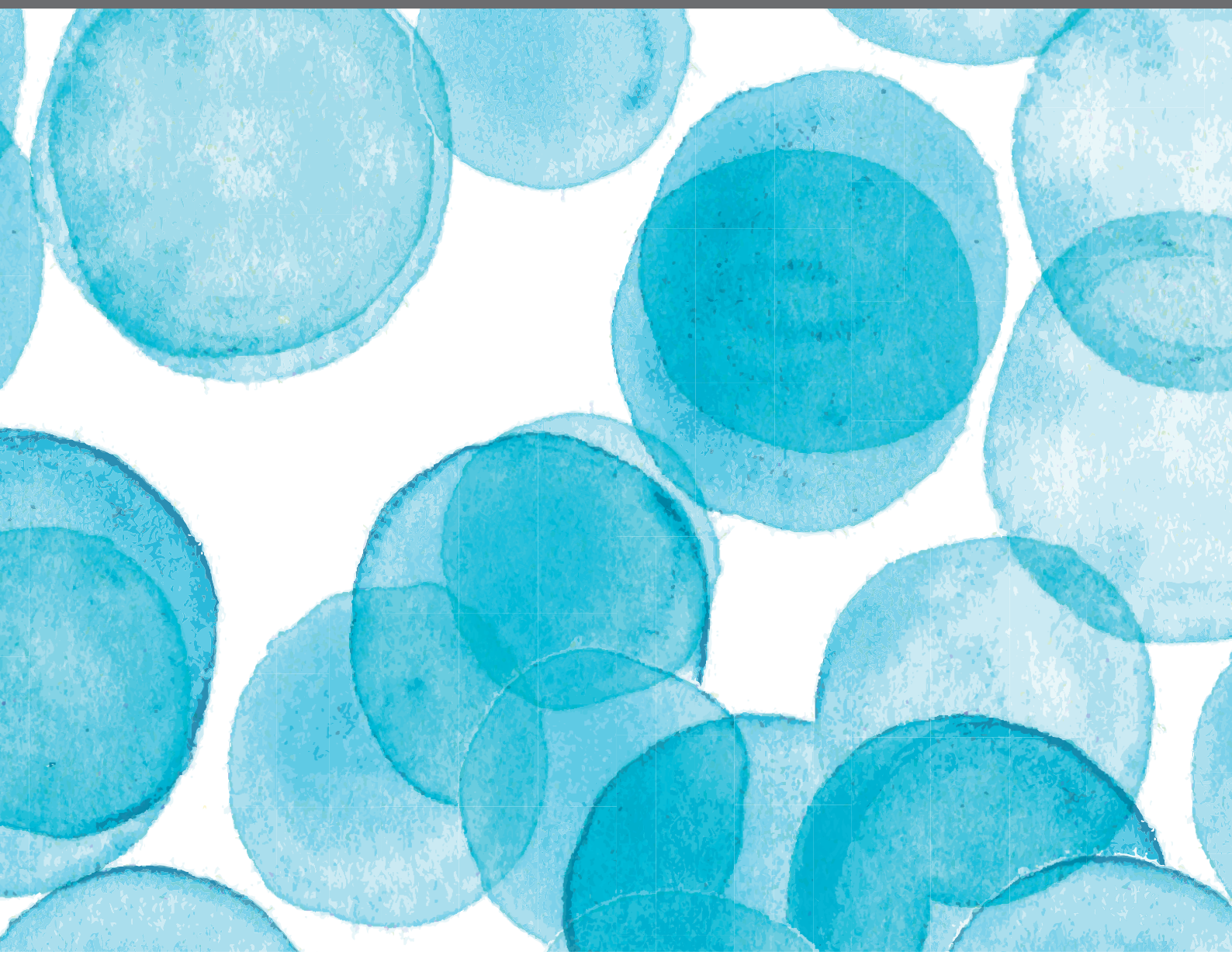
GENOMICS AND METAGENOMICS APPROACHES FOR FOOD VALUE CHAIN QUALITY, SAFETY, AND PRODUCT DEVELOPMENT

EDITED BY: Caroline Barretto, Nicholas Bokulich and Jerome Combrisson
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-855-8
DOI 10.3389/978-2-83250-855-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

GENOMICS AND METAGENOMICS APPROACHES FOR FOOD VALUE CHAIN QUALITY, SAFETY, AND PRODUCT DEVELOPMENT

Topic Editors:
Caroline Barretto, Nestle (Switzerland), Switzerland
Nicholas Bokulich, ETH Zürich, Switzerland
Jerome Combrisson, Mars (United States), United States

Citation: Barretto, C., Bokulich, N., Combrisson, J., eds. (2022). Genomics and Metagenomics Approaches for Food Value Chain Quality, Safety, and Product Development. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-855-8

Table of Contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 10 November 2022
DOI 10.3389/fmicb.2022.1056796

OPEN ACCESS

EDITED AND REVIEWED BY
Aldo Corsetti,
University of Teramo, Italy

*CORRESPONDENCE
Caroline Barretto
caroline.barretto@rdls.nestle.com

SPECIALTY SECTION
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 29 September 2022
ACCEPTED 21 October 2022
PUBLISHED 10 November 2022

CITATION
Barretto C, Combrisson J and Bokulich N (2022) Editorial: Genomics and metagenomics approaches for food value chain quality, safety, and product development. *Front. Microbiol.* 13:1056796. doi: 10.3389/fmicb.2022.1056796

Editorial: Genomics and metagenomics approaches for food value chain quality, safety, and product development

Caroline Barretto[1]*, Jerome Combrisson[2] and Nicholas Bokulich[3]

[1]Nestlé Research Center, Lausanne, Switzerland, [2]Mars, McLean, VA, United States, [3]Institute of Food, Nutrition, and Health, ETH Zürich, Zurich, Switzerland

KEYWORDS

genomics, metagenomics, food safety, food quality, product development

Editorial on the Research Topic
Genomics and metagenomics approaches for food value chain quality, safety, and product development

This Research Topic collects different contributions highlighting how genomics and metagenomics techniques are revolutionizing investigations of microbial activity along the complete food value chain, from farm to fork, advancing insight into functional roles of microbial strains and consortia in food quality and safety. Advances in 2nd- and 3rd-generation DNA sequencing technologies and computational methods for microbial data science have paved the way for multi-omics characterization of microbial communities in agriculture and food production, with a myriad of applications including pathogen tracing, functional characterization, and microbial ecology. Genomics applied to food production will become a new trend for routine food chain management, but critical assessment and future direction are needed, including better standardization, interoperability, and quality control in food applications.

Food safety and quality management regimes typically employ traditional methods for microbial detection and characterization, which have known limitations including levels of detection and resolution. Genomics and metagenomics methods have numerous benefits that overcome these limitations, offering potential applications that could improve food safety, quality, and sustainability. Nevertheless, other limitations must be addressed, including standardization and best practices for food applications.

The goal of this Research Topic is to critically evaluate applications of sequencing technologies and analysis techniques in food quality and safety throughout the food value chain. This encompasses genomics and metagenomics applications that span raw materials production to final products.

Manuscripts include applications of 2nd- and 3rd-generation sequencing technologies for genomics and metagenomics in food microbiology, e.g., genomics of foodborne pathogens or beneficial food microbes, bioinformatics techniques and

digital resources with specific applications in food quality and safety, foods as model systems for microbial ecology, and genomics and metagenomics applications in food fermentation.

The first article of this Research Topic (Djemiel et al.) gives an overview of -omics technologies contribution to the monitoring of the soil microbial quality, notably exploring the use of various methods to investigate the microbial diversity of soils. While showing the general principles, methodologies, specificities and benefits, the article also describes some of the challenges that need to be overcome to achieve a better understanding of soil microbial complexity.

Raw material monitoring is key to avoiding pathogens entering production in food facilities. The risks of having *Cronobacter* or *Salmonella* pathogens entering the food chain is highlighted in the article from Parra-Flores et al. The authors show the benefit of using whole genome sequencing (WGS), sequence typing prediction, determination of serotypes, and antibiotic resistance profiles on those pathogens.

Borges et al. describes how omics techniques have revolutionized food microbiology including bio-preservation to improve the microbial safety of various foods, the future being food microbiome engineering by considering food as a complex and dynamic microbiome, opening the field of innovation of production.

The genomes of six *Weissella* species were sequenced and analyzed by Fanelli et al. allowing the authors to demonstrate the strong carbohydrate utilization capabilities of the sequenced strains. The increasing availability of the genomic sequences of the *Weissella* species contributes to improving the knowledge about this genus and identifying the features defining its role in fermentative processes.

