling technique should be used, such as co-extraction techniques to enable simultaneous extraction of biomolecules for different omics layers from the same sample. For example, DNA and RNA can be simultaneously co-extracted using a solvent-based method or commercial co-extraction kits[59]. Similarly, metabolites and proteins can be co-extracted using biphasic solvent systems, which can separate proteins and metabolites into distinct phases, enabling their simultaneous extraction[60,61]. Finally, independent biological replicates, technical replicates and appropriate control samples (e.g., unfermented controls and negative controls) should be incorporated in the study for reliability and comparability[62]. Unfermented controls can be obtained by preparing and incubating sterilised substrates identically to fermented samples, but without the addition of an inoculum. Some computational considerations should be considered, including early planning of data requirements such as sequencing depths for DNA and RNA, whether to use targeted or untargeted metabolomic data, bioinformatic pipeline selection, and access to adequate infrastructure. Additionally, determining computational requirements for the analysis of each omics data or integrated analysis must also be taken into account[2]. After each omics dataset is generated, robust data preprocessing is necessary prior to integrated analysis. Omics data are often incomplete or contain errors due to the technical constraints and limitations of analytical techniques. Hence, raw data typically needs quality control, normalisation, filtering, and transformation to ensure compatibility for integrated analysis[63]. Once datasets are pre-processed, they can be integrated and analysed using appropriate integration strategies and tools (Fig. 1).

Based on the research goals and data types, integration strategies can broadly be divided into three main types: sequence- and genome-centric integration, data fusion, and knowledge-based integration. Sequence or genome-centric integration map metatranscriptomics and metaproteomics data to metagenomics and provide insights into gene expression and protein activity within the community. Data fusion integration approach links different meta-omics datasets, such as metagenomics and metabolomics, using statistical methods to identify meaningful biological connections between them. For example, network-based integration is a data fusion approach, where nodes represent the biological features such as genes, proteins, and metabolites, and edges represent the coordinated changes. This approach can integrate omics data and identify the biological features that change in a coordinated manner across the samples or conditions[64]. Knowledge-based integration maps different meta-omics data onto known metabolic pathways to model community function in-silico, allowing inference of active pathways and interactions within a microbial community[64,65].

Several tools implementing integration strategies and analysis techniques have been developed for an integrated multi-omics study. Each tool is designed to support different combinations of omics data with distinct integration strategies. Table 2 summarizes some potential multi-omics integration tools, outlining their principles, compatible omics data types, and their respective advantages and disadvantages.

Key challenges of multi-omics integration and circumvention strategies

While an integrated multi-omics approach facilitates the understanding of fermented food microbial communities by deciphering interactions between the microbiota, substrates, and the environment, their application remains limited owing to technical and logistical challenges, as well as the cost of data generation. This has led to reliance on single meta-omics approaches in the majority of recent studies. Here, the likely challenges associated with the integrated meta-omics approach and possible solutions to these challenges are discussed.

Data heterogeneity renders the integration challenging as each meta-omics platform generates data in different formats, scales, and units. This data heterogeneity results in uneven statistical power across the omics datasets, which may cause bias. This necessitates consistent sampling, appropriate sample handling, data preprocessing, and normalisation before the integration[66]. To maintain consistent statistical power, integration tools like MOFA[67] can process partially overlapping datasets and maintain statistical power.

The compositionality of data also makes the integration challenging. For example, metagenomics, metatranscriptomics, and metaproteomics describe relative abundances of DNA, RNA, and proteins, respectively, rather than absolute abundances, which can impact the

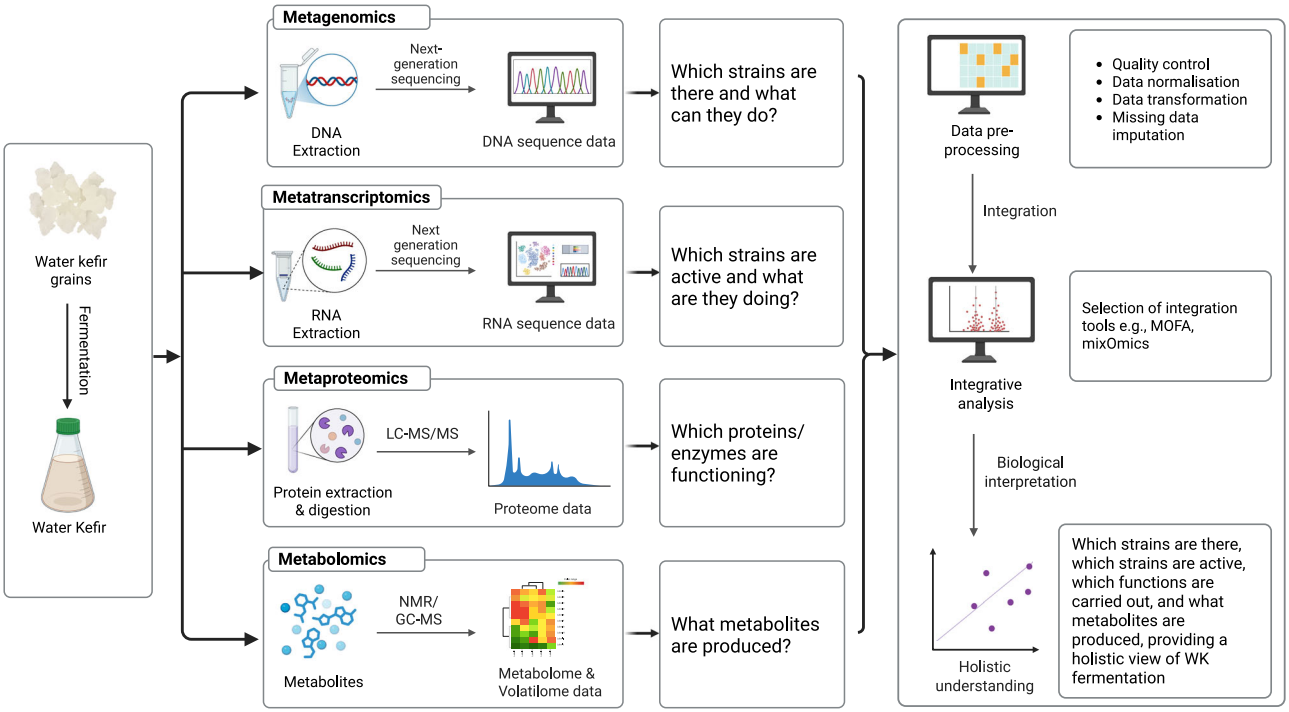

Fig. 1 | Conceptual framework for an integrated multi-omics study of fermented food microbiome, illustrated using the water kefir microbial community as a model system. The schematic diagram illustrates the conceptual framework for a multi-omics analysis of fermented food microbiome to achieve a holistic understanding of the microbial interactions, fermentation dynamics, and metabolic outputs. This figure used WK as a model system, and the process starts with sampling from the fermentation matrix and generation of different meta-omics datasets, followed by data preprocessing and integrative analysis of all omics data together using an appropriate multi-omics integration tool. (Created in BioRender. Khan, A. (2026) https://BioRender.com/dcmx99w).

integration. This can also complicate the statistical analysis and mislead the interpretation[2]. To overcome these challenges, experimental techniques can help to provide absolute abundance. For example, techniques like quantitative polymerase chain reaction (qPCR), plate counts, and flow cytometry, when combined with metagenomics, can provide an absolute abundance of DNA or microbes. However, these experimental techniques are time-consuming and laborious. A relatively new method, the use of synthetic DNA spike-in, is much more efficient for the absolute quantification of metagenomics data[68]. Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) and RNA spike-in can be used for the absolute quantification of transcripts[69–71]. Conversely, the use of isotope-labelled peptides in mass spectrometry can accurately measure the absolute abundance of proteins[72]. Additionally, tools that can handle compositional data should be used for integrative analysis. For instance, coda4microbiome[73] is an R package that utilizes Compositional Data Analysis (CoDA) techniques to make the compositional data suitable for statistical analysis.

Technical limitations of each omics platform are another challenge for integrated analysis[74]. For example, metagenomics can overestimate taxa due to the detection of DNA from dead cells, and metatranscriptomics may fail to detect the low-abundant or transient transcripts. Conversely, metabolomics and metaproteomics often does not detect all the metabolites or proteins due to the limitations in sample preparation, extraction methods, and instrument sensitivity. Some rare metabolites also may be detected in only a subset of samples at a specific time point. These technical limitations can lead to data sparsity and missing data, which can affect the integrative analysis. Careful experimental design can minimize technical limitations. For example, sample collection at multiple time points throughout the fermentation increases the chance of capturing low-abundance and transient biomolecules. Incorporating biological and technical replicates can enhance the data reliability and help in accurate biological interpretation. Simultaneous co-extraction of biomolecules from the same samples ensures consistency across omics datasets and reduces the chance of missing data. Additionally, statistical imputation techniques can be applied, which use a mathematical model to predict and estimate the missing value based on the observed patterns within the existing data[75]. Multi-omics integration tools, such as MOFA[67], use statistical imputation to complete the datasets before the integrative analysis.

The limited availability of up-to-date reference databases is a key challenge for multi-omics integrative studies. Many microbial genes, proteins, and metabolites of traditional fermented food communities like those of WK are not well characterized to date and are underrepresented in publicly available databases. This can hamper accurate functional assignment of genes, proteins, and metabolites across omics datasets and mislead the integrative analysis[76]. This underlines the importance of expanding and maintaining reference databases, particularly tailored for traditional fermented food communities like WK. Additionally, development of a sample-specific custom database for WK can provide complementary advantages by better capturing strain specific functional diversity.

Integrative analysis may face some key computational challenges, such as tracking molecular components across datasets, ensuring reliable data normalisation, and effective visualisation of results[77]. The lack of standardised workflow is a significant challenge. Aside from more generalised undertakings like the normalisation and visualisation of multi-omic datasets, a key challenge remains that of accurately integrating and labelling genes, proteins and compounds across the various omics layers in the food sample. For example, different approaches to taxonomic profiling, such as MetaPhlAn 4[78] and Kraken 2[79] rely on different classification strategies and databases, and delivers variable taxonomic assignments and performance with respect to strain separation. This principle extends to functional annotation of genomic and transcriptomic datasets, as well as the analysis of proteomic and metabolomic datasets. For example, across different metabolomic platforms, the same chemical compounds are sometimes listed under different names and therefore may require manual curation to allow comparisons between datasets. These inconsistencies can affect the multi-

Table 2 | Summary of potential tools for integrative meta-omics analysis outlining their principles, compatible omics data types, and respective advantages and disadvantages

Tool	Description of Integration Strategy	Omics Data Integrated	Advantages	Limitations	Reference
IMP	A reproducible pipeline for the integrated and reference-independent analysis of metagenomic and metatranscriptomics data.	Metagenomics and Metatranscriptomics	It is an open-source pipeline for standardized, automated, flexible, and reproducible large-scale integrated analysis. Also support robust read preprocessing, automated binning, as well as genomic signature-based visualizations.	It is computationally intensive and cannot integrate metaproteomics and metabolomics.	201
gNOMO	It can combine metagenomics, metatranscriptomics, and metaproteomics data from the same sample and create a proteogenomic database for integrated analysis.	Metagenomics, Metatranscriptomics, Metaproteomics	It is an automated, reproducible, and easily configurable pipeline for meta-omics data preprocessing, integrated analysis, and output visualization. It supports non-model organisms lacking well-annotated reference genomes or proteomes and enables the study of microbial communities within their biological and ecological contexts.	It does not support integration of metabolomics. Requires high computational resources and depends on other tools such as Snakemake.	202
gNOMO2	A comprehensive and modular pipeline for multi-omics analysis that used three different integration approaches, including proteogenomic database-based integration, differential abundance-based integration, and joint visualization-based integration	Metataxonomics, Metagenomics, Metatranscriptomics, Metaproteomics	It is relatively easy to use and improved compared to its predecessor, gNOMO. It can process raw data for integrated analysis and easily visualise the results	It requires high computational resources and expertise.	203
mmvec	A neural network that can predict the entire metabolic abundance profile from the microbial sequences and can learn the co-occurrence pattern between microbes and metabolites.	Metataxonomics, Metagenomics, Metabolomics	It is designed to identify the non-linear relationship between microbes and metabolites using deep learning and handle the compositional microbiome and metabolomics data.	It does not support metatranscriptomics and metaproteomics data and computationally intensive for large datasets.	204
NMFGOT	An unsupervised learning framework to integrate metagenome and metabolome data to identify a nonlinear relationship between microbiome and metabolites.	Metagenomics, Metabolomics	It can perform many functions, including sample clustering, visualisation, and downstream biological analysis.	It does not support metatranscriptomics and metaproteomicsdata.	205
MOFA	A statistical method for integration of multiple meta-omics data in an unsupervised fashion.	Metagenomics, Metataxonomics, Metatranscriptomics, Metaproteomics, Metabolomics	It can handle heterogeneous and missing data.	It is originally designed for a human dataset, hence no built-in features for microbial community data.	67
mixOmics	An R package designed for multivariate analysis of various omics datasets.	Metagenomics, Metataxonomics, Metatranscriptomics, Metaproteomics, Metabolomics	It allows statistical integration of various meta-omics datasets and allows data exploration, dimension reduction, and visualisation.	It requires preprocessing of input data (e.g., data normalisation).	206
mCIA	An exploratory data analysis method that identifies co-relationships between multiple omics datasets	Metagenomics, Metatranscriptomics, Metaproteomics	It can handle heterogeneous datasets and available as an R package.	It requires data normalisation.	207
MiBiOmics	A network-based approach for integrated analyses of multi-omics data.	Metagenomics, Metatranscriptomics, Metaproteomics, Metabolomics	It is a user-friendly web-based application and allows multi-omics data integration, exploration, and visualisation.	It is limited to the integration of up to three omics datasets at once and limited customisation.	208
PALM	A computational pipeline that integrates multi-omics data from longitudinal microbiome studies using dynamic Bayesian networks.	Metagenomics, Metatranscriptomics, Metabolomics	It allows integration of time-series multi-omics datasets.	It is computationally complex and does not support metaproteomics data.	209
OmicsAnalyst	A data-driven approaches for multi-omics integration, analysis, and visually exploring their results in a clear and meaningful manner.	Any normalised meta-omics data	It is a web-based, user-friendly tool with a simple interface and allows rapid statistical analyses and visual exploration.	It is not suitable for time-series data and requires appropriate preprocessing for all omics data types.	210
ViMO	It has separate workflows for metagenomics, metatranscriptomics, and metaproteomics, but also allow integrative analyses.	Metagenomics, Metatranscriptomics, Metaproteomics,	It is a relatively user-friendly tool and allows cross omics mapping and interactive visualisation.	It depends on other tools for multi-omics integrative analyses.	211

omics integration and eventually influence the prediction of metabolic interactions by GSMM and selection of strains, thereby affecting the reliability of the designed SynComs for biotechnological applications.

In this context, WK has the potential to offer a simplified model system to evaluate and mitigate these multi-omics integration challenges. WK offers advantages due to the functional redundancy between closely related strains that are observed in WK communities[29]. Minimal functional communities of fully sequenced WK microorganisms can produce omics datasets with less taxonomic noise and potentially streamline the data integration. It also offers an opportunity to systematically compare the analytical pipelines and reference databases to reveal how different pipelines influence the biological interpretation. A further advantage is the well-documented succession patterns seen in WK and corresponding community states at different points of the fermentation, enabling timepoint-driven analysis of minimal communities and with a view to gaining novel biological and mechanistic insights. Furthermore, while the metabolomic space is likely still large, it is likely to be greatly reduced when compared with more complex plant and animal-derived substrates seen in other fermented food matrices, as WK is generally composed of sucrose and water, with occasional addition of dried or fresh fruits (e.g., dried figs, raisins). Non-omic data relating to WK could be included in integration models, such as sensory profile, degree of carbonation, or batch number, where alternative substrates are used. This approach could further aid in the biological interpretation of the model food system, connecting -omics data to relevant biotechnological attributes.

Finally, use of data reduction techniques, development of user-friendly tools and a standardized workflow for integrative analysis can improve efficiency and reproducibility of multi-omics analysis, particularly in controlled system such as WK. Advanced integration tools (e.g., machine learning or network-based methods) may detect complex patterns across omics datasets; however, it is often challenging to extrapolate these patterns to their biological activity. Combining advanced tools with simpler well-understood methods like the statistical correlation method can make the biological implications more clear[76]. Therefore, through careful experimental design, proper data handling, and selection of the appropriate integration tools, many major hurdles of multi-omics integration for WK may be overcome.

Genome-scale metabolic modelling for the rational design of microbial communities in fermented foods

GSMM has received attention in the context of the design of SynComs. It allows the prediction of microbe-microbe interactions, and microbe-habitat interactions[8]. These insights facilitates the rational design and optimisation of SynComs through identification of optimal species composition and metabolic pathways, tailored for a specific outcome[80]. For instance, GSMM has been applied to develop a plant growth-promoting community for *Campos rupestres* using 270 metagenome assembled genomes (MAGs), which was less complex and improved crop productivity under stress conditions[20]. Similarly in fermented foods, GSMM was used to design 120 SynComs for sourdough with different combinations of LAB and yeast and predicted their growth dynamics, CO_2 production, and volatile profiles. Experimental validation confirmed enhanced fermentation performance with the SynComs that predicted to have high yeast growth and CO_2 production, compared to the SynComs that predicted to have low yeast growth and CO_2 production[21]. However, this study compared the fermentation performance across the SynComs without directly comparing with natural sourdough communities. Similarly, in cocoa bean fermentation, GSMM-guided defined SynCom was successful in reproducing the key metabolic and sensory attributes of fine flavour chocolate under controlled fermentation conditions, supporting its functional equivalence to natural fermentation[81].

Microbial communities such as those in WK contain diverse species of microorganisms with different functions throughout the fermentation process. However, not all the species present in a community contribute to a given functionality. Therefore, it is possible to design a function-specific minimal community, which is a subset of the larger community. While other fermented foods (e.g., milk kefir, sourdough) have been extensively studied using omics approaches and SynCom studies[82,83], many such systems are less flexible in the use of alternative substrates. For example, milk kefir relies on milk and sourdough on flour-based substrates. This limits their manipulation and renders them less flexible for controlled analysis of how substrates influence the community function and metabolic outputs. Conversely, WK communities are relatively simple, yet functionally diverse, with higher experimental flexibility in accommodating a wide range of plant-based substrates for fermentation, including substrates from food waste streams. This makes WK microbial communities a perfect paradigm to design GSMM-based minimal SynComs.

A complete workflow of microbial community design for water kefir using genome-scale metabolic modelling

The complete workflow of GSMM-guided design of microbial communities for WK can be divided into six steps: (1) defining desired and undesired metabolites (2) selection of candidate strains and their genome sequencing (3) genome annotation and GEM reconstruction of candidate strains (4) computation of collective metabolic capabilities and cooperation potential, (5) identification of minimal communities required for the target functions, (6) Experimental validation and iterative refinement of minimal communities.

Defining desired and undesired metabolites. The first step of GSMM-guided SynCom design for WK is defining the functional targets of the community. These functions may include the improvement of flavour profile and health-promoting properties, or the utilisation of alternate substrates such as plant-based milks or food waste streams. For the SynCom design, it is essential to define the desired functions in terms of target or non-target metabolites. Metabolomics data from traditional WK can help in identifying the key metabolites that must present to keep the basic characteristics of traditional WK, as well as to identify the metabolites associated with undesirable traits, such as off-flavours or safety concerns. For example, to enhance the flavour profile of WK, desirable metabolites may include lactic acid, mannitol, and some ester compounds (e.g., ethyl acetate), while undesirable metabolites may consist of compounds that can create off-flavour, such as sulphur compounds[84]. In terms of health-promoting properties, desirable metabolites could include short-chain fatty acids, vitamins, EPS, and gamma-aminobutyric acid while undesirable metabolites could include biogenic amines, such as histamine and tyramine that can cause adverse health effects[85]. Depending on the GSMM approach, it may be necessary to limit the scope of metabolites to those that can be predicted by the tools. For example, bioactivity of EPS such as levans may depend on their structural conformation, which cannot be accurately predicted by current metabolic modelling techniques. Esters such as ethyl acetate may be desirable or unpalatable depending on their concentration and co-occurrence with other volatile molecules[86].

Selection of candidate strains and their genome sequencing. Once the target metabolites are defined, the candidate bacteria and yeast from traditional WK community should be selected to include them in the GSMM-based SynCom design. When designing SynComs for WK, core microbial member selection is crucial, as these core members are necessary to mimic the fermentation dynamics and sensory properties of traditional WK. If the core members are not present in the SynCom, it can lead to the production of a beverage that lacks the basic characteristics of a traditional WK and can also impact microbial interactions. Metagenomics data from traditional WK fermentation can help to identify the core taxa, offering guidance on which strains should be cultivated for inclusion in the SynComs[29]. Large-scale metagenomic study of WK communities suggested that the communities can be assembled into multiple, distinct community states at the genus level, indicating that different combinations of core taxa may result in functionally similar but compositionally different WK fermentations[29].

From a practical perspective, the first consideration should be the selection of cultivable strains. SynCom is only useful for WK fermentation if the required strains are cultivable in the laboratory, because non-cultivable microbes have limited technological use. A notable advantage of fermented food microbial communities is that they consist of predominantly cultivable microbes compared to microbial communities from other environments (e.g., soil, gut, sea)[87].

It is also crucial to consider the environmental and nutritional requirements for the selected strains. WK fermentation is usually performed at room temperature with a drop in pH to around 3.5-4.5. The selection of strains from the natural WK community is not an issue since they are already adapted to acidic conditions. However, strains from other environments could be selected based on their compatibility with the specific application. This can be achieved through microbial growth profiling under relevant fermentation conditions, such as pH and temperature. SynCom design for novel substrates requires finding metabolically compatible key strains that can efficiently metabolise the available nutrients in the substrates. Metabolic compatibility can be evaluated through several experimental approaches. For example, the Biolog Phenotype MicroArray can be used to assess novel substrate utilisation profiles, isothermal calorimetry can measure microbial metabolic activity, and growth profiling in the novel substrates can assess the performance under relevant fermentation conditions[88,89]. Regulatory guidelines may also limit the application of strains without a safe history of use in food in the development of SynComs (discussed further below).

Additionally, strains from different origins must be evaluated for any potential antagonistic activity with each other. For example, if strains from different WKG or strains from different fermented foods are used in the SynComs, they may show antagonistic behaviour towards each other. This antagonism may arise from the production of inhibitory compounds such as organic acids, bacteriocins, and hydrogen peroxide. Antagonism may also arise from the competition between microbes for essential nutrients and due to the presence of bacteriophages that may target a specific strain. If any antagonistic activity is present between the microbes of the SynCom, it may inhibit one or more microbes of the community, leading to the destabilization of the community. Experimental techniques including spot-on-lawn assays, well diffusion assays, co-culture experiments, and plaque assays can be used to identify any potential antagonistic activity between the selected strains[90,91].

Once the candidate strains are selected, their whole genomes need to be sequenced to adequate depth and quality metrics to allow their assembly using an appropriate algorithm or software[92–94]. Following the sequencing and assembly, these genomes can be used for assessment of candidate strains by screening genes associated to antimicrobial resistance, toxin productions and virulence factors. The process of strain selection often becomes an iterative process, as the selected strains may require replacement based on the in vitro assessment of safety, growth, metabolic compatibility, community stability, and susceptibility to bacteriophage infection.

Genome annotation and GEM reconstruction. Once the assembly of whole-genome sequences (WGS) of candidate strains obtained, they need to be annotated. It is a crucial step as the accuracy of genome annotation directly affects the quality of resulting GEMs[8]. Genome annotation first predicts the location of open reading frames and subsequently assigns them with metabolic functions. Typically, this is achieved by comparing sequences with databases of genes and proteins of known function. There are several annotation tools available that can be used for the genome annotation of both prokaryotes and eukaryotes found in WK. Tools for prokaryotic genome annotation that are currently available include Prokka[95], PGAP standalone[96], MicrobeAnnotator[97] and Bakta[98], which balance speed of execution with level of detail and degree of cross-referencing to various databases. Gene prediction in yeasts is complicated by the presence of genes containing introns, which may require additional training using known transcriptomic or proteomic data. Key eukaryotic gene prediction algorithms include GeneMark-ETP[99] and AUGUSTUS[100], which have been amalgamated into pipelines such as MAKER2[101], BRAKER2[102], and Funannotate[103]. The Yeast Genome Annotation Pipeline (YGAP)[104] is another tool which annotates open reading frames based on gene homology and synteny to known yeast species and is particularly useful where transcriptomic data is not available. Ultimately, manual curation of automated annotations using domain-level expertise may be required, particularly for non-model organisms.

Once genomes are annotated, Genome-Scale Metabolic Network (GSMN) are reconstructed for each individual candidate strain, which represents the entire set of biochemical reactions that an organism can carry out[105]. The GSMN can then be converted into GEM by translating the reconstructed network into a mathematical framework[8]. GEM reconstruction is a time-consuming process, but it can be expedited by automated and semi-automated tools (Table 3). These tools can automatically perform GEM reconstruction and certain tasks in the reconstruction process, such as genome annotation, generation of a draft GSMN, and gap-filling[106]. There are two main approaches of GEM reconstruction: bottom-up and top-down, which may be implemented in standalone tools or suites with additional features for curation and analysis of models[106]. The bottom-up approach starts with an annotated genome of an organism, from which a draft GEM is generated. Conversely, top-down approaches use a curated universal model as a template, from which pathways containing genes absent from the annotated genome are removed[3].

Draft GEMs often contain gaps due to incomplete annotation, and which result in missing reactions. Therefore, tools with gap-filling algorithms have been developed to improve the quality of GEMs[107]. Gap-filling algorithms are applied to both bottom-up and top-down draft GEMs to address missing reactions and dead-end metabolites[107], such as MENECO[108], fastGapFill[109], and DNNGIOR[110]. Manual curation of draft GEMs and automated gap-filling suggestions is often required to correct for missing or incorrect pathways, based on experimental data and expert knowledge for that taxon[111]. Recent tool development has also focused on methods to amalgamate GEMs produced by multiple pipelines into a single consensus model like GEMsembler[112], modelBorgifier[113],and mergem[114], as well as automated pipelines to identify putative errors in GEMs, such as MACAW[115]. While several tools allow direct community modelling using collections of GEMs, there are also wrapper tools such as MetaGEM[116], which implements CarveMe[117] within a full pipeline from metagenomes to community models.