Wambui et al. reports a comprehensive genomic analysis of the *Clostridium estertheticum* complex (CEC) belonging to a diverse group of bacteria associated with spoilage of chilled vacuum-packed meat. The comprehensive comparative genomics of 50 genomes supports the reconstruction and comparison of intra- and interspecies metabolic pathways linked to meat spoilage.

Braley et al. addressed the gut microbiota of meat-producing animals as being one of the most important sources of surface contamination of processed carcasses, a concern for both food safety and for the shelf life of pork meat products. Based on their observations, even with optimal primary processing practices, gut microbiota modifications may not have any profound effects on carcass microbiota. This is an important learning for the control of bacterial carcass surface contamination in slaughterhouses.

Using both culture-dependent methods and 16S rRNA gene sequencing, Liao and Wang illustrated that there was no significant difference in bacterial abundance, diversity, and composition of bacterial communities in three brands of Romain Lettuce (RL) on the "Use By" dates (UBD) and on the 5 days after the "Use By" dates (UBD5), suggesting that the microbial quality of RL remained the same at two storage time points. Interestingly, bacterial communities present in RL samples were influenced by "brand" and "harvesting season" factors.

Ishiya et al. determined the whole genome sequences of three bacterial strains, named FNDCR1, FNDCF1, and FNDCR2, isolated from a practical nata-de-coco producing bacterial culture. The 16S rDNA sequence and phylogenetic analysis revealed that all strains belonged to a different clade within the *Komagataeibacter* genus.

The aim of the Nouws et al. study was to assess the potential benefits of WGS compared with conventional molecular methods currently used in the investigation of Staphylococcal food poisoning (SFP) outbreaks. Complete Staphylococcal enterotoxin (*se*) gene profiling of isolates from various sources, and relatedness between isolates were investigated using WGS, enabling complete *se* gene profiling with high performance, in contrast to PCR-based *se* gene detection for which primers sometimes showed to be non-specific.

Finally, Buytaers et al. assessed new methodologies to investigate food samples in the event of food-borne outbreaks. The authors present a proof-of-concept using shiga toxin-producing *Escherichia coli* for a strain-level metagenomics food-borne outbreak investigation method using the MinION and Flongle flow cells from Oxford Nanopore Technologies. Strain-level characterization could be achieved, linking the food containing a pathogen to the related human isolate of the affected patient, by means of a SNP-based phylogeny, with the use of long reads applying a bioinformatics method.

Together, this Research Topic demonstrates applications of genomics and metagenomics approaches to better elucidate the role of microbial ecology on food safety and quality assurance from farm to fork. In addition to showcasing new technologies and applications for food chain monitoring, several articles in this Research Topic highlight the need for discussion within the international scientific community to overcome several challenges in utilizing and translating these technologies, including the need for well-validated international standards for harmonization and interoperability.

Author contributions

CB, JC, and NB wrote the manuscript. All authors made a direct and intellectual contribution to the work and approved the final version for publication.

Conflict of interest

Author CB is employed by Nestlé. Author JC is employed by Mars.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Microbiology

ORIGINAL RESEARCH
published: 02 November 2021
doi: 10.3389/fmicb.2021.750278

Whole Genome Sequencing Provides an Added Value to the Investigation of Staphylococcal Food Poisoning Outbreaks

*Stéphanie Nouws[1,2,3], Bert Bogaerts[1,2,3], Bavo Verhaegen[4], Sarah Denayer[4], Lasse Laeremans[5], Kathleen Marchal[2,3,6], Nancy H. C. Roosens[1], Kevin Vanneste[1] and Sigrid C. J. De Keersmaecker[1]**

[1] Transversal Activities in Applied Genomics, Sciensano, Brussels, Belgium, [2] IDLab, Department of Information Technology, Ghent University - IMEC, Ghent, Belgium, [3] Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, [4] National Reference Laboratory for Foodborne Outbreaks (NRL-FBO) and for Coagulase Positive Staphylococci (NRL-CPS), Foodborne Pathogens, Sciensano, Brussels, Belgium, [5] Organic Contaminants and Additives, Sciensano, Brussels, Belgium, [6] Department of Genetics, University of Pretoria, Pretoria, South Africa

OPEN ACCESS

Edited by:
Caroline Barretto,
Nestle (Switzerland), Switzerland

Reviewed by:
Roger Stephan,
University of Zurich, Switzerland
Valentine Usongo,
Health Canada, Canada
Hisaya Ono,
Kitasato University, Japan

***Correspondence:**
Sigrid C. J. De Keersmaecker
Sigrid.DeKeersmaecker@Sciensano.be

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: 30 July 2021
Accepted: 04 October 2021
Published: 02 November 2021

Citation:
Nouws S, Bogaerts B, Verhaegen B, Denayer S, Laeremans L, Marchal K, Roosens NHC, Vanneste K and De Keersmaecker SCJ (2021) Whole Genome Sequencing Provides an Added Value to the Investigation of Staphylococcal Food Poisoning Outbreaks. Front. Microbiol. 12:750278. doi: 10.3389/fmicb.2021.750278

Through staphylococcal enterotoxin (SE) production, *Staphylococcus aureus* is a common cause of food poisoning. Detection of staphylococcal food poisoning (SFP) is mostly performed using immunoassays, which, however, only detect five of 27 SEs described to date. Polymerase chain reactions are, therefore, frequently used in complement to identify a bigger arsenal of SE at the gene level (*se*) but are labor-intensive. Complete *se* profiling of isolates from different sources, i.e., food and human cases, is, however, important to provide an indication of their potential link within foodborne outbreak investigation. In addition to complete *se* gene profiling, relatedness between isolates is determined with more certainty using pulsed-field gel electrophoresis, *Staphylococcus* protein A gene typing and other methods, but these are shown to lack resolution. We evaluated how whole genome sequencing (WGS) can offer a solution to these shortcomings. By WGS analysis of a selection of *S. aureus* isolates, including some belonging to a confirmed foodborne outbreak, its added value as the ultimate multiplexing method was demonstrated. In contrast to PCR-based *se* gene detection for which primers are sometimes shown to be non-specific, WGS enabled complete *se* gene profiling with high performance, provided that a database containing reference sequences for all *se* genes was constructed and employed. The custom compiled database and applied parameters were made publicly available in an online user-friendly interface. As an all-in-one approach with high resolution, WGS additionally allowed inferring correct isolate relationships. The different DNA extraction kits that were tested affected neither *se* gene profiling nor relatedness determination, which is interesting for data sharing during SFP outbreak investigation. Although confirming the production of enterotoxins remains important for SFP investigation, we delivered a proof-of-concept that WGS is a valid alternative and/or complementary tool for outbreak investigation.

Keywords: whole genome sequencing, *Staphylococcus aureus*, staphylococcal food poisoning (SFP), outbreak investigation, enterotoxin gene profiling, relatedness determination, DNA extraction kit

INTRODUCTION

Through the production of staphylococcal enterotoxins (SEs), *Staphylococcus aureus* is globally one of the most common causative agents responsible for food poisoning outbreaks [European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC), 2021]. Detection of staphylococcal food poisoning (SFP) is, therefore, mainly based on the detection of SEs in food leftovers, the enumeration of $>10^5$ *S. aureus* CFU/g food, and the isolation of *S. aureus* from the poisoned food source and affected human cases (Hennekinne et al., 2010).

For SE detection, commercial kits using immunological methods have been developed, such as the automated VIDAS® immunoanalyzer or the Staphylococcal Enterotoxin Reversed Passive Latex Agglutination (SET-RPLA) kit, and are widely used by National Reference Laboratories (NRLs) and Centers, often under standardized ISO 19020:2017 environments (Hennekinne et al., 2010). However, these kits only allow detection of at most five SEs (the classical five SE: SEA to SEE), whereas an arsenal of 27 different enterotoxins is described to date (Merda et al., 2020). Moreover, due to insertion, deletion, duplication, and recombination events in their encoding genes, new SEs [such as SELV and SELU arising from fusion between *sem* and *sei* (Thomas et al., 2006) and Ψ*ent1* and Ψ*ent2* pseudogenes (Letertre et al., 2003b; Thomas et al., 2006), respectively] and SE variants (such as SEC_{1-4}, SEC_{bovine}, SEC_{ovine}, SELU2, and others) are continuously generated. Many of the described SEs are determined to exhibit emetic properties [in alphabetical order: SEA to SEE, SEG to SEI, SEK to SET, and SEY (Spaulding et al., 2013; Johler et al., 2015; Wilson et al., 2018; Zhang et al., 2018; Aung et al., 2020)]. Moreover, it is most likely that multiple of these toxins are produced during SFP and play a (contributing) role in the provoked illness (Umeda et al., 2017; Fisher et al., 2018). Therefore, it is important to detect all toxins involved, not only during SFP outbreak investigation, but also during SFP surveillance, to assess the possible risk of strains isolated from food matrices in view of illness prevention. Because the development of new immunological assays remains difficult (Fisher et al., 2018), molecular methods, such as polymerase chain reactions (PCRs) are frequently used to detect SE encoding genes (*se*) (Fisher et al., 2018). Although no information on the eventual production of the toxin is obtained, detection at the genotypic level demonstrates the potential for toxin production. In response, many PCR primer pairs have been developed detecting the different *se* genes (for references, see **Supplementary Table 1**). However, the fact that multiple primer sets per se gene are described across the literature and not all are exhaustively investigated for their specificity impedes the selection of the most appropriate primers for routine applications. Indeed, the high sequence similarity between *se* (pseudo)genes and mutations in the primer binding sites of *se* variants increases the risk of false positives and negatives, respectively. Moreover, despite efforts in developing multiplex PCR assays targeting multiple *se* genes per assay (Becker et al., 1998; Schmitz et al., 1998; Monday and Bohach, 1999; Mehrotra et al., 2000; Sharma et al., 2000; Letertre et al., 2003a; Shylaja et al., 2010; Nagaraj et al., 2014), detecting the full arsenal of *se* genes currently described (from now on referred to as "complete *se* gene profiling") remains a labor-intensive and time-consuming process.