Multi-omics integration is increasingly used to refine GEMs, such as enzyme-constrained approaches that are employed in GECKO[118]. Proteomics data offers relative quantification of proteins and enzymes, which enable understanding of functional roles of species, and the interspecies interactions within the community[119]. The integration of enzyme abundance data with GEM helps to refine the GEM by setting realistic limits on reaction rates, thus improving the accuracy of condition-specific flux predictions and enhancing the predictions of interspecies interactions. For example, integration of proteomics data with GEM of *Aspergillus niger* significantly improved the model's ability to predict metabolic fluxes under different growth conditions[120].

Transcriptomics data can also significantly improve metabolic prediction by GEM. By integrating gene expression data, static GEMs can be transformed into dynamic, condition-specific GEMs. This allows simulation of actual microbial metabolism under varying environmental conditions, such as nutrient changes, thereby providing insights into how environmental perturbations affect metabolic activities at both cellular and community levels[77].For example, Zampieri et al. integrated condition-specific gene expression data with GEM to predict real-time metabolic activity of anaerobic digestion microbial community[121]. This could be applied to WK production, to model community metabolism under different phases of microbial succession, including nutrient, pH and temperature conditions. For instance, yeast strains can vary significantly in their adaptability to osmotic stress and nitrogen limitation-conditions which are directly relevant to WK production.

Table 3 | Details of tools for GEM reconstruction and analysis

Tool	Underlying Approach	Genome Annotation	Gap-filling	Community modelling	Database	Reference
Pathway Tools	Bottom-up	No	Yes, via MetaFlux module	Yes	Metacyc[212]	213
ModelSEED	Bottom-up	Yes, via RAST[214]	Yes	Not directly	ModelSEED[215]	216
ModelSEED 2.0 (available through KBase)	Bottom-up	Yes, via RAST[214]	Yes, highly conservative and limited by organism-specific templates	Not directly	ModelSEED[215]	217
CarveMe	Top-down	No	Yes	Yes	BiGG[218]	117
gapseq	Bottom-up	No	Yes, using network and sequence homology to reference database and proprietary Linear Programming-based algorithm	Yes	Own databases, built on UniProt[219] and TCDB[220] as well as the ModelSEED database[215]	107
RAVEN Toolbox 2.0	Bottom-up	No	Yes	No	MetaCyc[212] and KEGG[221]	222
Merlin 4.0	Bottom-up	Yes, via plugins	Yes, via plugins	Not directly	KEGG[221], ModelSEED[215], UniProtKB[163], TCDB[220]	223
Reconstructor	Bottom-up	No	Yes, via parsimonious flux balance analysis (pFBA)[224]	Not directly	Modified ModelSEED[215] database	225
AuReMe	Bottom-up	Yes, via Pathway Tools outputs	Yes, via Meneco[108]	Yes, via Metage2Metabo[126]	BiGG[218], ModelSEED[215], MetaCyc[212], KEGG[221]	226

Step 3: computation of metabolic capabilities and cooperation potential within the community. After reconstruction of GEMs, they can be analysed to predict the metabolic behaviour of individual microbes and their potential interaction(s) in the community. The aim is to understand the individual and collective metabolic potential of species and gain insights into the complementarity of metabolic pathways within the community, which ultimately aids the SynCom design. Several approaches have been developed for GEM analysis, including a constraint-based approach and a graph-based approach as described below[6].

A majority of tools adopt a constraint-based approach via Constraint-Based Reconstruction and Analysis (COBRA) algorithms (Table 4), which use stoichiometric coefficients to describe the proportions of reactants and products in the GSMN. This approach allows quantitative prediction of metabolic genotype-phenotype relationships[7]. The most commonly used COBRA method is Flux Balance Analysis (FBA)[122], which assumes steady-state conditions with constant uptake (such as found in continuous flow production systems) and progression of each pathway at equal rate. Dynamic flux balance analysis (dFBA)[123] extends FBA by incorporating uptake kinetics, accounting for temporal changes in metabolite concentration as occur in batch systems, which allow understanding of microbial interactions such as cross-feeding, competition, and succession within the community under dynamic conditions. However, COBRA-based methods require highly accurate GEMs and do not scale linearly, making them only suitable for communities with up to ten members, as larger communities significantly increase the computational demand[6].

Another approach for the analysis involves graph-based approaches, where GSMNs are examined in terms of their topology. It is a complementary alternative to quantitative constraint-based methods, which de-emphasise stoichiometry and use GSMNs to provide insights into microbial interactions and ranges of producible metabolites at the community-level[124,125]. Consequently, it is suitable for analysing non-model organisms, GSMNs reconstructed from draft genomes or MAGs, and larger community-level metabolic models. However, this approach deprioritises reaction stoichiometry, thus producing less accurate predictions of metabolic fluxes[126,127]. Network expansion algorithms, as implemented in Metage2Metabo, for example, use a graph-based approach to predict the set of metabolites producible based on the composition of the user-defined growth medium, also known as seeds[124,126].

From a nutrition perspective, GSMM also offers the potential to analyse GEMs from food microbes for potential impacts on the host. This has been explored in the cattle rumen[126] and in the design of probiotics for human consumption[128], wherein microbial GEMs or simply the metabolites they are predicted to produce, are integrated with a metabolic model for the host.

Step 4: identification of minimal synthetic communities and key species required to produce target metabolites. Minimal SynComs are of interest for traditional fermented foods like WK because it simplifies the pool of strains from a biotechnological perspective and allows more robust and predictable fermentations[129]. After the prediction of individual and collective metabolic capabilities of the candidate strains, the next step is to identify the key strains that can collectively produce the predefined target metabolites. Several tools can aid the identification of keystone species and the development of minimal communities. For instance, network expansion-based tools CoMiDA[130] can identify a minimum set of species that can collectively produce a set of desired target metabolites from a user-defined set of available substrates. On the other hand, MiSCoTo[131] applies constraint-based logic to identify the combination of species that can collectively produce the target metabolites from predefined substrates. Both tools can classify species into essential and alternative symbionts. Essential symbionts are the strains that must be present in the community to produce the target metabolites, while alternative symbionts are the strains that can be substituted with each other in the communities. Inclusion of multiple alternative

Table 4 | Overview of main approaches and tools for GEMs analysis to predict collective metabolic capabilities and cooperation potential in microbial communities

Approach	Description	Tool	Implementation	Reference
Constraint-based approach (COBRA)	This approach converts GSMNs into mathematical models, also known as GEMs, relying on stoichiometric matrix and constraints, such as mass balance, enzyme capacity, and nutrient availability. This approach is used to quantitatively predict the metabolic fluxes and the interspecies interactions under a defined environmental condition. Commonly used COBRA-based algorithms are flux balance analysis (FBA), dynamic flux balance analysis (dFBA), and flux variability analysis (FVA). However, the COBRA-based approach requires highly accurate GEMs, which are challenging, particularly for less characterised species. Additionally, current COBRA-based tools can only be applied to communities with few members, as the computational demand increases exponentially with the number of members in the community.	COBRA Toolbox	Multiple algorithms available, including FBA, dFBA, FVA etc.	227
		SteadyCom	Plugin for COBRA Toolbox; Generalisation of FBA incorporating specific growth rates and their time-averaged equality.	228
		Microbiome Modeling Toolbox 2.0	Plugin for COBRA Toolbox; Predicts microbial interactions and metabolic fluxes from microbiome data.	229
		OptCom	FBA applied to individual species, as well as community-level exchange.	230
		DyMMM	Plugin to COBRA Toolbox; Extension of dFBA algorithm; models small competitive and cross-feeding communities but is unsuitable for large and/or mutualistic ones. Highly accurate GEMs are required.	231
		DFBAlab	A MATLAB-based implementation of the LP feasibility problem with lexicographic optimization. It scales efficiently based on the number of species in the community but requires high-quality GEMs for accuracy.	232
		d-OptCom	A dynamic version of OptCom; predicts metabolic interactions and community composition changes in small, well-characterised communities.	233
		SMETANA	Mixed-integer linear programming; Evaluates possible interspecies interactions between community members via interaction scores for metabolic exchanges.	234
Graph-based approach	This approach represents GSMNs as graphs, where nodes are metabolites or reactions and edges are their relationships. Two main methods are used under the graph-based approach. Network expansion algorithms determine producible metabolites from a defined set of nutrients, while the network topology methods explore metabolic complementarities. This approach is more robust to incomplete/inaccurate GSMNs and can be applied to complex communities.	MeneTools	Network Expansion tools that predict producible metabolites by a microbial community from a set of compounds, given as growth medium or seeds. Implemented in Metage2Metabo.	226
		NetSeed	Network Topology; A web-based tool also available as open-source Perl module that can identify minimal set of seed compounds required to activate a specific metabolic pathway.	235
		NetCooperate	Network Topology; A web-based tool designed to predict microbe-microbe and microbe-host interactions and provide insight into their cooperative potential. It can be used for both small and large communities.	236

symbionts offers functional redundancy, which allows the design of robust communities that are resilient to environmental perturbations.

Experimental validation and iterative refinement of minimal synthetic communities. After the design of minimal SynComs, they need to be experimentally validated through trial fermentations, incorporating microbiological assay (e.g., viable colony forming unit count), physicochemical characterisation (e.g., pH monitoring) and meta-omics approaches (e.g., metagenomics, metabolomics, volatimomics). This step is crucial for the validation and refinement of the SynComs, following the Design-Build-Test-Learn (DBTL) cycle approach[132]. Trial fermentations should be performed using the substrates that were applied to GSMM and the fermentation samples then need to be analysed to ensure the SynComs function as intended under the real-world conditions (e.g., the production of the desired metabolites). However, during the experimental validation some ecological factors must be taken into considerations. For example, difference in microbial growth rates, initial inoculation densities, and the order of strain introduction can significantly affect the community stability and functions. In the traditional WKGs, microbes are spatially organised within the polysaccharide matrix, creating distinct ecological niches which led to the temporal succession of activity[133]. In contrast, when using SynComs all the strains are typically inoculated at the start and rely on their natural succession, with diffusion of intermediate metabolites unconstrained by the grains. It remains to be explored whether inoculation of these SynComs will lead to grain growth over time. However, inoculation of individual strains may provide additional advantages in commercial WK fermentation over the grains, as it allows more controllable and reproducible fermentation through the inoculation of defined strains at precise ratios (e.g., controlling ethanol production by modulating yeast abundance).

Omics-based investigation can be used to track the fermentation using the SynComs, such as metagenomics to track the community composition and metatranscriptomics to analyse gene expression patterns during different stages of fermentation[45,134]. This enables the verification of community composition and the activity of key metabolic pathways that were predicted in silico. Conversely, metaproteomics allows relative quantification of functional proteins and enzymes that drive the fermentation. Finally, metabolomics can identify and quantify the biochemical compounds produced in the fermentation with the SynComs. Targeted metabolomics should be performed to quantify the predefined target metabolites used to design the SynComs. This can validate the prediction of GSMM, ensuring that the target metabolites are indeed produced during the fermentation. A comprehensive untargeted metabolomics study should also be performed to find out the complete metabolite profile that positively or negatively contributing to the sensory and functional properties of the beverage. Characterisation of these metabolites provides a thorough understanding of the chemical outcome of WK fermentation using the designed SynComs. Integration of these omics data can further help to follow the microbial succession and track the changes in metabolic function during the fermentation, enabling validation and refinement of GSMM predictions, as recently demonstrated for WK[135]. Finally, the predation and horizontal gene transfer by bacteriophage and viruses can affect the stability of SynComs and introduce novel metabolic functions[136,137]. Therefore, metaviromics can also be utilised to uncover the bacteriophages and other viruses within the SynComs. Integration of all meta-omics data from these trial fermentations can provide a comprehensive molecular understanding of microbe-microbe and microbe-habitat interactions within the SynComs. This, in turn, allows the iterative refinement of SynComs tailored for the desired functional outcomes.

The visual synopsis of the steps involved in the GSMM-based SynCom design for WK is outlined in Fig. 2.

Why use genome-scale metabolic modelling to design microbial communities for water kefir?

GSMM has the potential to develop a stable, defined, and minimal SynComs tailored for WK fermentation, which can optimise the fermentation process and enhance the potential health-promoting effects and organoleptic properties of the beverage. Challenges associated with traditional WKG-based fermentation include excessive ethanol production, residual sugar due to incomplete utilisation, off-flavour formation, and variable microbial succession patterns driven by WKG structure and substrate composition. GSMM-guided SynCom design provide a route to address these limitation through functional customisation and provide additional benefits compared to conventional methods of community design (Fig. 3).

As mentioned above, traditionally WK fermentation relies on complex microbial consortia, while SynComs may be applied to establish defined, less complex cultures, thereby enhancing the controllability and reproducibility while simultaneously allowing the customisation of functional outcomes[138]. SynComs have already demonstrated their advantages in food fermentations by shortening fermentation time, enhancing the flavour profiles, modulating the nutritional profile, and improving the quality of foods[21,81,139–141]. Recent efforts have been made to develop SynComs for WK fermentation, which simplified WK communities through empirical strain selection with the aim of achieving stable and reproducible fermentation[47,48]. However, these studies primarily focused on establishing functional communities without providing in-depth analysis of the underlying microbial interactions.