Complete *se* gene profiling can, however, also provide an initial indication on the potential link between the contaminated food source and the human case(s) (Denayer et al., 2017). Yet, because food sources are sometimes contaminated with several *S. aureus* strains, the level of isolate discrimination based on the *se* gene profile might not be sufficient. Indeed, accurate epidemiological typing methods are a prerequisite to distinguish between isolated strains and aiding in identifying the causal strain for SFP outbreaks. For this purpose, pulsed-field gel electrophoresis (PFGE), multiple-locus variable number of tandem repeats analysis (MLVA), and phage typing are still regarded as the gold standard to study isolate relatedness (Blair and Williams, 1961; Wildemauwe et al., 2004; Roussel et al., 2015). However, these methods have shown some issues with interpretation and/or interlaboratory comparison of obtained profiles (Cookson et al., 1996; Murchan et al., 2003; Price et al., 2013). Therefore, DNA sequence-based methods, such as *Staphylococcus* protein A gene (*spa*) typing and multilocus sequence typing (MLST) are frequently applied as well because their analysis is less ambiguous. Also for MLVA, DNA sequencing of alleles has been used to increase the initially limited transferability of obtained data by confirming the deduced number of repeats (Schouls et al., 2009). However, all these conventional methods are recently shown to struggle in distinguishing closely related *S. aureus* isolates (Moore et al., 2015; Roussel et al., 2015; Cunningham et al., 2017; Cremers et al., 2020).

Because whole genome sequencing (WGS) analyzes the entire genome of bacteria, it enables complete genotypic characterization of isolates. Thanks to its unparalleled resolution standing out compared with conventional methods (Anderson et al., 2014; Moore et al., 2015; Cunningham et al., 2017; Cremers et al., 2020), it has moreover become the ultimate tool for inferring phylogenetic relationships between bacterial isolates, using core genome MLST (cgMLST) or single nucleotide polymorphism (SNP) analysis. Although WGS was already used to analyze Methicillin-resistant *S. aureus* strains responsible for nosocomial infections (Anderson et al., 2014; Moore et al., 2015; Peacock and Paterson, 2015; Cunningham et al., 2017; Sabat et al., 2017), it has been scarcely applied for SFP investigation despite the advantages highlighted above (Merda et al., 2020; Schwendimann et al., 2021).

For WGS to be applied in the analysis of *S. aureus* isolates for se gene profiling, it is important to have a DNA sample representative for the isolate so that WGS data are comparable between (inter)national laboratories. Data sharing is indeed indispensable in the scope of outbreak investigations. For *S. aureus*, this means that the extracted genome must contain DNA of mobile genetic elements, such as plasmids, as they often harbor *se* genes (Argudín et al., 2010). Some frequently used, commercial DNA extraction kits have, however, been described as having decreased plasmid extraction performances (Becker et al., 2016; Nouws et al., 2020a), which can potentially impact complete *se* gene profiling. The successful integration of WGS-based *se* gene profiling is, additionally, hindered by the lack of comprehensive and/or freely and easily accessible databases.

Indeed, commonly used databases for virulence gene detection in *S. aureus* do not contain or correctly annotate the full arsenal of *se* genes currently described, are not publicly available through an easily accessible online resource, and/or are not always linked to a user-friendly interface to query WGS data for *se* genes (Joensen et al., 2015; Liu et al., 2019; Sayers et al., 2019; Merda et al., 2020).

This study aimed to assess whether WGS would have an added value for application in the investigation of SFP outbreaks. For this purpose, WGS was used for *se* gene detection with a custom database [based on the frequently used Virulence Factor Database (VFDB)] of in-house sequenced isolates, complemented with publicly available WGS data. The determined *se* profiles were compared with those previously obtained with conventional methods supplemented with *in silico* PCR. Different DNA extraction kits were moreover compared for their appropriateness to use for WGS-based *se* gene profiling. The performance of WGS in *S. aureus* isolate relatedness determination and the potential influence of using different DNA extraction kits on this analysis was evaluated by comparison with *a priori* known relationships.

MATERIALS AND METHODS

Selected Isolates and Their Characteristics

Criteria for Isolate Selection

To evaluate the performance of WGS in determining *se* gene profiles, isolates (**Table 1**) were selected to cover a repertory as extensive as possible of different *se* genes. For each of the enterotoxins (genes) whose presence was assessed previously, an in-house available isolate was included, i.e., *sea* to *see*, *seg* to *selj*, *sep*, and *ser* as determined by the multiplex PCR of the European Union Reference Laboratory for Coagulase Positive Staphylococci (EURL-CPS) (Roussel et al., 2015) at genotypic level and/or SEA to SEE as determined by the immunoassays SET-RPLA (detection and separate typing of SEA to SED) and VIDAS (detection of SEA to SEE) analyses at the phenotypic level. Isolates that were part of an SFP outbreak [(Denayer et al., 2017), i.e., outbreak A], previously confirmed based on PFGE profiles, were also included to allow investigating if WGS has an added value in resolving the SFP outbreak as proof of concept. All isolates were provided by the Belgian NRL for Foodborne Outbreaks (NRL-FBO) and Coagulase Positive Staphylococci (NRL-CPS), from which two isolates were initially received from the EURL-CPS as reference strains, i.e., TIAC3971 (S-6) and TIAC3972 (FRI-362).

For *se* genes not covered within the selected isolates, public WGS raw reads and/or assemblies of *S. aureus* isolates were incorporated (**Table 1**) to cover the complete arsenal of *se* genes. Accession numbers of the pubic WGS data employed in this study are provided in **Supplementary Table 2**.

Ethics Approval

This study includes *S. aureus* strains isolated from human feces or swabs from human cases and food matrices. Besides the origin of the strain, no human data were applied in this study, and thus, no ethical approval or written informed consent was required.

Determining the Complete *se* Gene Profile

Because no information concerning the presence/absence of *se*/SE other than those tested with the EURL-CPS multiplex PCR and/or SET-RPLA/VIDAS commercial kits was initially available (i.e., limited to a maximum of 11 out of the 27 *se*/SE), *in silico* PCR was performed to obtain the complete *se* gene profile of each isolate. The same was done for the isolates for which publicly available WGS data were used as a confirmation of the reported *se* genes in the corresponding publications [(Fursova et al., 2020; Merda et al., 2020); see **Table 1** and **Supplementary Table 2**]. For this *in silico* PCR, the literature was searched to collect all reported conventional PCR primer sets (n: 155) for detection of 27 *se* genes and two pseudogenes (**Supplementary Table 1**).

Each primer pair described in the literature was checked for its specificity to the target gene (i.e., no aspecific detection of other staphylococcal genes with similar amplicon size) and, if not yet known, its amplicon size, by aligning the primer pairs against the NCBI nucleotide collection (nt) for *S. aureus* (taxid: 1280) using the Primer-BLAST tool of NCBI (Ye et al., 2012) through its online resource[1] (see **Supplementary Data**). Besides the criteria previously described (Vanneste et al., 2018), default parameters were applied that assessed whether a primer pair allowed the formation of an *in silico* amplicon.

Based on the investigated specificity of all primer sets, 89.7% of the complete list of collected primers were used for the *in silico* PCR. For this purpose, these 139 primer pairs were aligned against the assemblies of the 13 in-house sequenced *S. aureus* isolates (see **Supplementary Data**) processed with the GenElute Bacterial gDNA kit and the downloaded assemblies of the three isolates for which WGS data were publicly available, using Primer-BLAST. A gene was considered present when at least one of the gene-specific primer pairs yielded an *in silico* product with correct amplicon size according to the preset criteria. Only when a gene was not detected by any of the primer sets was it considered absent.