GSMM-guided microbial community design offers a comprehensive understanding of microbial interactions and allows rational design of substrate-specific minimal communities in silico that may retain the key interactions of traditional WKG. It can predict the functional potential of a microbial community from its collection of genomes, enabling in silico simulation to achieve a more accurate prediction of microbe-microbe interactions and microbe-habitat interactions. This facilitates rapid prototyping, substrates-specific community design, optimisation, and functional customisation of the community. Therefore, GSMM-guided SynComs has the potential to reduce both time and cost associated with the extensive trial-and-error experimental approaches required in conventional community design to develop desirable traits in the final product[8].

GSMM-guided SynComs also have the capacity to be developed into patentable novel WK communities. Novel fermentation may be defined as the fusion of traditional fermentation practices with rational formulation of microbial communities to broaden and specifically direct the applicability of traditional fermentation[142]. The concept of novel fermentation has emerged as a response to converging factors, including increasing demand for diverse and functional fermented foods, increasing interest in health and nutrition, and advances in microbiome studies. Additionally, global challenges associated with food security, increasing food demand for population growth, efforts to look for novel solutions to global food shortage, and the shift towards more sustainable and plant-based diets has led to the search for innovative fermentation solutions. Several experimental strategies can be used to develop novel fermentations, including applying different substrates for diverse flavours, engrafting target species for desired traits, and assembling whole-community chimeras for synergy[142]. This concept of novel fermentation offers tremendous possibilities for innovation in the fermented food space. Several studies have explored WK fermentation with varieties of alternative substrates, including coconut, cherry, red plum, soy, quinoa beet, mandarin, and persimmon[42,143–151]. These fermentations using WKGs resulted in the production of novel WK or WK-like beverages with promising organoleptic properties.

However, scaling up these fermentations with undefined WKGs for industrial-scale production remains a challenge[24]. This is primarily because WKG microbial communities may be evolved to grow well in specific substrates, making them less predictable in novel substrates. Such complexity makes it challenging to predict microbial dynamics and interactions, and consequently to control the fermentation process in an industrial setting to achieve consistent products with desired sensory properties and

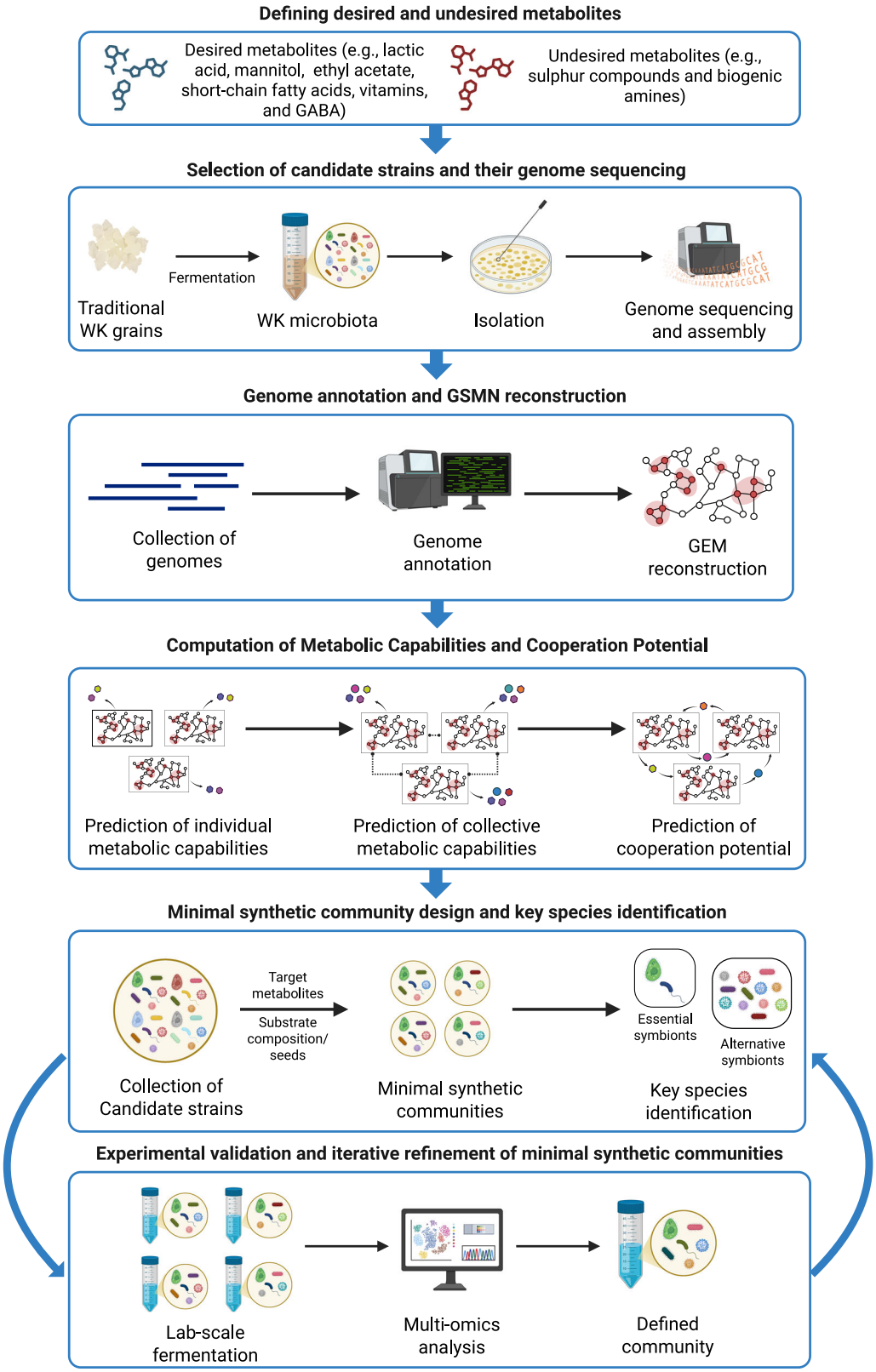

Fig. 2 | Workflow for the design and experimental validation of minimal synthetic community for water kefir fermentation using genome-scale metabolic modelling. The process starts with defining desirable and undesirable metabolites that must be present in the WK to achieve the target functions. Candidate strains are then cultivated from traditional WKGs, and their genomes are sequenced and assembled. Genome assemblies of each strain are then annotated using appropriate tools and are used for GEM reconstruction. Subsequently, the GEM are analysed to predict the individual and collective metabolic capabilities and compute the cooperation potential among the strains. Based on these predictions, minimal SynComs and key species are identified from the collection of candidate strains that can ensure the production of predefined target metabolites from the given substrates. Afterwards, these minimal SynComs are experimentally validated using lab-scale fermentation and multi-omics analysis, which allow iterative refinement of the community composition and fermentation conditions for the target functions. The outcome is a minimal defined microbial community with the potential of industrial-scale production of water kefir. Created in BioRender. Khan, A. (2026) https://BioRender.com/s4d929p.

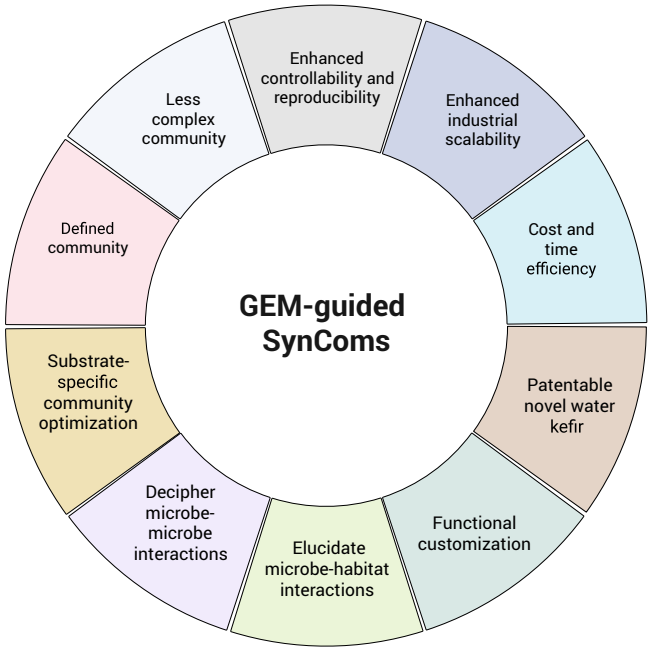

Fig. 3 | Advantages of genome-scale metabolic modelling for rational design of microbial communities in water kefir fermentation. This figure illustrates the key benefits of using GSMM in the rational design of microbial communities for water kefir fermentation. These benefits collectively make GSMM a powerful tool to design SynComs. (Created in BioRender. Khan, A. (2026) https://BioRender.com/t07bew3).

functionalities across batches[24]. GSMM may overcome the challenge by developing robust, substrate-specific defined communities with the potential to enhance industrial scalability.

Challenges and future directions in microbial community design using genome-scale metabolic modelling

Developing a functional and implementable microbial community for WK requires cultivation of different groups of key species from the traditional WKGs. Many species found in WK are difficult to cultivate due to the lack of optimised cultivation techniques. For example, Zanirati et al. detected some LAB, including *L. hilgardii*, *L. nagelii*, and *Oenococcus kitaharae* using culture-independent methods, but they were not detected using culture-dependent methods[152]. Additionally, AAB is another important component of traditional WKGs, but they are well known for their special ability to stay in a viable but not-culturable state, which rendering their biological evaluation challenging[153]. Analysis of MAGs of these difficult-to-culture microbes can provide valuable genomic insights on their primary metabolism, substrate utilization, oxygen requirements, antimicrobial resistance, which can facilitate the design of selective culture media and optimise their growth condition[154]. Additionally, high-throughput and machine learning-guided robotic culturomics platform integrated with rapid identification methods and advanced imaging can facilitate the isolation of these difficult-to-culture microbes from WK[155–157].

High-quality complete genome sequences are crucial for GSMM. Low-quality or incomplete genome sequence can lead to inaccurate and incomplete genome annotation, missing key pathways in WK microbes (e.g., sugar metabolism, EPS production, vitamin biosynthesis). Until recently, short-read and long-read hybrid assembly was needed to generate highly accurate microbial genomes[158,159]. However, recent developments in long-read sequencing technology, such as those developed by Pacific Biosciences and Oxford Nanopore, means one can now produce highly accurate, near-complete genome assemblies without the extra polishing steps required in hybrid assemblies[160–162].

Many WK-derived microbes are poorly characterised to date. For example, a recent study detected 18 novel species and identified many species that had not been cultivated previously[29]. This indicates that a substantial number of WK-derived microbes remain to be characterised in terms of their genetic diversity as well as experimental validation of functional annotations. Consequently, genes can be misannotated and labelled as hypothetical proteins with unknown functions. The use of less rapid annotation tools for functional annotation of predicted proteins, in combination with manual curation and possibly using curated databases like SwissProt[163], can improve the quality of GEMs in such cases[164]. The advancement of metagenome sequencing and assembly can produce highly accurate microbial MAGs that can guide the targeted isolation of newly discovered strains[165]. Additionally, targeted isolation and phenotypic characterisation of such strains can help to improve functional annotations and decipher their roles in WK fermentation.

The lack of information on gene function, including secondary metabolic pathways of microbes further complicates the GEM reconstruction process. WK fermentation involves a complex microbial community that produces a diverse range of secondary metabolites contributing to its flavour, aroma, and health-promoting effects. Biosynthetic gene clusters (BGCs) responsible for secondary metabolism evolve rapidly and often vary between species and strains, and their poor characterisation remain a major challenge in microbial biotechnology[166]. This can increase inaccuracies in annotations and downstream GEM reconstruction[5,105]. To improve the identification of secondary metabolic pathways, tools like antiSMASH[167] and DeepBGC[168] can help in identifying BGCs. For novel and poorly characterised BGCs, metatranscriptomics allows the initial identification of gene clusters that are active during fermentation and integration of culture-dependant genomics and metabolomics data with experimental validation can further confirm their activity[169].

GEM reconstruction of yeast poses additional challenges, especially for non-conventional yeast found in WK, such as *Brettanomyces bruxellensis*, and *Zygotorulaspora florentina*[170]. Yeast genomes are typically significantly larger than their bacterial counterparts, containing multiple chromosomes, introns, non-coding regions, and extensive repeat regions, thus requiring higher sequencing depth for better coverage and optimal genome assembly[171]. Long-read sequencing can significantly improve the genome assembly of yeasts[172]. However, genome annotation of yeast genomes remains a major bottleneck due to their complex genome sequence, unknown gene functions, and eukaryotic cellular structure, which affect the GEM reconstruction[104]. Enhancing GEM for non-conventional yeast requires comprehensive manual curation with the knowledge from biochemical databases and literature[173]. Experimental validation of pathways with culture-dependent omics can further improve the accuracy of GEMs of yeast[174].

Simulation of dynamic microbial interactions within the WK microbial community is also challenging. During the fermentation, WK community members engage in complex interactions, including cross-feeding, cooperation, competition through antibiotics and other inhibitory compounds, competition for resources, differences in growth rates, gene regulation such as quorum sensing, small molecule signalling, and environmental modifications such as pH change[28]. These interactions regulate microbial behaviour across spatial and temporal scales, which helps microbes to adapt to varying environmental conditions such as pH, oxygen levels, sugar types, and nutrient composition. GSMM often fails to capture these microbial interactions, hindering accurate simulation of interspecies metabolic exchange and community-level metabolic fluxes. Without capturing all relevant interactions, strains predicted to be compatible in silico may not perform as expected during real fermentation[175]. Experimental co-culture studies under controlled conditions and a combination of omics data from different time points, can capture these interactions that GEM alone cannot capture[36]. To date, efforts to integrate co-culture data with GEM have not yet been reported for WK. However, understanding of these complex interactions may provide an opportunity for iterative refinement of GEM

and experimental validation, enabling the design of a more stable community.