Because conventional PCR does not enable discriminating between *selu*, its variant *selu2*, and the pseudogenes Ψ*ent1* and Ψ*ent2* due to high-sequence similarity (Heymans et al., 2010; Liang et al., 2016), their detection is generally followed by DNA Sanger sequencing for confirmation (Collery and Smyth, 2007). Therefore, when the primer pairs indicated the presence of *selu*, *selu2*, or Ψ*ent1-2*, *in silico* PCR was repeated using a primer pair [forward: 5′– TGA TAA TTA GTT TTA ACA CTA AAA TGC G-3′; reverse: 5′– CGT CTA ATT GCC ACG TTA TAT CAG T-3′; (Letertre et al., 2003b)] targeting the complete gene length. The amplicon sequence was extracted and aligned against those of NCBI sequences (see **Supplementary Figure 1**) harboring *selu* (Genbank reference AY158703.1), *selu2* (Genbank reference MN450302.1), and Ψ*ent1-2* (Genbank reference MN450303.1) using CLC Sequence Viewer 8.0 to type the specific gene based on similarity with the reference sequences.

[1] https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi

TABLE 1 | Characteristics of the selected in-house isolates and isolates for which WGS data were publicly available.

Isolate name/ BioSample accession	Relation	Original matrix	Detection at protein level					Detection at gene level																														
			SEA	SEB	SEC	SED	SEE	*sea*	*seb*	*sec*	*sed*	*see*	*seg*	*seh*	*sei*	*sej*	*sek*	*sel*	*sem*	*sen*	*seo*	*sep*	*seq*	*ser*	*ses*	*set*	*selu*		*selv*	*selw*	*selx*	*sey*	*selz*	*sel26*	*sel27*	*Ψent1*	*Ψent2*
																										u	*u2*											
TIAC1840	NO	Food	0	0	0	0	0	0*	0*	0*	0*	0*	0*	0*	0*	0*	0	0	0	0	0	1*	0	0*	0	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC1848	O	Human	1	0	0	1	0	1*	0*	0*	1*	0*	1*	0*	1*	1*	0	0	1	1	1	0*	0	1*	0	0	0	0	0	1	1	0	0	0	0	1	1
TIAC1991	NO	Human	1	0	0	0	0	1*	0*	0*	0*	0*	0*	1*	0*	0*	1	0	0	0	0	0*	1	0*	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC1992	NO	Human	0	0	1	0	0	0*	0*	1*	0*	0*	1*	0*	1*	0*	0	1	1	1	1	0*	0	0*	0	0	0	1	0	1	1	0	0	0	0	0	0
TIAC1993°	NO	Human	1	0	0	0	0	1*	0*	0*	0*	0*	0*	1*	0*	0*	1	0	0	0	0	0*	1	0*	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC1994°	NO	Human	1	0	0	0	0	1*	0*	0*	0*	0*	0*	1*	0*	0*	1	0	0	0	0	0*	1	0*	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC2001	NO	Food	0	0	0	0	0	0*	0*	0*	0*	0*	0*	0*	0*	0*	0	0	0	0	0	1*	0	0*	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC3462	NO	Food	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
TIAC3971	NO	Food	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC3972	NO	Food	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0
TIAC1798	O	Food	1	0	0	0†	0	1*	0*	0*	1*	0*	1*	0*	1*	1*	0	0	1	1	1	0*	0	1*	0	0	0	0	0	1	1	0	0	0	0	1	1
TIAC1847	O	Human	1	0	0	1	0	1*	0*	0*	1*	0*	1*	0*	1*	1*	0	0	1	1	1	0*	0	1*	0	0	0	0	0	1	1	0	0	0	0	1	1
TIAC3152	NO	Food	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0
SAMN02391177 (Merda et al., 2020)	UN	Wild boar	UN	UN	UN	UN	UN	0	0	0	0	0	1	0	1	0	0	0	1**	1	1	0	0	0	0	0	0	1	0	1**	1	1	0	1	1	0	0
SAMN02403200 (Merda et al., 2020)	UN	Human	UN	UN	UN	UN	UN	0	0	0	0	0	1	0	1	1	0	0	1	1	1	0	0	1	1	1	0	1	0	1**	1**	0	0	0	0	0	0
SAMN13134218 (Fursova et al., 2020)	UN	Food	UN	UN	UN	UN	UN	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	0	0	0	1	1	0**	1**	1**	0	0	0	0	0	0

The table contains an overview of all characteristics of the selected isolates for this study (in-house isolates have a name starting with "TIAC," and publicly available data are listed with their NCBI BioSample accession), the relatedness between the isolates as determined previously with PFGE {indicated as being part of an outbreak (O; and isolate name in bold) [(Denayer et al., 2017); Outbreak A], not being part of an outbreak (NO), or unknown (UN)}, their origin ("original matrix"), and results on their determined SE/se profile. TIAC1993 and TIAC1994 were isolated from nose and throat samples, respectively, from the same person (indicated with "°"). TIAC3971 and TIAC3972 are reference strains S-6 (Lopes et al., 1993) and FRI-326 (Bergdoll et al., 1971), respectively, sent by the EURL-CPS. Results of the immunological assays performed in routine are shown if known (UN: unknown), i.e., "Detection at protein level." Results on the previously performed multiplex PCR (targeted genes are indicated with ""), in silico PCR (this study), and previously in literature reported WGS analysis [discrepant results with the in silico PCR are indicated with "**" (Fursova et al., 2020; Merda et al., 2020)] are indicated as "Detection at gene level." Toxin or gene presence is indicated with "1" in a green box, absence with "0" in a pink box. Subtyping of se variants was not performed with the routine analyses. Only when selu or selu2 were detected with in silico PCR, their presence was further investigated through comparing the amplicon sequence with those of selu (Genbank reference AY158703.1), selu2 (Genbank reference MN450302.1), and Ψent1-2 (Genbank reference MN450303.1) NCBI reference sequences. The in silico PCR results were fully congruent with those obtained from the previously performed multiplex PCR. The depicted in silico PCR result was only discrepant for some genes (**) in comparison with the previously reported WGS-based se profile (Fursova et al., 2020; Merda et al., 2020), i.e., "1**" in a green box indicated the gene was present with in silico PCR but not reported by the previous study, whereas "0**" in a pink box indicated the gene was identified as absent with in silico PCR but was detected by the previous study. † SE not detected with SET-RPLA but detected with the multiplex PCR. The result of the in silico PCR confirmed the result of the multiplex PCR and, thus, the presence of sed in TIAC1798.*

DNA Preparation and Quality Control

All 13 in-house sequenced isolates were preserved in a glycerol-brain heart infusion (BHI) broth stock (40.0%) at −80°C until analysis. A loopful of each stock was grown overnight (16 h at 37°C) on nutrient agar plates, and a single colony was then inoculated in 10 ml of BHI and grown overnight with shaking at 37°C and 200 rpm.

DNA extraction was performed on all 13 cultures and a blank sample (BHI incubated overnight) using the GenElute Bacterial Genomic DNA (gDNA) kit (Sigma-Aldrich, Missouri, United States) according to the manufacturer's protocol for Gram positive bacteria. Three of the 13 isolates (TIAC1798, TIAC1847, and TIAC3152) selected based on previously determined presence of *se*/SE exclusively located/encoded on plasmids [i.e., *sed*, *selj*, *ser, ses,* and *set* (Bayles and Iandolo, 1989; Zhang et al., 1998; Omoe et al., 2003; Ono et al., 2008)] and on mutual relationships (i.e., two outbreak isolates and one non-outbreak isolate) were also prepared using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) and the Wizard gDNA Purification kit (Promega, Wisconsin, United States) (**Table 2**). These kits were selected based on their frequent use for WGS of *S. aureus* isolates (Gordon et al., 2014; Lee et al., 2015; Ekkelenkamp et al., 2018) and recommendations by leading institutes [EURL-CPS; (Merda et al., 2020)] in the field, respectively. The kits were used according to the manufacturer's instructions and proposed protocols (Merda et al., 2020), respectively. Moreover, the GenElute Bacterial gDNA kit was tested on the same three isolates according to the specific protocol for Staphylococcal species, which requires the use of the expensive lysostaphin (Sigma-Aldrich, Missouri, United States) during bacterial cell lysis.

DNA concentration, purity, and integrity were determined with, respectively, the dsDNA HS and BR assay kits for the Qubit 4 fluorometer (Thermo Fisher Scientific, Schwerte, Germany), NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Schwerte, Germany), and the Genomic DNA ScreenTape and Reagent kits for TapeStation 4200 electrophoresis (Agilent Technologies, Santa Clara, CA), according to the manufacturer's recommendations.