The composition of substrates plays a crucial role in WK fermentation. Both carbon and nitrogen sources are required for the microbial metabolism of the WK community[176]. It is extremely difficult to accurately and completely define the composition of the substrates used in GSMM, potentially biasing the analysis. Variability in nutritional composition of substrates due to the source and seasonality may affect community interactions. For example, it has recently been demonstrated that switching substrates changes the yeast dominance pattern in WK fermentations[32]. Unlike traditional WKG communities, GSMM-designed communities are simplified, highlighting the inter-dependency of these microbes. As a result, community interactions may fail if even one of the members behaves unpredictably in varying substrate conditions. To maintain the community stability under variable substrate conditions, GSMM should be use to design substrate-specific robust communities considering the detailed nutritional composition of substrates[126]. Additionally, multiple functionally redundant key strains can be included in the community to ensure community stability under varying substrate compositions. This will ensure that if one strain fails or becomes less active under certain substrates, others can compensate for the function and maintain the interspecies interactions necessary for community stability.

Besides these technical challenges, the application of multi-omics and GSMM approaches to design SynComs for WK faces additional challenges in the industrial implementation of these methods. In the modern fermented food industry, several advanced approaches are adopted to monitor production processes and improve the quality, safety, and efficiency of fermented foods (e.g., monitoring pH and sugar consumption, targeted metabolite analysis to quantify sugar, organic acids, and ethanol, automated control of defined fermentation conditions). However, integration of multi-omics and GSMM to iteratively optimise the SynComs require additional expertise, research facilities, and computational resources, increasing the associated cost and thereby, limiting their implementation at industrial scale. Therefore, these approaches are currently positioned as research-driven, long-term innovation tools rather than routine industrial implementation.

Regulatory and safety considerations for using genome-scale metabolic modelling-based synthetic communities in water kefir production

Local safety and regulatory compliance are critical in the design of GSMM-guided SynComs for the development of novel fermented foods. While GSMM approaches can automate SynCom design, they typically do not capture safety-relevant traits of individual strains. Therefore, additional safety assessments are essential to confirm that the selected strains are suitable for food use, as microbes in fermented foods can directly impact consumer health and product safety.

Foods must meet the local regulatory requirements, which vary between regions[177]. For example, in the European Union (EU), any food without a significant history of consumption in the EU before May 15, 1997, is considered a novel food and they are covered by regulation (EU) 2015/2283. Fermented food-associated microbial cultures are regarded as a type of food ingredient, and when new starter cultures are introduced, such as GSMM-guided SynComs, they will be covered by novel food regulations and require risk assessments. The European Food Safety Authority (EFSA) maintains a Qualified Presumption of Safety (QPS) list of microbes that have well-documented scientific evidence for safe use in food or feed. Therefore, the first step is to determine whether strains included in the SynComs have an established safety status in the QPS list for the intended use case[178]. For instance, some species on the QPS list are approved for the use during food fermentations on the condition that they are removed from the final food product ("production strains"), and others are approved for animal feedstuffs but not foods destined for human consumption. Even if a species is QPS-listed, it must generally be free of acquired antimicrobial resistance (AMR) to clinically relevant antimicrobials. Some species carry additional qualifications, for example, requirements for the absence of viable cells in the final product or demonstration of no toxigenic activity. Similarly, in the United States (U.S.), microbial strains used in foods are regulated by the Generally Recognized as Safe (GRAS) inventory of the Food and Drug Administration (FDA)[179].

Currently, more than 100 species hold EFSA's QPS status. WK-associated species on the list include LAB such as Lacticaseibacillus paracasei, Lentilactobacillus diolivorans, Lentilactobacillus hilgardii, Leuconostoc citreum, Leuconostoc mesenteroides, Leuconostoc pseudomesenteroides, and Oenococcus oeni; yeasts such as Hanseniaspora uvarum, Saccharomyces bayanus, Saccharomyces cerevisiae, and Schizosaccharomyces pombe; and, among AAB, only a few species, with Gluconobacter oxydans being the only QPS-listed representatives from WK.

If a strain is not included on the QPS list, a rigorous risk assessment must be performed to confirm its suitability for the use in food[180]. EFSA safety assessments for non-QPS species are triggered by application for market authorisation for products containing the strain(s). EFSA provides detailed guidance for the WGS analysis of microorganisms intentionally used in the food chain[181]. According to these guidelines, WGS data must be evaluated for the presence of genes encoding known virulence factors, toxins, invasion or adhesion factors, metabolic pathways relevant to toxigenicity, and antimicrobial resistance (AMR) to clinically relevant antimicrobials.

Microbes with acquired AMR genes could interact with the human microbiome and the wider environment[182]. Such strains may also disseminate AMR genes via horizontal gene transfer. Consequently, comprehensive AMR assessment using both phenotypic and genomics-based methods is essential. Phenotypic assays are useful for initial screening but determining whether AMR genes are intrinsic or acquired requires in-depth bioinformatics analysis. EFSA guidelines specify that intrinsic AMR genes (those naturally occurring and non-transferable) are generally acceptable for food use[183].

From both a consumer safety and regulatory perspective, fermentation substrates must be free from harmful contaminants, including residual pesticides, mycotoxins, and antibiotic contamination[184–187]. For plant-based substrates used in WK, pesticides and mycotoxins are a potential concern. GSMM guided SynCom design provides an excellent opportunity to incorporate strains with the capacity to degrade and detoxify such contaminants of substrates, and potentially enhancing the safety of WK.

Although SynComs for WK can be developed using naturally occurring microbes from WKGs, Genetically Modified Microorganisms (GMOs) can be useful in metabolic modelling, enabling controlled manipulation of a specific metabolic pathways and validation of model predictions. From an industrial perspective, the use of GMOs in SynComs may be useful in WK production by improving fermentation performance (e.g., enhancing substrate utilization), controlling metabolic outputs (e.g., knocking out pathways for undesirable byproducts), and enhancing sensory profiles (e.g., manipulating pathways for increased production of aroma compounds)[188]. However, the inclusion of GMOs is subject to regulatory constraints in many regions and likely to conflict with consumers' natural perceptions of fermented foods, reducing acceptance[189]. If GMOs are considered, their potential benefits must be weighed against consumer acceptance and compliance with any additional regional or national regulations that may apply[190]. Regardless of whether naturally occurring or genetically modified strains are used, all food production processes should adhere to established safety and quality frameworks, including *Good Manufacturing Practices*, *Hazard Analysis and Critical Control Point* systems, and *Sanitation Standard Operating Procedures*[191]. In the EU, EFSA requires risk assessment of genetically modified microbes under Regulation (EC) No 1829/2003 before they are authorised to be used in food and feed to ensure their safety for humans, animals, and the environment. Similarly, in the US, the FDA works with the Environmental Protection Agency and the U.S. Department of

Agriculture to ensure genetically modified microbes used in food are safe for people and the environment[192].

Another important regulatory consideration for most fermented beverages is ethanol content. Producers must comply with applicable national or regional legislation. For example, in the EU, alcohol content must be declared if it exceeds 1.2% alcohol by volume (Regulation (EU) No 1169/2011). A recent survey by the Food Safety Authority of Ireland on unpasteurised, plant-based fermented beverages (including kombucha, water kefir, and ginger soda) found that four out of 32 products contained undeclared ethanol levels above this threshold[193]. Similar findings also observed internationally. For example, a survey of 232 plant-based fermented beverages (including kombucha, water kefir, and ginger beer) in Australia revealed 52 of them had an alcohol content exceeding limit of 1.15% alcohol by volume[194]. While excessive ethanol levels present a challenge for the industry, they also offer an opportunity for SynCom-based optimisation. For instance, standard yeasts such as *S. cerevisiae*, commonly detected in these beverages, could be substituted with low-ethanol-producing yeast strains[195,196]. Moreover, microbial communities can be designed to actively reduce ethanol levels by including AAB capable of metabolising ethanol[197,198]. Conversely, based on the FDA regulations, if beverage exceeding 0.5% alcohol by volume, fall within the alcoholic beverage category under the Alcohol and Tobacco Tax and Trade Bureau[177].

Additionally, when selecting microbial strains for commercial starter cultures, producers must be aware of obligations under the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization[199], if their country is a signatory of the protocol. This international agreement, implemented in the EU via Regulation (EU) No 511/2014, may require permits or mutually agreed terms when accessing genetic resources from signatory countries. Such obligations can extend to both newly isolated strains and those obtained from third parties. Culture collections often have their own terms of use, which may restrict commercial applications or require benefit-sharing agreements. Additionally, strains sourced from starter culture providers may be subject to intellectual property rights or licensing conditions, which can limit their use in product development and commercial distribution. Careful verification of strain provenance and usage rights is therefore essential before incorporation into a commercial fermentation process.

Finally, in the context of fermented foods, the term "synthetic microbial community" may imply something unnatural. While the term synthetic microbial community is commonly used in synthetic ecology research, it may reduce consumer acceptance when it comes to food. As these communities are an assemblage of naturally occurring microbes derived from traditional fermented foods then perhaps to explain the process and improve the consumer acceptance a more appropriate term might be "defined community", similar to defined starter cultures often used in cheese production.

Outlook

Despite the advancement of community-level meta-omics approaches over the last few years, the application of comprehensive multi-omics on fermented foods remains limited. Recent large-scale study on WK by integrating multiple omics data demonstrated remarkable success in providing insight into the microbial diversity, interactions, and their metabolite profile[29]. Future research on WK and similar fermented foods using comprehensive multi-omics analysis would further elucidate the complex microbial interactions, fermentation dynamics, key metabolites, and their potential health impacts. Establishing and adopting a standardized and easy-to-follow workflow for fermented food multi-omics studies across research groups would ensure consistency, reproducibility, and comparability between the studies. It will significantly enhance the reliability and impact of findings. As with WK, many newly discovered fermented food-associated microbes remain uncultivated and uncharacterised[29,200]. Along with the multi-omics study of fermented food microbial communities, expanding cultivation and characterisation efforts of these newly discovered and uncharacterised species would provide insight into their functional roles.

Expanding the application of GSMM across a range of traditional fermented foods offers a promising approach for developing region-specific, novel, and healthy fermented foods using diverse sustainable substrates. This approach will contribute to ensuring global food and nutrition security and meeting the growing demand for sustainable functional foods. The systematic integration of multi-omics data with GSMM would unlock its full potential and help in the experimental validation of these novel foods.

Incorporating safety assessment features, such as screening of genes related to AMR, toxin production genes, virulence, and other potentially harmful traits, alongside current GSMM tools would improve the food-specific application. Ranking strains based on these safety features and whether the strains are listed on the QPS list would streamline the safe strain section for novel fermented food microbial communities. Additionally, globally harmonized regulations for the commercial production and labelling of these foods would improve safety and consumer acceptance. Finally, future research in fermented food could avoid using the term 'synthetic communities' and instead use 'defined communities' to overcome any potential consumer concerns.

Data availability

No datasets were generated or analysed during the current study.

Received: 26 November 2025; Accepted: 30 March 2026;
Published online: 15 April 2026

References

1. Aguiar-Pulido, V. et al. Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis: supplementary issue: bioinformatics methods and applications for big metagenomics data. *Evol. Bioinform.* **12s1**, EBO.S36436 (2016).
2. Arıkan, M. & Muth, T. Integrated multi-omics analyses of microbial communities: a review of the current state and future directions. *Mol. Omics* **19**, 607–623 (2023).
3. Tarzi, C., Zampieri, G., Sullivan, N. & Angione, C. Emerging methods for genome-scale metabolic modeling of microbial communities. *Trends Endocrinol. Metab.* **35**, 533–548 (2024).
4. Ankrah, N. Y. D. et al. Enhancing microbiome research through genome-scale metabolic modeling. *mSystems* **6**, e00599–21 (2021).
5. Fang, X., Lloyd, C. J. & Palsson, B. O. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* **18**, 731–743 (2020).
6. Kumar, M., Ji, B., Zengler, K. & Nielsen, J. Modelling approaches for studying the microbiome. *Nat. Microbiol.* **4**, 1253–1267 (2019).
7. Colarusso, A. V., Goodchild-Michelman, I., Rayle, M. & Zomorrodi, A. R. Computational modeling of metabolism in microbial communities on a genome-scale. *Curr. Opin. Syst. Biol.* **26**, 46–57 (2021).
8. Du, H., Li, M. & Liu, Y. Towards applications of genome-scale metabolic model-based approaches in designing synthetic microbial communities. *Quant. Biol.* **11**, 15–30 (2023).
9. Singh, D. & Lercher, M. J. Network reduction methods for genome-scale metabolic models. *Cell. Mol. Life Sci.* **77**, 481–488 (2020).
10. Edwards, J. S. & Palsson, B. O. Systems properties of the *Haemophilus influenzaeRd* metabolic genotype. *J. Biol. Chem.* **274**, 17410–17416 (1999).
11. Cautha, S. C. et al. Model-driven design of a *Saccharomyces cerevisiae* platform strain with improved tyrosine production capabilities. *IFAC Proc.* **46**, 221–226 (2013).
12. Guan, N. et al. Comparative genomics and transcriptomics analysis-guided metabolic engineering of *Propionibacterium acidipropionici* for improved propionic acid production. *Biotechnol. Bioeng.* **115**, 483–494 (2018).
13. McAnulty, M. J., Yen, J. Y., Freedman, B. G. & Senger, R. S. Genome-scale modeling using flux ratio constraints to enable