Library Preparation and Sequencing

One nanogram (in 5 μl) of each DNA extract was used for Nextera XT library preparation (Illumina, San Diego, CA). All libraries were sequenced on a MiSeq instrument (Illumina, San Diego, CA) using the MiSeq V3 chemistry to produce 2 × 250 bp paired-end reads, aiming at a theoretical sequencing depth of 60-fold.

Whole Genome Sequencing Data Analysis

All bioinformatic analyses were performed using the respective tools on an in-house instance of Galaxy (Afgan et al., 2018). An online user-friendly Galaxy interface to perform raw read trimming, assembly and *se* gene detection (among other functionalities) using the bioinformatic methods and criteria as described in this manuscript is publicly available upon registration.[2]

Read Trimming and Assembly

The raw reads were trimmed using Trimmomatic 0.38 (Bolger et al., 2014) by removing Nextera XT adaptors and other Illumina-specific sequences ("Illuminaclip" set to value "NexteraPE-PE.fa:2:30:10"), removing low-quality residues at the start and end of the reads ("leading:10" and "trailing:10"), clipping reads when average Q-scores dropped below 20 over a sliding window of four residues ("slidingwindow:4:20"), and dropping reads shorter than 40 bases after processing ("minlen:40"). Trimmed reads were *de novo* assembled using

[2]https://galaxy.sciensano.be/tool_runner?tool_id=pipeline_staphylococcus

TABLE 2 | Characteristics of selected DNA extraction kits.

DNA extraction kit	Price per sample* (€)	Average DNA conc. (ng/μL) ± s.d.	Average DNA yield (mg/ml starting material) ± s.d.	DNA purity (average ± s.d.)		Length range of DNA fragments (kb)	Remark
				A260/280	A260/230		
GenElute - NL	3.51	11.73 ± 2.23	2.93 ± 0.56	1.35 ± 0.02	0.80 ± 0.08	[55.78, >60.00]	Used by Belgian NRL-FBO (Nouws et al., 2020b)
GenElute	6.58	38.02 ± 9.47	9.50 ± 2.37	1.82 ± 0.03	1.87 ± 0.23	[30.18, >60.00]	
DNeasy	3.77	4.45 ± 2.46	0.78 ± 0.43	1.68 ± 0.22	0.96 ± 0.29	[20.38, >60.00]	Frequently used for WGS of *S. aureus* (Ekkelenkamp et al., 2018)
Wizard	6.49	15.90 ± 11.11	1.59 ± 1.11	2.08 ± 0.16	2.44 ± 0.12	>60.00	Used by EURL-CPS (Merda et al., 2020)

*The names of the DNA extraction kits are abbreviated, i.e., GenElute-NL: GenElute Bacterial gDNA kit using the protocol for Gram positive bacteria (without lysostaphin); GenElute: GenElute Bacterial gDNA kit using the protocol for Staphylococcal species (with lysostaphin); DNeasy: DNeasy Blood & Tissue kit; Wizard: Wizard gDNA Purification kit. For each of the used DNA extraction kits, averages and standard deviations (s.d.) of obtained DNA concentrations, yield, and purity were calculated from the three isolates (TIAC1798, TIAC1847, and TIAC3152). Fragment lengths are shown as ranges because the TapeStation Genomic DNA ScreenTape only gives exact measurements until 60 kb. *Prices as of February 2021 (excl. TVA, shipping, and handling costs). Prices were calculated from kits with highest throughput. With the exception of the cost of lysostaphin, the cost of extra products or materials required but not provided with the kit were not taken into account.*

SPAdes 3.13.0 (Bankevich et al., 2012) setting the "–careful" and "–cov-cutoff 10" options to reduce mismatches and short indels and remove low coverage contigs, respectively. Based on advice from a technical report by the ECDC, contigs below 1000 bp in length were removed using Seqtk seq 1.2[3] using the "-L" option to improve assembly quality (ECDC, 2019). Relevant assembly statistics (N50, number of contigs, and median coverage against assembly) were calculated with Quast 4.4 (Gurevich et al., 2013) and are shown in **Supplementary Table 3**.

Database for *se* Gene Detection

The different publicly available databases for virulence gene detection in *S. aureus*, i.e., the Victors database (Sayers et al., 2019), VirulenceFinder database (Joensen et al., 2014), and VFDB_Full (Liu et al., 2019), were checked for the presence of *se* gene reference sequences. The VFDB_Full was selected because it was the most complete in containing reference sequences for the 27 *se* genes and was retrieved from its respective source[4] on November 16, 2020. Some sequences were manually removed from and/or (re)added to the extended VFDB_Full database (i.e., *sep*, *ses*, *set*, *selw*, *selx*, *sey*, *selz*, *sel26*, and *sel27*; see **Supplementary Data**). Variants of *sec* (not separately annotated in the original VFDB_Full database) were also added so that variant subtyping of *sec* and *selu* (separately annotated in the original VFDB_Full database) was enabled. More information on sequences and accession numbers kept within (from the original VFDB_Full) and newly added to the extended VFDB_Full database is provided in **Supplementary Table 4**. The **Supplementary Data** also contains a FASTA file with sequences for all *se* genes present in the extended VFDB_Full database. Database sequences were clustered with an 85.0% sequence identity cutoff using the "cd-hit-est" function from CD-HIT 4.6.8 (Li and Godzik, 2006) to limit the detection of genes to one per cluster. This clustered database is also integrated in the online user-friendly interface for *se* gene detection.

Detection of *se* Genes Using the Extended Virulence Factor Database_Full Database

All samples were genotypically characterized for the presence of *se* genes, using two methods: (i) Aligning assemblies with BLAST + 2.6.0 (Camacho et al., 2009) and (ii) mapping trimmed reads with SRST2 0.2.0 (Inouye et al., 2014) with the options "–max-divergence 10," "–min-coverage 60," "–gene-max-mismatch 10," and "–max-unaligned-overlap 150," both against the extended VFDB_Full database. For the isolate with BioSample accession SAMN13134218, only the assembly was publicly available, limiting *se* gene detection to the assembly based approach using BLAST+ (i.e., no gene detection with SRST2 could be performed). Hits identified with <60.0% query coverage and/or >10.0% sequence divergence for SRST2 or with <60.0% query coverage and/or <90.0% sequence identity for BLAST+ were omitted. The best hit for each detected database cluster for BLAST+ was determined based on a previously described allele scoring method (Larsen et al., 2012). In the case of unexpected results, manual contig alignment against the corresponding reference gene using BLAST (Altschul et al., 1990) was performed to identify assembly artifacts. A gene was considered present with WGS when it could be detected with SRST2 and/or BLAST+.

Phylogenetic Analysis of the *S. aureus* Isolates Using Core Genome Multilocus Sequence Typing-Typing

In silico cgMLST-typing was performed as previously described (Bogaerts et al., 2021) by aligning assemblies using BLAST+ against the *S. aureus* cgMLST scheme of PubMLST (Jolley et al., 2018) containing 2208 loci (downloaded on January 3, 2021). Only exact allele calls (i.e., requiring a full-length, 100.0% identical match) were accepted. For tree construction, loci called in <80.0% of samples were stripped from the allele call matrix. A minimum spanning tree based on the allele call matrix was created using GrapeTree 1.5.0 (Zhou et al., 2018) with the "method" option set to "MSTreeV2′," and afterward visualized using FigTree 1.4.4 (Rambaut, 2007).

Identifying the Completeness of *se* Gene Profiling and Evaluating the Performance of Whole Genome Sequencing in Inferring Phylogenetic Relationships

To evaluate if WGS can identify the complete *se* gene profile of *S. aureus* isolates, the *se* gene profiles determined with SRST2 and BLAST+ against the extended VFDB_Full database were compared with the expected profiles obtained using *in silico* PCR. From this comparison, WGS-based *se* gene detection results were classified as either true positives (TPs), false positives (FPs), true negatives (TNs), or false negatives (FNs), for which definitions are shown in **Supplementary Table 5**.

Moreover, phylogenetic relationships inferred from the cgMLST profiles determined with WGS were compared to the *a priori* known relationships to evaluate the additional benefit of WGS in SFP investigation as a proof of concept.

The potential influences of the tested DNA extraction kits and protocols on both *se* gene detection and inferring phylogenetic relationships were assessed similarly.

RESULTS

Selected Isolates and Their Characteristics

To identify the benefits of WGS in SFP investigation, 13 in-house sequenced isolates (**Table 1**) were selected based on *a priori* known relationships to include outbreak (n: 3) and non-outbreak (n: 10) isolates and on the previously assessed presence of *se* and/or SE (i.e., *sea* to *see*, *seg* to *selj*, *sep*, and *ser*, and/or SEA to SEE). The selection of isolates included isolates of food (n: 7) and human (n: 6) origin, including two (TIAC1993 and TIAC1994) that were isolated from nose and throat of one human case. Because the previously identified *se*/SE profiles of the in-house sequenced isolates was limited to a maximum 11 out of 27 described *se* genes, the complete *se* gene profile per isolate

[3]https://github.com/lh3/seqtk

[4]http://www.mgc.ac.cn/VFs/download.htm

was extended with *in silico* PCR (**Supplementary Table 6**). To cover the complete repertoire of currently described *se* genes (*ses*, *set*, *selv*, *sey*, *sel26*, and *sel27* that were missing in the in-house sequenced isolates), a publicly available WGS data of three isolates were added (**Table 1**). The *in silico* PCR was similarly applied for the publicly available WGS data to confirm the previously reported *se* genes (Fursova et al., 2020; Merda et al., 2020). Based on the investigated specificity of all primer sets published across the literature (see **Supplementary Data** and **Supplementary Table 1**), 139 out of 155 primer pairs (i.e., 89.7%) were selected for the *in silico* PCR. From the *in silico* PCR results, seven sets (i.e., 5.0% of the 139 used sets) were, moreover, identified to not consistently detect their target genes even when present (see **Supplementary Data** and **Supplementary Table 6**). This is most likely because these primers were not designed to anneal in sufficiently conserved gene regions. More information on the *in silico* PCR can be found in the **Supplementary Data** and **Supplementary Table 6**. For *se* genes or SEs whose presence was identified with conventional methods (EURL-CPS multiplex PCR assays and/or VIDAS/SET-RPLA), results of the *in silico* PCR were fully congruent (**Table 1**). For all *se* genes that were previously reported to be present in the publicly available WGS data [see **Table 1** and **Supplementary Table 2** (Fursova et al., 2020; Merda et al., 2020)], their presence was accordingly identified with *in silico* PCR except for *selv* that was not detected in SAMN13134218 based on the *in silico* PCR (see **Supplementary Table 6** and **Table 1**) although it was previously reported to be present (Fursova et al., 2020). Moreover, across these three isolates analyzed with public data, *in silico* PCR enabled the detection of three additional *se* genes (i.e., *sem* in SAMN02391177, *selx* in SAMN02403200 and SAMN13134218, and *selw* in the three isolates; **Supplementary Table 6** and **Table 1**), previously not identified to be present (Fursova et al., 2020; Merda et al., 2020).

Identifying the Completeness of Whole Genome Sequencing-Based *se* Gene Profiling

By comparison with the expected *se* gene profiles (**Table 1**), it was assessed whether WGS could identify the complete *se* gene repertoire (reportable range) of *S. aureus* isolates using the extended VFDB_Full database (**Table 3**). The presence or absence of, in total, 27 different *se* genes, among which are *sec* and *selu* variants, and of two pseudogenes was determined. Compared with solely using the VFDB_Full database, the presence or absence of eight extra genes (*ses*, *set*, *selw* to *selz*, *sel26*, and *sel27*) and *sec* variants could, hence, be determined.

For *se* gene detection, the WGS-determined profile per isolate (**Table 3**) was identical to what was expected (i.e., based on the conventional methods and *in silico* PCR as indicated in **Table 1**). Indeed, across the, in total, 432 observations (27 *se* genes for each of the 16 isolates), i.e., 127 positive and 305 negative observations, there were no FPs or FNs. In comparison with the analyses performed during the initial SFP outbreak investigation [**Table 1**; (Denayer et al., 2017)], WGS provided information on the presence/absence of 17 extra *se* genes and additionally allowed subtyping of *sec* and *selu* variants (**Table 3**). For the isolates for which public WGS data were available (Fursova et al., 2020; Merda et al., 2020), the WGS-based *se* gene detection methods applied in this study using the extended VFDB_Full database enabled the detection of three extra *se* genes (i.e., *sem* in SAMN02391177, *selx* in SAMN02403200 and SAMN13134218, and *selw* in the three isolates) in line with the expected characteristics within **Table 1**. The *selv* gene that was previously detected in SAMN13134218 (Fursova et al., 2020) was missed with the WGS method using the extended VFDB_Full database in this study. However, manual alignment of the contigs with the NCBI reference sequence of *selv* (Genbank reference EF030427.1) with BLAST confirmed its absence as was similarly found with the *in silico* PCR and, hence, the results of this study. Between the *se* gene profiles obtained with SRST2 and with BLAST+, only one difference was observed for *se* gene detection, related to a fragmentation of the assembly in the *sec* gene for TIAC1992 that led to its missed detection only with BLAST+. By using the extended VFDB_Full database, WGS, thus, enabled complete *se* gene profiling using both the read mapping-based (SRST2) and assembly based (BLAST+) approach.

For the WGS-based detection of the Ψ*ent* pseudogenes, there were some discrepancies in isolates TIAC1798, TIAC1847, and TIAC1848 compared with what was expected (cfr. **Table 1**). One of both pseudogenes was systematically missed with BLAST+ and SRST2, which could be explained by the clustering of both genes in the database that limits their detection to only one of them. Differences in the detected Ψ*ent* pseudogene in TIAC1848 by BLAST+ and SRST2 can be explained by the fact that both methods use different allele scoring methods. However, manual alignment of the respective contigs with a Ψ*ent1-2* reference sequence (Genbank reference MN450303.1) confirmed the presence of both pseudogenes in the TIAC1798, TIAC1847, and TIAC1848 isolates with 100.0% sequence identity and query coverage (data not shown). Different DNA extraction kits and protocols (i.e., including lysostaphin in the bacterial lysis step) were tested for their influence on detecting *se* genes. Although lysostaphin increased the DNA yield, sufficient amounts and concentrations were obtained for Nextera XT library preparation without its usage [**Table 2**; (Illumina, 2018)]. For the detection of the 27 *se* genes, identical profiles were obtained with WGS for all samples (i.e., WGS data of an isolate processed with different DNA extraction kits) per isolate, irrespective of the DNA extraction kit, when compared with the expected *se* gene profiles. No problems in detecting any of the exclusively plasmid-encoded *se* genes (i.e., *sed*, *selj*, *ser*, *ses*, and *set*) were observed. When comparing *se* gene profiles obtained with SRST2 and BLAST+ separately for the samples per isolate, only one difference was obtained, related to the missed assembly-based detection of *sei* due to assembly fragmentation in TIAC1847 processed by the Wizard gDNA Purification kit. Therefore, no influence of the kit on the detection of *se* genes, whether or not plasmid-encoded, could be identified. Difficulties with the detection of both Ψ*ent* pseudogenes could again be observed across all samples of TIAC1798 and TIAC1847 with both detection methods, explained by the clustering of the pseudogenes in the database in combination

TABLE 3 | WGS-based *se* gene profile of the selected in-house isolates and influence of the DNA extraction kits on se gene profiling.

Isolate name/ Biosample accession	DNA extraction kit	*se* genes																														Pseudo-genes	
		sea	*seb*	*sec*						*sed**	*see*	*seg*	*seh*	*sei*	*sej**	*sek*	*sel*	*sem*	*sen*	*seo*	*sep*	*seq*	*ser**	*ses**	*set**	*selu*		*selv*	*selw*	*selx*	*sey*	*selz*	*sel26*	*sel27*	*ψent1*	*ψent2*
				c1	*c2*	*c3*	*c4*	*c bovine*	*c ovine*																		*u*	*u2*								
TIAC1840	GenElute-NL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC1848	GenElute-NL	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1[A]	1[B]
TIAC1991	GenElute-NL	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC1992	GenElute-NL	0	0	1[A]	0	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0		0	1	1	0	0	0	0	0	0
TIAC1993	GenElute-NL	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC1994	GenElute-NL	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC2001	GenElute-NL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC3462	GenElute-NL	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	0	0	0	0		0	1	0	0	0	0	0	0	0
TIAC3971	GenElute-NL	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
TIAC3972	GenElute-NL	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0
TIAC1798	GenElute-NL	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0[C]	1
	GenElute	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0[C]	1
	DNeasy	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1[A]	1[B]
	Wizard	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1[A]	1[B]
TIAC1847	GenElute-NL	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0[C]	1
	GenElute	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1[A]	1[B]
	DNeasy	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1[A]	1[B]
	Wizard	1	0	0	0	0	0	0	0	1	0	1	0	1[A]	1	0	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1[A]	1[B]
TIAC3152	GenElute-NL	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0
	GenElute	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0
	DNeasy	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0
	Wizard	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0
SAMN02391177	UN	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	0	0	0	0	1	0	1	1	1	0	1	1	0	0
SAMN02403200	UN	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	1	1	0	0	1	1	1	0	1	0	1	1	0	0	0	0	0	0
SAMN13134218[D]	UN	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0