metabolic engineering of clostridial metabolism in silico. *BMC Syst. Biol.* **6**, 42 (2012).
14. Folger, O. et al. Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7**, 501 (2011).
15. Mardinoglu, A. et al. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083 (2014).
16. Giordano, N. et al. Genome-scale community modelling reveals conserved metabolic cross-feedings in epipelagic bacterioplankton communities. *Nat. Commun.* **15**, 2721 (2024).
17. Levy, R. & Borenstein, E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci.* **110**, 12804–12809 (2013).
18. Mataigne, V., Vannier, N., Vandenkoornhuyse, P. & Hacquard, S. Multi-genome metabolic modeling predicts functional inter-dependencies in the *Arabidopsis* root microbiome. *Microbiome* **10**, 217 (2022).
19. Zomorrodi, A. R. & Segrè, D. Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nat. Commun.* **8**, 1563 (2017).
20. Gonçalves, O. S., Creevey, C. J. & Santana, M. F. Designing a synthetic microbial community through genome metabolic modeling to enhance plant–microbe interaction. *Environ. Microbiome* **18**, 81 (2023).
21. Sabach, O. et al. The rise of the sourdough: genome-scale metabolic modeling-based approach to design sourdough starter communities with tailored-made properties. *Int. J. Food Microbiol.* **407**, 110402 (2023).
22. Großkopf, T. & Soyer, O. S. Synthetic microbial communities. *Curr. Opin. Microbiol.* **18**, 72–77 (2014).
23. Shi, H. et al. Advances in fermented foods revealed by multi-omics: a new direction toward precisely clarifying the roles of microorganisms. *Front. Microbiol.* **13**, 1044820 (2022).
24. Bozkir, E., Yilmaz, B., Sharma, H., Esatbeyoglu, T. & Ozogul, F. Challenges in water kefir production and limitations in human consumption: a comprehensive review of current knowledge. *Heliyon* **10**, e33501 (2024).
25. Vinderola, G. et al. Fermented foods: a perspective on their role in delivering biotics. *Front. Microbiol.* **14**, 1196239 (2023).
26. Bechtner, J., Wefers, D., Schmid, J., Vogel, R. F. & Jakob, F. Identification and comparison of two closely related dextransucrases released by water kefir borne *Lactobacillus hordei* TMW 1.1822 and *Lactobacillus nagelii* TMW 1.1827. *Microbiology* **165**, 956–966 (2019).
27. Waldherr, F. W., Doll, V. M., Meißner, D. & Vogel, R. F. Identification and characterization of a glucan-producing enzyme from Lactobacillus hilgardii TMW 1.828 involved in granule formation of water kefir. *Food Microbiol.* **27**, 672–678 (2010).
28. Lynch, K. M., Wilkinson, S., Daenen, L. & Arendt, E. K. An update on water kefir: microbiology, composition and production. *Int. J. Food Microbiol.* **345**, 109128 (2021).
29. Breselge, S. et al. The core microbiomes and associated metabolic potential of water kefir as revealed by pan multi-omics. *Commun. Biol.* **8**, 415 (2025).
30. Gulitz, A., Stadie, J., Wenning, M., Ehrmann, M. A. & Vogel, R. F. The microbial diversity of water kefir. *Int. J. Food Microbiol.* **151**, 284–288 (2011).
31. Laureys, D., Aerts, M., Vandamme, P. & De Vuyst, L. The buffer capacity and calcium concentration of water influence the microbial species diversity, grain growth, and metabolite production during water kefir fermentation. *Front. Microbiol.* **10**, 2876 (2019).
32. Laureys, D., Aerts, M., Vandamme, P. & De Vuyst, L. Oxygen and diverse nutrients influence the water kefir fermentation process. *Food Microbiol* **73**, 351–361 (2018).
33. Laureys, D. & De Vuyst, L. The water kefir grain inoculum determines the characteristics of the resulting water kefir fermentation process. *J. Appl. Microbiol.* **122**, 719–732 (2017).
34. Laureys, D. & De Vuyst, L. Microbial species diversity, community dynamics, and metabolite kinetics of water kefir fermentation. *Appl. Environ. Microbiol.* **80**, 2564–2572 (2014).
35. Marsh, A. J., O'Sullivan, O., Hill, C., Ross, R. P. & Cotter, P. D. Sequence-based analysis of the microbial composition of water kefir from multiple sources. *FEMS Microbiol. Lett.* **348**, 79–85 (2013).
36. Stadie, J., Gulitz, A., Ehrmann, M. A. & Vogel, R. F. Metabolic activity and symbiotic interactions of lactic acid bacteria and yeasts isolated from water kefir. *Food Microbiol.* **35**, 92–98 (2013).
37. Verce, M., De Vuyst, L. & Weckx, S. Shotgun metagenomics of a water kefir fermentation ecosystem reveals a novel *Oenococcus* species. *Front. Microbiol.* **10**, 479 (2019).
38. Breselge, S. et al. Bifidobacterium fermentum sp. nov. and *Bifidobacterium aquikefiricola* sp. nov., isolated from water kefir. *Int. J. Syst. Evol. Microbiol*. **74**, 006549 (2024).
39. Leech, J. et al. Fermented-food metagenomics reveals substrate-associated differences in taxonomy and health-associated and antibiotic resistance determinants. *mSystems* **5**, e00522–20 (2020).
40. Patel, S. H. et al. A temporal view of the water kefir microbiota and flavour attributes. *Innov. Food Sci. Emerg. Technol.* **80**, 103084 (2022).
41. Arrieta-Echeverri, M. C. et al. Multi-omics characterization of the microbial populations and chemical space composition of a water kefir fermentation. *Front. Mol. Biosci.* **10**, 1223863 (2023).
42. Azi, F. et al. Metabolite dynamics and phytochemistry of a soy whey-based beverage bio-transformed by water kefir consortium. *Food Chem.* **342**, 128225 (2021).
43. Gökırmaklı, Ç, Yüceer, Y. K. & Guzel-Seydim, Z. B. Chemical, microbial, and volatile changes of water kefir during fermentation with economic substrates. *Eur. Food Res. Technol.* **249**, 1717–1728 (2023).
44. Wang, Y. et al. Volatile flavor characteristics and microbial fermentation differences of water kefir grains from different origins. *Eur. Food Res. Technol.* https://doi.org/10.1007/s00217-025-04823-w (2025).
45. Rios, D. L. et al. Comparative metatranscriptome analysis of Brazilian milk and water kefir beverages. *Int. Microbiol.* **27**, 807–818 (2023).
46. Gulitz, A., Stadie, J., Ehrmann, M. A., Ludwig, W. & Vogel, R. F. Comparative phylobiomic analysis of the bacterial community of water kefir by 16S rRNA gene amplicon sequencing and ARDRA analysis. *J. Appl. Microbiol.* **114**, 1082–1091 (2013).
47. Köhler, S. et al. Tradition as a stepping stone for a microbial defined water kefir fermentation process: insights in cell growth, bioflavoring, and sensory perception. *Front. Microbiol.* **12**, 732019 (2021).
48. Duboux, S., Zhang, J., Börner, R. A., Priour, S. & Tan, J. P. Symbiotic culture of bacteria and yeast for production of water kefir. WO2023072715A1, World Intellectual Property Organization (2023).
49. Xu, D. et al. Label-free quantitative proteomic analysis reveals the lifestyle of Lactobacillus hordei in the presence of *Sacchromyces cerevisiae*. *Int. J. Food Microbiol.* **294**, 18–26 (2019).
50. Yap, M. et al. Next-generation food research: use of meta-omic approaches for characterizing microbial communities along the food chain. *Annu. Rev. Food Sci. Technol.* **13**, 361–384 (2022).
51. Rodríguez, M. A. et al. Bacterial diversity using metagenomics of 16s rDNA in water kefir, an innovative source of probiotics for bee nutrition. *Rev. Argent. Microbiol.* **56**, 191–197 (2024).
52. Yerlikaya, O., Akan, E. & Kinik, Ö The metagenomic composition of water kefir microbiota. *Int. J. Gastron. Food Sci.* **30**, 100621 (2022).
53. Chen, W. Demystification of fermented foods by omics technologies. *Curr. Opin. Food Sci.* **46**, 100845 (2022).

54. Zhang, Y., Thompson, K. N., Huttenhower, C. & Franzosa, E. A. Statistical approaches for differential expression analysis in metatranscriptomics. *Bioinformatics* **37**, i34–i41 (2021).
55. Yang, L., Fan, W. & Xu, Y. Metaproteomics insights into traditional fermented foods and beverages. *Compr. Rev. Food Sci. Food Saf.* **19**, 2506–2529 (2020).
56. Ma, D., Wang, B., Xiao, S. & Wang, J. Non-targeted metabolomics unveils metabolic spectrum characteristics and core metabolite interaction networks in water kefir fermentation. *LWT* **195**, 115840 (2024).
57. Wen, L. et al. Applications of multi-omics techniques to unravel the fermentation process and the flavor formation mechanism in fermented foods. *Crit. Rev. Food Sci. Nutr.* **64**, 8367–8383 (2024).
58. Tarazona, S. et al. Harmonization of quality metrics and power calculation in multi-omic studies. *Nat. Commun.* **11**, 3092 (2020).
59. Woodland, B. et al. Sample preparation for multi-omics analysis: considerations and guidance for identifying the ideal workflow. *Proteomics* e13983, https://doi.org/10.1002/pmic.13983 (2025).
60. Brockbals, L., Ueland, M., Fu, S. & Padula, M. P. Development and thorough evaluation of a multi-omics sample preparation workflow for comprehensive LC-MS/MS-based metabolomics, lipidomics and proteomics datasets. *Talanta* **286**, 127442 (2025).
61. Kang, J., David, L., Li, Y., Cang, J. & Chen, S. Three-in-one simultaneous extraction of proteins, metabolites and lipids for multi-omics. *Front. Genet.* **12**, 635971 (2021).
62. Zakharkin, S. O., Kim, K., Bartolucci, A. A., Page, G. P. & Allison, D. B. Optimal allocation of replicates for measurement evaluation studies. *Genom. Proteom. Bioinform.* **4**, 196–202 (2006).
63. Zhou, R., Ng, S. K., Sung, J. J. Y., Goh, W. W. B. & Wong, S. H. Data pre-processing for analyzing microbiome data – a mini review. *Comput. Struct. Biotechnol. J.* **21**, 4804–4815 (2023).
64. Ferrocino, I. et al. The need for an integrated multi-OMICs approach in microbiome science in the food system. *Compr. Rev. Food Sci. Food Saf.* **22**, 1082–1103 (2023).
65. Heintz-Buschart, A. & Westerhuis, J. A. A beginner's guide to integrating multi-omics data from microbial communities. *Biochemist* **44**, 23–29 (2022).
66. Lee, B., Zhang, S., Poleksic, A. & Xie, L. Heterogeneous multi-layered network model for omics data integration and analysis. *Front. Genet.* **10**, 1381 (2020).
67. Argelaguet, R. et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, e8124 (2018).
68. Zaramela, L. S., Tjuanta, M., Moyne, O., Neal, M. & Zengler, K. synDNA—a synthetic DNA spike-in method for absolute quantification of shotgun metagenomic sequencing. *mSystems* **7**, e00447–22 (2022).
69. Bong, D., Sohn, J. & Lee, S.-J. V. Brief guide to RT-qPCR. *Mol. Cells* **47**, 100141 (2024).
70. Green, M. R. & Sambrook, J. Quantification of RNA by real-time reverse transcription-polymerase chain reaction (RT-PCR). *Cold Spring Harb. Protoc.* **2018**, pdb.prot095042 (2018).
71. Owens, N. D. L. et al. Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell Rep.* **14**, 632–647 (2016).
72. Kettenbach, A. N., Rush, J. & Gerber, S. A. Absolute quantification of protein and post-translational modification abundance with stable isotope–labeled synthetic peptides. *Nat. Protoc.* **6**, 175–186 (2011).
73. Calle, M. L., Pujolassos, M. & Susin, A. coda4microbiome: compositional data analysis for microbiome cross-sectional and longitudinal studies. *BMC Bioinform.* **24**, 82 (2023).
74. Morabito, A., De Simone, G., Pastorelli, R., Brunelli, L. & Ferrario, M. Algorithms and tools for data-driven omics integration to achieve multilayer biological insights: a narrative review. *J. Transl. Med.* **23**, 425 (2025).
75. Song, M. et al. A review of integrative imputation for multi-omics datasets. *Front. Genet.* **11**, 570255 (2020).
76. Luo, Y., Zhao, C. & Chen, F. Multiomics research: principles and challenges in integrated analysis. *BioDesign Res.* **6**, 0059 (2024).
77. Fondi, M. & Liò, P. Multi -omics and metabolic modelling pipelines: challenges and tools for systems microbiology. *Microbiol. Res.* **171**, 52–64 (2015).
78. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
79. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
80. Wu, S. et al. Deciphering and designing microbial communities by genome-scale metabolic modelling. *Comput. Struct. Biotechnol. J.* **23**, 1990–2000 (2024).
81. Gopaulchan, D. et al. A defined microbial community reproduces attributes of fine flavour chocolate fermentation. *Nat. Microbiol.* **10**, 2130–2152 (2025).
82. Rappaport, H. B., Senewiratne, N. P. J., Lucas, S. K., Wolfe, B. E. & Oliverio, A. M. Genomics and synthetic community experiments uncover the key metabolic roles of acetic acid bacteria in sourdough starter microbiomes. *mSystems* **9**, e00537–24 (2024).
83. Guzel-Seydim, Z. B., Gökırmaklı, Ç. & Greene, A. K. A comparison of milk kefir and water kefir: physical, chemical, microbiological and functional properties. *Trends Food Sci. Technol.* **113**, 42–53 (2021).
84. Ustaoğlu-Gençgönül, M., Gökırmaklı, Ç., Üçgül, B., Karagül-Yüceer, Y. & Guzel-Seydim, Z. B. Chemical, microbial, and volatile compounds of water kefir beverages made from chickpea, almond, and rice extracts. *Eur. Food Res. Technol.* **250**, 2233–2244 (2024).
85. Alvarez, M. A. & Moreno-Arribas, M. V. The problem of biogenic amines in fermented foods and the use of potential biogenic amine-degrading microorganisms as a solution. *Trends Food Sci. Technol.* **39**, 146–155 (2014).
86. Laureys, D. et al. The type and concentration of inoculum and substrate as well as the presence of oxygen impact the water kefir fermentation process. *Front. Microbiol.* **12**, 628599 (2021).
87. Wolfe, B. E. & Dutton, R. J. Fermented foods as experimentally tractable microbial ecosystems. *Cell* **161**, 49–55 (2015).
88. Braissant, O., Bachmann, A. & Bonkat, G. Microcalorimetric assays for measuring cell growth and metabolic activity: methodology and applications. *Methods* **76**, 27–34 (2015).
89. Mackie, A. M., Hassan, K. A., Paulsen, I. T. & Tetu, S. G. Biolog phenotype microarrays for phenotypic characterization of microbial cells. in *Environmental Microbiology* (eds. Paulsen, I. T. & Holmes, A. J.) Vol. 1096 123–130 (Humana Press, 2014).
90. Hockett, K. L. & Baltrus, D. A. Use of the soft-agar overlay technique to screen for bacterially produced inhibitory compounds. *J. Vis. Exp.* 55064, https://doi.org/10.3791/55064 (2017).
91. Temkin, M. I., Carlson, C. M., Stubbendieck, A. L., Currie, C. R. & Stubbendieck, R. M. High throughput co-culture assays for the investigation of microbial interactions. *J. Vis. Exp.* 60275, https://doi.org/10.3791/60275 (2019).
92. Wick, R. R. et al. Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol.* **22**, 266 (2021).
93. Zhang, X. et al. Benchmarking of long-read sequencing, assemblers and polishers for yeast genome. *Brief. Bioinform.* **23**, bbac146 (2022).
94. Cosma, B.-M. et al. Evaluating long-read de novo assembly tools for eukaryotic genomes: insights and considerations. *GigaScience* **12**, giad100 (2022).
95. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
96. Tatusova, T. et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).

97. Ruiz-Perez, C. A., Conrad, R. E. & Konstantinidis, K. T. MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinform.* **22**, 11 (2021).
98. Schwengers, O. et al. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification: Find out more about Bakta, the motivation, challenges and applications, here. *Microb. Genomics* **7**, (2021).
99. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**, 757–768 (2024).
100. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).
101. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**, 491 (2011).
102. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).
103. Palmer, J. M. & Stajich, J. Funannotate v1.8.1: Eukaryotic genome annotation. Zenodo. https://doi.org/10.5281/ZENODO.4054262 (2020).
104. Proux-Wéra, E., Armisén, D., Byrne, K. P. & Wolfe, K. H. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinform.* **13**, 237 (2012).
105. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
106. Faria, J. P., Rocha, M., Rocha, I. & Henry, C. S. Methods for automated genome-scale metabolic model reconstruction. *Biochem. Soc. Trans.* **46**, 931–936 (2018).
107. Zimmermann, J., Kaleta, C. & Waschina, S. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol.* **22**, 81 (2021).
108. Prigent, S. et al. Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLoS Comput. Biol.* **13**, e1005276 (2017).
109. Thiele, I., Vlassis, N. & Fleming, R. M. T. fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* **30**, 2529–2531 (2014).
110. Boer, M. D. et al. Improving genome-scale metabolic models of incomplete genomes with deep learning. *iScience* **27**, 111349 (2024).
111. Mendoza, S. N., Olivier, B. G., Molenaar, D. & Teusink, B. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.* **20**, 158 (2019).
112. Matveishina, E. K., Bartmanski, B. J., Benito-Vaquerizo, S. & Zimmermann-Kogadeeva, M. GEMsembler: consensus model assembly and structural comparison of genome-scale metabolic models across tools improve functional performance. *mSystems* **10**, e00574–25 (2025).
113. Sauls, J. T. & Buescher, J. M. Assimilating genome-scale metabolic reconstructions with modelBorgifier. *Bioinformatics* **30**, 1036–1038 (2014).
114. Hari, A., Zarrabi, A. & Lobo, D. *mergem*: merging, comparing, and translating genome-scale metabolic models using universal identifiers. *NAR Genom. Bioinform.* **6**, lqae010 (2024).
115. Moyer, D. C., Reimertz, J., Segrè, D. & Fuxman Bass, J. I. MACAW: a method for semi-automatic detection of errors in genome-scale metabolic models. *Genome Biol.* **26**, 79 (2025).
116. Zorrilla, F., Buric, F., Patil, K. R. & Zelezniak, A. metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Res.* **49**, e126–e126 (2021).
117. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
118. Domenzain, I. et al. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nat. Commun.* **13**, 3766 (2022).
119. Kleiner, M. Metaproteomics: much more than measuring gene expression in microbial communities. *mSystems* **4**, e00115–19 (2019).
120. Zhou, J., Zhuang, Y. & Xia, J. Integration of enzyme constraints in a genome-scale metabolic model of *Aspergillus niger* improves phenotype predictions. *Microb. Cell Factories* **20**, 125 (2021).
121. Zampieri, G., Campanaro, S., Angione, C. & Treu, L. Metatranscriptomics-guided genome-scale metabolic modeling of microbial communities. *Cell Rep. Methods* **3**, 100383 (2023).
122. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
123. Mahadevan, R., Edwards, J. S. & Doyle, F. J. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* **83**, 1331–1340 (2002).
124. Ebenhöh, O., Handorf, T. & Heinrich, R. Structural analysis of expanding metabolic networks. *Genome Inform. Int. Conf. Genome Inform.* **15**, 35–45 (2004).
125. Handorf, T., Ebenhöh, O. & Heinrich, R. Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J. Mol. Evol.* **61**, 498–512 (2005).
126. Belcour, A. et al. Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. *eLife* **9**, e61968 (2020).
127. Aittokallio, T. Graph-based methods for analysing networks in cell biology. *Brief. Bioinform.* **7**, 243–255 (2006).
128. Lee, Y. Q. et al. Genome-scale metabolic model-guided systematic framework for designing customized live biotherapeutic products. *npj Syst. Biol. Appl.* **11**, 73 (2025).
129. Jin, R. et al. Synthetic microbial communities: novel strategies to enhance the quality of traditional fermented foods. *Compr. Rev. Food Sci. Food Saf.* **23**, e13388 (2024).
130. Eng, A. & Borenstein, E. An algorithm for designing minimal microbial communities with desired metabolic capacities. *Bioinformatics* **32**, 2008–2016 (2016).
131. Frioux, C., Fremy, E., Trottier, C. & Siegel, A. Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics* **34**, i934–i943 (2018).
132. Liu, R., Bassalo, M. C., Zeitoun, R. I. & Gill, R. T. Genome scale engineering techniques for metabolic engineering. *Metab. Eng.* **32**, 143–154 (2015).
133. Michielsen, S., Vercelli, G. T., Cordero, O. X. & Bachmann, H. Spatially structured microbial consortia and their role in food fermentations. *Curr. Opin. Biotechnol.* **87**, 103102 (2024).
134. Srinivas, M., O'Sullivan, O., Cotter, P. D., Sinderen, D. V. & Kenny, J. G. The application of metagenomics to study microbial communities and develop desirable traits in fermented foods. *Foods* **11**, 3297 (2022).
135. Breselge, S. et al. Water kefir multi-omics reveals functional redundancies despite taxonomic differences and the underappreciated contribution of yeast. *npj Sci. Food* **9**, 265 (2025).
136. Cantuti Gendre, J. et al. Exploring viral diversity in fermented vegetables through viral metagenomics. *Food Microbiol.* **128**, 104733 (2025).
137. Zhang, H., Zhang, H., Du, H., Yu, X. & Xu, Y. The insights into the phage communities of fermented foods in the age of viral metagenomics. *Crit. Rev. Food Sci. Nutr.* **65**, 1656–1668 (2025).

138. Li, C. et al. A systematic discussion and comparison of the construction methods of synthetic microbial community. *Synth. Syst. Biotechnol.* **9**, 775–783 (2024).
139. Peng, M. et al. Effects of a mixed koji culture of *Aspergillus oryzae* HG-26 and *Aspergillus niger* HG-35 on the levels of enzymes, antioxidants and phenolic compounds in soy sauce during the fermentation process. *Int. J. Food Sci. Technol.* **52**, 1585–1593 (2017).
140. Shi, S. et al. Microbial metabolic transformation and antioxidant activity evaluation of polyphenols in kombucha. *Food Biosci.* **51**, 102287 (2023).
141. Wang, H. et al. Core microbes identification and synthetic microbiota construction for the production of Xiaoqu light-aroma Baijiu. *Food Res. Int.* **183**, 114196 (2024).
142. Arrigan, D., Kothe, C. I., Oliverio, A., Evans, J. D. & Wolfe, B. E. Novel fermentations integrate traditional practice and rational design of fermented-food microbiomes. *Curr. Biol.* **34**, R1094–R1108 (2024).
143. Alves, V. et al. Development of fermented beverage with water kefir in water-soluble coconut extract (*Cocos nucifera L.*) with inulin addition. *LWT* **145**, 111364 (2021).
144. Da Costa, M. R., De Alencar, E. R., Dos Santos Leandro, E., Mendonça, M. A. & De Souza Ferreira, W. F. Characterization of the kefir beverage produced from yam (*Colocasia esculenta L.*), sesame seed (*Sesamum indicum L.*) and bean (*Phaseolus vulgaris L.*) extracts. *J. Food Sci. Technol.* **55**, 4851–4858 (2018).
145. Fiorda, F. A. et al. Development of kefir-based probiotic beverages with DNA protection and antioxidant activities using soybean hydrolyzed extract, colostrum and honey. *LWT Food Sci. Technol.* **68**, 690–697 (2016).
146. Güzel-Seydim, Z. B., Şatır, G. & Gökırmaklı, Ç Use of mandarin and persimmon fruits in water kefir fermentation. *Food Sci. Nutr.* **11**, 5890–5897 (2023).
147. Limbad, M. et al. Microbial and chemical changes during fermentation of coconut water kefir beverage. *Appl. Sci.* **13**, 7257 (2023).
148. Ozcelik, F., Akan, E. & Kinik, O. Use of Cornelian cherry, hawthorn, red plum, roseship and pomegranate juices in the production of water kefir beverages. *Food Biosci.* **42**, 101219 (2021).
149. Sanches, F. L. et al. Study and characterization of a product based on a vegetable extract of quinoa fermented with water kefir grains. *World J. Microbiol. Biotechnol.* **40**, 118 (2024).
150. Tavares, P. P. L. G. et al. Non-conventional sucrose-based substrates: development of non-dairy kefir beverages with probiotic potential. *Fermentation* **9**, 384 (2023).
151. Wang, X. & Wang, P. Red beetroot juice fermented by water kefir grains: physicochemical, antioxidant profile and anticancer activity. *Eur. Food Res. Technol.* **249**, 939–950 (2023).
152. Zanirati, D. F. et al. Selection of lactic acid bacteria from Brazilian kefir grains for potential use as starter or probiotic cultures. *Anaerobe* **32**, 70–76 (2015).
153. De Roos, J. & De Vuyst, L. Acetic acid bacteria in fermented foods and beverages. *Curr. Opin. Biotechnol.* **49**, 115–119 (2018).
154. Liu, S. et al. Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. *Microbiome* **10**, 76 (2022).
155. Chen, C. et al. Function-driven high-throughput screening and isolation of ester-producing strains for glutinous rice wine fermentation. *Food Res. Int.* **199**, 115393 (2025).
156. Huang, Y. et al. High-throughput microbial culturomics using automation and machine learning. *Nat. Biotechnol.* **41**, 1424–1433 (2023).
157. Xu, J. et al. Culturing bacteria from fermentation pit muds of Baijiu with culturomics and amplicon-based metagenomic approaches. *Front. Microbiol.* **11**, 1223 (2020).
158. Kyoui, D., Mikami, N., Yamamoto, H., Kawarai, T. & Ogihara, H. Complete genome sequence of *Lactobacillus curvatus* nfh-km12, isolated from the Japanese traditional fish fermented food Kabura-zushi. *Microbiol. Resour. Announc.* **7**, e00823–18 (2018).
159. Wick, R. R., Judd, L. M. & Holt, K. E. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLOS Comput. Biol.* **19**, e1010905 (2023).
160. Wick, R. R., Howden, B. P. & Stinear, T. P. Autocycler: long-read consensus assembly for bacterial genomes. *Bioinformatics* **41**, btaf474 (2025).
161. Schiffer, A. M., Rahman, A., Sutton, W., Putnam, M. L. & Weisberg, A. J. A comparison of short- and long-read whole-genome sequencing for microbial pathogen epidemiology. *mSystems* e01426-25, https://doi.org/10.1128/msystems.01426-25 (2025).
162. Sereika, M. et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat. Methods* **19**, 823–826 (2022).
163. The UniProt Consortium et al. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617 (2025).
164. Griesemer, M., Kimbrel, J. A., Zhou, C. E., Navid, A. & D'haeseleer, P. Combining multiple functional annotation tools increases coverage of metabolic annotation. *BMC Genom.* **19**, 948 (2018).
165. Han, Y. et al. Unlocking the potential of metagenomics with the PacBio high-fidelity sequencing technology. *Microorganisms* **12**, 2482 (2024).
166. Lai, Q. et al. Deciphering the biosynthetic potential of microbial genomes using a BGC language processing neural network model. *Nucleic Acids Res.* **53**, gkaf305 (2025).
167. Blin, K. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
168. Hannigan, G. D. et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* **47**, e110–e110 (2019).
169. De Oliveira, A. C. F. M. et al. A metabologenomics approach reveals the unexplored biosynthetic potential of bacteria isolated from an Amazon Conservation Unit. *Microbiol. Spectr.* **13**, e00996–24 (2025).
170. De Almeida, E. L. M., Kerkhoven, E. J. & Da Silveira, W. B. Reconstruction of genome-scale metabolic models of non-conventional yeasts: current state, challenges, and perspectives. *Biotechnol. Bioprocess Eng.* **29**, 35–67 (2024).
171. Renzi, S., Nenciarini, S., Bacci, G. & Cavalieri, D. Yeast metagenomics: analytical challenges in the analysis of the eukaryotic microbiome. *Microbiome Res. Rep*. **3**, 2 (2023).
172. Balarezo-Cisneros, L. N. et al. High quality de novo genome assembly of the non-conventional yeast Kazachstania bulderi describes a potential low pH production host for biorefineries. *Commun. Biol.* **6**, 918 (2023).
173. Carter, E. L., Constantinidou, C. & Alam, M. T. Applications of genome-scale metabolic models to investigate microbial metabolic adaptations in response to genetic or environmental perturbations. *Brief. Bioinform.* **25**, bbad439 (2023).
174. Ventorim, R. Z., Ferreira, M. A. D. M., De Almeida, E. L. M., Kerkhoven, E. J. & Da Silveira, W. B. Genome-scale metabolic model of oleaginous yeast *Papiliotrema laurentii*. *Biochem. Eng. J.* **180**, 108353 (2022).
175. Nikoloudaki, O., Aheto, F., Di Cagno, R. & Gobbetti, M. Synthetic microbial communities: A gateway to understanding resistance, resilience, and functionality in spontaneously fermented food microbiomes. *Food Res. Int.* **192**, 114780 (2024).
176. Zannini, E. et al. Influence of substrate on the fermentation characteristics and culture-dependent microbial composition of water kefir. *Fermentation* **9**, 28 (2022).
177. Mukherjee, A., Gómez-Sala, B., O'Connor, E. M., Kenny, J. G. & Cotter, P. D. Global regulatory frameworks for fermented foods: a review. *Front. Nutr.* **9**, 902642 (2022).