*For each isolate, the se gene profile determined with WGS based on the extended VFDB_Full database is indicated. Isolates originating from the same outbreak are shown in bold. Gene presence is indicated as "1" in a green box, absence as "0" in a pink box. When a gene is detected only by SRST2 or BLAST+, it is shown with "1" in a light green box. When a gene was missed compared to the in silico PCR results (**Table 1**), it is indicated with "0" in a light pink box. Gene names with an asterisk "*" are genes exclusively encoded on plasmids. Gene names indicated in blue are genes whose presence/absence could solely be identified by using the extended VFDB_Full database (because these genes were absent in the original VFDB_Full database). Enterotoxin genes whose variants were determined with the extended VFDB_Full database, are shown. For the three isolates that were tested with the different DNA extraction kits, the se gene profiles are also shown. The names of the DNA extraction kits are abbreviated, i.e., GenElute-NL: GenElute Bacterial gDNA kit using the protocol for Gram positive bacteria (without lysostaphin); GenElute: GenElute Bacterial gDNA kit using the protocol for Staphylococcal species (with lysostaphin); DNeasy: DNeasy Blood & Tissue kit; Wizard: Wizard gDNA Purification kit; UN: unknown.*

[A]Gene not detected with BLAST+.

[B]Gene not detected with SRST2.

*[C]Discrepant result for the pseudogenes, compared to the expected se gene profile as determined with in silico PCR (**Table 1**).*

[D]Only the assembly was publicly available to use for se gene detection in this study. Therefore, only BLAST+ could be used.

with differences in the allele scoring methods of BLAST+ and SRST2. Manual alignment of the contigs against the Ψ*ent1-2* reference (Genbank reference MN450303.1), confirmed the presence of both pseudogenes with 100.0% sequence identity and query coverage in both isolates for the different extraction kits (data not shown).

Performance Evaluation of Whole Genome Sequencing-Based Isolate Relatedness Determination

The performance of WGS in isolate relatedness determination was evaluated through comparison of inferred phylogenetic relationships by cgMLST-typing with the *a priori* known relationships of the in-house sequenced isolates. Of all 2208 core gene loci, on average, 91.2 ± 5.7% could be detected with 100.0% query coverage and sequence identity across all isolates. **Figure 1** visualizes the relatedness between the *S. aureus* isolates, based on cgMLST. The tree demonstrates that the outbreak strain isolated from the food matrix (TIAC1798) clustered together with those of the human cases (TIAC1847 and TIAC1848) in a single clade carried by one branch, being separated from all other non-outbreak isolates. Similarly, both strains isolated from an identical human case at different locations (i.e., TIAC1993 and TIAC1994) also clustered together on one branch. No cgMLST allele differences were identified between any of the outbreak isolates within the outbreak clade or between the two isolates from the same human case. WGS-based cgMLST analyses, therefore, provided inferred phylogenetic relationships between the *S. aureus* isolates that were in accordance with the *a priori* known relationships.

Moreover, the used DNA extraction kit and workflow had no influence on the obtained cgMLST profiles or on the retrieved relationships between isolates. Indeed, no cgMLST allele differences were obtained between samples from the same isolate and, thus, clustered together per isolate in one single clade carried by a single branch, irrespective of the applied DNA extraction kit.

DISCUSSION

The goal of this study was to assess the potential benefits of WGS compared with conventional molecular methods currently used in the investigation of SFP outbreaks. For this purpose, WGS was employed for *se* gene detection applying a custom database, and its profiles were compared with those previously obtained by routine methods (the EURL-CPS multiplex PCR and/or SET-RPLA/VIDAS analyses) and extended by *in silico* PCR. To cover the complete arsenal of currently described *se* genes, we included publicly available WGS data (Fursova et al., 2020; Merda et al., 2020) that were also verified with *in silico* PCR. Moreover, the performance of WGS-based isolate relatedness determination was evaluated using cgMLST on a data set with *a priori* known phylogenetic relationships. Within the scope of SFP investigation, different DNA extraction kits were tested for their influence on the comparability of WGS data.

This study demonstrates that WGS presents a valid alternative to molecular methods, serving as the ultimate multiplexing approach for *se* gene detection in *S. aureus* isolates. WGS enabled complete *se* gene profiling within one single assay. In contrast to WGS, PCR-based methods for the analysis of *se* gene profiles tend to miss the detection of genes that were not targeted by the approach, or when using primer pairs not targeting sufficiently conserved gene regions (see **Supplementary Table 1**). As a result, SFP outbreaks caused by non-classical enterotoxins that cannot be detected with existing commercial kits (i.e., SEs other than SEA to SEE) risk being left unresolved when also not targeted with PCR. However, complete PCR-based *se* gene profiling is very time-consuming and labor-intensive. Besides giving a first indication on the possible causality of a strain toward SFP outbreaks, knowledge on the complete *se* gene profile of isolates is also important in SFP surveillance. Its combination with clinical data of human cases can help in risk assessment of *S. aureus* strains to predict the potential pathogenicity of isolates. By analyzing isolated strains of SFP outbreaks for their complete *se* gene profile, more knowledge can moreover be acquired on *se* genes most commonly involved in SFP outbreaks.

Both SRST2 and BLAST+ showed highly similar WGS-based *se* gene detection outputs. Across all isolates, there was only one *se* gene left undetected because of assembly fragmentation when using BLAST+. The application of SRST2 read mapping is, hence, preferred for WGS-based *se* gene profiling. Although not encountered in this study, assembly issues might not only result in BLAST+ being more prone to the missed detection of *se* genes, but also the false positive detection of (pseudo)genes (Mason et al., 2018) that have recombined to new *se* genes. Indeed, *selv* and *selu* genes have been described to be formed from recombination events in *sem* and *sei* genes, or Ψ*ent1* and Ψ*ent2* pseudogenes, respectively (Thomas et al., 2006). More extensive performance evaluation of applying BLAST+ and SRST2 for *se* gene detection, which was, however, not the main goal of this study, would require a higher number of *S. aureus* isolates to be analyzed. Although previous research shows performance differences to be limited for gene detection in other bacterial species when using high-coverage data sets (Bogaerts et al., 2021), it is an interesting future research topic to also verify for *S. aureus* enterotoxin gene detection.

Whole genome sequencing also allows subtyping of *se* variants. Recently, an increasing number of SE variants are being described (Blaiotta et al., 2004; Fernández et al., 2006; Kohler et al., 2012; Johler et al., 2016; Aziz et al., 2020; Etter et al., 2020; Merda et al., 2020). Multiple variants (such as for SEC) are already shown to exhibit different structural and superantigenic features (Etter et al., 2020), but the influence on their emetic activity has not yet been investigated. However, in the future, based on this knowledge, subtyping of variants can become increasingly important in predicting the emetic potential of *S. aureus* strains. In this study, subtyping of *sec* and *selu* was done as proof of concept, but the applied *se* gene detection thresholds (i.e., >60.0% query coverage and >90.0% sequence identity for BLAST+ and >60.0% query coverage and <10.0% sequence divergence for SRST2) allow also detecting other *se* variants, including novel ones, without subtyping. However, WGS offers