178. EFSA Panel on Biological Hazards (BIOHAZ) et al. Update of the list of qualified presumption of safety (QPS) recommended microbiological agents intentionally added to food or feed as notified to EFSA 22: suitability of taxonomic units notified to EFSA until March 2025. *EFSA J*. **23**, e9510 (2025).
179. U.S. Food and Drug Administration. Generally Recognized as Safe (GRAS). *U.S. Food & Drug Administration* https://www.fda.gov/food/food-ingredients-packaging/generally-recognized-safe-gras (2023).
180. EFSA Panel on Nutrition, Novel Foods and Food Allergens (NDA) et al. Guidance on the scientific requirements for an application for authorisation of a novel food in the context of regulation (EU) 2015/2283. *EFSA J*. **22**, e8961 (2024).
181. European Food Safety Authority (EFSA). EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. *EFSA J*. **22**, e8912 (2024).
182. Wolfe, B. E. Are fermented foods an overlooked reservoir of antimicrobial resistance? *Curr. Opin. Food Sci.* **51**, 101018 (2023).
183. EFSA Panel on Biological Hazards (BIOHAZ) et al. Statement on how to interpret the QPS qualification on 'acquired antimicrobial resistance genes'. *EFSA J*. **21**, e05134 (2023).
184. Battilani, P. et al. Mycotoxin mixtures in food and feed: holistic, innovative, flexible risk assessment modelling approach: *EFSA Support*. *Publ*. **17**, 1757E (2020).
185. European Food Safety Authority (EFSA) et al. The 2023 European Union report on pesticide residues in food. *EFSA J*. **23**, 4038 (2025).
186. U.S. Food & Drug Administration. Chemical contaminants & pesticides. *U.S. Food & Drug Administration.* https://www.fda.gov/food/chemical-contaminants-pesticides (2025).
187. U.S. Food & Drug Administration. Mycotoxins. *U.S. Food & Drug Administration.* https://www.fda.gov/food/natural-toxins-food/mycotoxins (2024).
188. Mirsalami, S. M. & Mirsalami, M. Advances in genetically engineered microorganisms: transforming food production through precision fermentation and synthetic biology. *Future Foods* **11**, 100601 (2025).
189. Ramírez Rojas, A. A., Swidah, R. & Schindler, D. Microbes of traditional fermentation processes as synthetic biology chassis to tackle future food challenges. *Front. Bioeng. Biotechnol.* **10**, 982975 (2022).
190. Hanlon, P. & Sewalt, V. GEMs: genetically engineered microorganisms and the regulatory oversight of their uses in modern food production. *Crit. Rev. Food Sci. Nutr.* **61**, 959–970 (2021).
191. De Oliveira, C. A. F., Da Cruz, A. G., Tavolaro, P. & Corassin, C. H. Food safety: good manufacturing practices, sanitation standard operating procedures, hazard analysis, and critical control point. in 163–173. https://doi.org/10.1016/B978-0-323-90747-7.00011-9 (Elsevier, 2025).
192. U.S. Food & Drug Administration. How GMOs are regulated in the United States. *U.S. Food & Drug Administration.* https://www.fda.gov/food/agricultural-biotechnology/how-gmos-are-regulated-united-states (2024).
193. Food Safety Authority of Ireland. *Survey of Plant-Based Fermented Beverages Available on the Irish Market* (Food Safety Authority of Ireland, 2021).
194. Food Standards Australia New Zealand. *Survey of Alcohol Content and Labelling of Fermented Soft Drinks*. https://www.foodstandards.gov.au/science-data/surveillance/survey-of-alcohol-content-and-labelling-of-fermented-soft-drinks?utm_source=chatgpt.com (2019).
195. Bellut, K. & Arendt, E. K. Chance and challenge: non-*Saccharomyces* yeasts in nonalcoholic and low alcohol beer brewing – a review. *J. Am. Soc. Brew. Chem.* **77**, 77–91 (2019).
196. Contreras, A. et al. Evaluation of non-*Saccharomyces* yeasts for the reduction of alcohol content in wine. *Appl. Environ. Microbiol.* **80**, 1670–1678 (2014).
197. He, Y. et al. Oxidative fermentation of acetic acid bacteria and its products. *Front. Microbiol.* **13**, 879246 (2022).
198. Raspor, P. & Goranovič, D. Biotechnological applications of acetic acid bacteria. *Crit. Rev. Biotechnol.* **28**, 101–124 (2008).
199. Smith, D., Da Silva, M., Jackson, J. & Lyal, C. Explanation of the Nagoya Protocol on access and benefit sharing and its implication for microbiology. *Microbiology* **163**, 289–296 (2017).
200. Carlino, N. et al. Unexplored microbial diversity from 2500 food metagenomes and links with the human microbiome. *Cell* **187**, 5775–5795.e15 (2024).
201. Narayanasamy, S. et al. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol.* **17**, 260 (2016).
202. Muñoz-Benavent, M. et al. gNOMO: a multi-omics pipeline for integrated host and microbiome analysis of non-model organisms. *NAR Genom. Bioinform.* **2**, lqaa058 (2020).
203. Arikan, M. & Muth, T. gNOMO2: a comprehensive and modular pipeline for integrated multi-omics analyses of microbiomes. *GigaScience* **13**, giae038 (2024).
204. Morton, J. T. et al. Learning representations of microbe–metabolite interactions. *Nat. Methods* **16**, 1306–1314 (2019).
205. Ma, Y. & Liu, L. NMFGOT: a multi-view learning framework for the microbiome and metabolome integrative analysis with optimal transport plan. *npj Biofilms Microbiomes* **10**, 135 (2024).
206. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
207. Meng, C., Kuster, B., Culhane, A. C. & Gholami, A. M. A multivariate approach to the integration of multi-omics datasets. *BMC Bioinform.* **15**, 162 (2014).
208. Zoppi, J., Guillaume, J.-F., Neunlist, M. & Chaffron, S. MiBiOmics: an interactive web application for multi-omics data exploration and integration. *BMC Bioinform.* **22**, 6 (2021).
209. Ruiz-Perez, D. et al. Dynamic bayesian networks for integrating multi-omics time series microbiome data. *mSystems* **6**, https://doi.org/10.1128/msystems.01105-20 (2021).
210. Zhou, G., Ewald, J. & Xia, J. OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res.* **49**, W476–W482 (2021).
211. Schiml, V. C. et al. Integrative meta-omics in Galaxy and beyond. *Environ. Microbiome* **18**, 56 (2023).
212. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.* **48**, D445–D453 (2020).
213. Karp, P. D. et al. Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief. Bioinform.* **17**, 877–890 (2016).
214. Aziz, R. K. et al. The RAST server: rapid annotations using subsystems technology. *BMC Genom.* **9**, 75 (2008).
215. Seaver, S. M. D. et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* **49**, D575–D588 (2021).
216. Henry, C. S. et al. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* **28**, 977–982 (2010).
217. Faria, J. P. et al. ModelSEED v2: high-throughput genome-scale metabolic model reconstruction with enhanced energy biosynthesis pathway prediction. Preprint at https://doi.org/10.1101/2023.10.04.556561 (2023).
218. Norsigian, C. J. et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res*. gkz1054, https://doi.org/10.1093/nar/gkz1054 (2019).
219. The UniProt Consortium et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

220. Saier, M. H. et al. The Transporter Classification Database (TCDB): 2021 update. *Nucleic Acids Res.* **49**, D461–D467 (2021).
221. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
222. Wang, H. et al. RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput. Biol.* **14**, e1006541 (2018).
223. Capela, J. et al. *merlin*, an improved framework for the reconstruction of high-quality genome-scale metabolic models. *Nucleic Acids Res.* **50**, 6052–6066 (2022).
224. Lewis, N. E. et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390 (2010).
225. Jenior, M. L., Glass, E. M. & Papin, J. A. Reconstructor: a COBRApy compatible tool for automated genome-scale metabolic network reconstruction with parsimonious flux-based gap-filling. *Bioinformatics* **39**, btad367 (2023).
226. Aite, M. et al. Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models. *PLoS Comput. Biol.* **14**, e1006146 (2018).
227. Heirendt, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).
228. Chan, S. H. J., Simons, M. N. & Maranas, C. D. SteadyCom: predicting microbial abundances while ensuring community stability. *PLoS Comput. Biol.* **13**, e1005539 (2017).
229. Heinken, A. & Thiele, I. Microbiome modelling toolbox 2.0: efficient, tractable modelling of microbiome communities. *Bioinformatics* **38**, 2367–2368 (2022).
230. Zomorrodi, A. R. & Maranas, C. D. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput. Biol.* **8**, e1002363 (2012).
231. Zhuang, K. et al. Genome-scale dynamic modeling of the competition between *Rhodoferax* and *Geobacter* in anoxic subsurface environments. *ISME J.* **5**, 305–316 (2011).
232. Gomez, J. A., Höffner, K. & Barton, P. I. DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis. *BMC Bioinform.* **15**, 409 (2014).
233. Zomorrodi, A. R., Islam, M. M. & Maranas, C. D. d-OptCom: dynamic multi-level and multi-objective metabolic modeling of microbial communities. *ACS Synth. Biol.* **3**, 247–257 (2014).
234. Zelezniak, A. et al. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci.* **112**, 6449–6454 (2015).
235. Carr, R. & Borenstein, E. NetSeed: a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment. *Bioinformatics* **28**, 734–735 (2012).
236. Levy, R., Carr, R., Kreimer, A., Freilich, S. & Borenstein, E. NetCooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinform.* **16**, 164 (2015).

Acknowledgements

This study is part of the DOMINO project. This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101060218.

Author contributions

A.K. and S.B. wrote the original draft of the manuscript. A.K. made the figures, and tables and edited the manuscript. J.G.K. and J.M. conceived the work and revised and edited the manuscript. A.K.O., P.C., O.S. and S.M. revised the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Anik Khan, Jennifer Mahony or John G. Kenny.