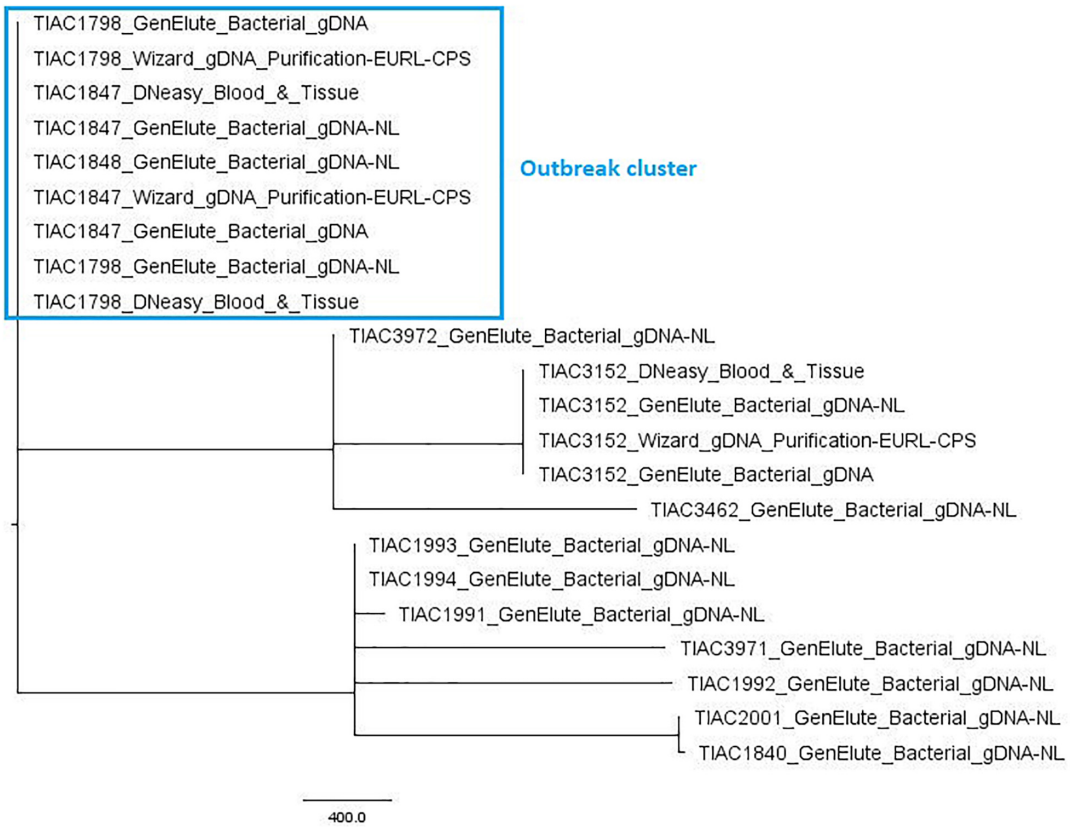


FIGURE 1 | Minimum spanning tree based on cgMLST for *S. aureus* isolates. A minimum spanning tree was created with GrapeTree using the MSTreeV2 method on all in-house sequenced outbreak and non-outbreak samples, i.e., all isolates, some of which processed with different DNA extraction kits. The sample names in the figure consist of the name of the respective isolate and an abbreviation of the applied DNA extraction kit, i.e., GenElute_Bacterial_gDNA-NL: GenElute Bacterial gDNA kit using the protocol for Gram-positive bacteria [without (i.e., no) lysostaphin, NL]; GenElute_Bacterial_gDNA: GenElute Bacterial gDNA kit using the protocol for Staphylococcal species (with lysostaphin); DNeasy_Blood_&_Tissue: DNeasy Blood & Tissue kit; Wizard_gDNA_Purification: Wizard gDNA Purification kit. The outbreak samples (outlined in the blue box) from food origin (TIAC1798) consistently cluster together with those of human cases (TIAC1847 and TIAC1848) while non-outbreak samples (TIAC1840, TIAC1991, TIAC1992, TIAC1993, TIAC1994, TIAC2001, TIAC3152, TIAC3462, TIAC3971, and TIAC3972) are separated from the outbreak cluster and delineated per isolate (except for TIAC1993 and TIAC1994 sampled from the same person that also cluster together because of their identical cgMLST profiles). The scale bar represents the number of cgMLST allele differences between samples. No cgMLST allele differences were identified between the outbreak isolates, nor isolates processed with different DNA extraction kits.

the potential to expand *se* subtyping to all variants when annotated as such in the applied gene detection database. This is an additional added value of WGS compared with conventional PCR-based methods, in which extra Sanger sequencing analyses of the obtained amplicons are often required for further subtype identification (Collery and Smyth, 2007).

For WGS to provide full *se* gene profiling, the application of a database containing reference sequences for all *se* genes and, if subtyping is of interest, also all variants is indispensable. However, publicly available reference databases for virulence gene detection in *S. aureus* isolates (Joensen et al., 2014; Liu et al., 2019; Sayers et al., 2019; Merda et al., 2020) do not enable complete *se* gene profiling (e.g., *ses, set*, *selw* to *selz*, *sel26* and *sel27* are missing in the VFDB_Full database, and *ses, set*, *selv* to *selz*, *sel26* and *sel27* in the VirulenceFinder DataBase) or subtyping of *se* variants (because they are not or ambiguously annotated in the databases). Studies using these databases as such (Huang et al., 2017; Petit and Read, 2018; Fursova et al., 2020; Karki et al., 2020; Merda et al., 2020; Schwendimann et al., 2021), thus risk underestimating the prevalence of *se* genes, potentially affecting obtained results. Indeed, more *se* genes were reported to be present in the isolates with publicly available WGS data after reanalysis with our methods compared with their previously WGS-determined *se* gene profiles, mainly because of the lack of complete databases in the other studies (Fursova et al., 2020; Merda et al., 2020). Recently, the EURL-CPS developed a workflow for *S. aureus* analysis in the scope of SFP investigation, i.e., NAuRA, using a database that allows almost complete (*selw* is missing and no variant subtyping) *se* gene detection (Merda et al., 2020). NAuRA is available on github, and the Uniprot accession numbers of SE protein sequences used as reference in their genomic analyses are available in the supplementary data of the corresponding publication (Merda et al., 2020). NAuRA can be locally installed by an experienced bioinformatician, who might not always be available in each NRL. Nevertheless, the implementation of a publicly available

database that is easily accessible and queryable is favorable in a routine setting because it supports the comparability of WGS data results, especially important during SFP investigation. To make WGS more approachable, it is, thus, important that reference databases enabling full *se* gene profiling are made publicly available through online open repositories. Moreover, these databases should be continuously updated with the state-of-the-art knowledge so that full *se* gene profiling and variant subtyping is possible. A FASTA file containing the *se* gene sequences and their corresponding accession numbers used for the database constructed in this manuscript was added to the **Supplementary Data**. This database was also integrated in the open access online interface[5] (registration required), using the parameters as described in this manuscript, to moreover allow user-friendly *se* gene detection.

Although present in the extended VFDB_Full database, in our study, discrepant observations were identified with the detection of Ψ*ent* pseudogenes when using WGS. This was related to (i) database sequence clustering, which limits gene detection to one per cluster, and (ii) differences in the allele scoring methods of SRST2 and BLAST+ that decide upon the detected gene (allele) per cluster. Because both Ψ*ent* pseudogenes coexist on the enterotoxin gene cluster (*egc*) (Jarraud et al., 2001), one of them risks being left undetected when using BLAST+ or SRST2. However, because pseudogenes are non-functional (Letertre et al., 2003b; Liang et al., 2016), the impact of these discrepancies can be regarded as negligible. Nevertheless, if their detection would be of special interest, we suggest performing manual alignment of the assembly with a Ψ*ent1-2* reference sequence to identify their potential joint presence as performed in this study. This extra alignment step still remains much less time-consuming compared with the Sanger sequencing analyses that would be required when using conventional PCR methods to distinguish pseudogene presence from those of the *selu* gene (Collery and Smyth, 2007). Sequence clustering of gene variants is, however, important when applying large databases, such as the VFDB_Full, to aid in identifying the most appropriate gene (allele) within the cluster. Without clustering of gene variants in combination with default gene detection thresholds, a hit to all of those variants might be picked up if the gene is present, whereas manual inspection would reveal that it concerns only one gene (variant) that is present. Based on sequence similarity between all *se* genes and variants (Merda et al., 2020), we decided to apply a sequence identity threshold of 85.0% to cluster the extended VFDB_Full database, combined with default gene detection thresholds (i.e., >60.0% query coverage and >90.0% sequence identity for BLAST+ and >60.0% query coverage and <10.0% sequence divergence for SRST2). Most *se* genes and variants are consequently clustered together per gene except for *selu* and the pseudogenes or *sem* and *selv* that are present in the same cluster together. With the combination of both clustering and gene detection thresholds, a sufficient level of sensitivity and specificity is acquired to detect novel *se* variants and avoid FP and FN gene detection, respectively. The importance of appropriate gene detection (and database sequence clustering) thresholds is also shown by analyzing publicly available WGS data. Based on *in silico* PCR and manual inspection of the contigs, the *selv* gene was found to have been previously falsely detected within these public WGS data [SAMN13134218; (Fursova et al., 2020)]. This could likely be explained by the application of default SRST2 gene detection thresholds (i.e., >90.0% query coverage and <10.0% sequence divergence), which are not stringent enough to distinguish between the present *sem* and *sei*, and the absent *selv*, in the absence of a database clustering step. Indeed, with these detection criteria, *sem*, *sei*, and *selv* were detected in the previous study (Fursova et al., 2020) although only *sem* and *sei* were determined to be present here. Although the gene detection criteria in our study were looser, the presence and absence of these genes could be correctly identified because they were combined with a database whose sequences were clustered. Of all 27 enterotoxins described to date, *selv* was, thus, determined not to be present in any of the selected strains/public data. Expanding the number of isolates in this study would likely also increase the potential of covering *selv*. However, because this gene is not targeted by the VIDAS/SET-RPLA immunoassays or EURL-CPS multiplex PCR applied routinely in-house and is found to be only scarcely prevalent (Merda et al., 2020), including an isolate containing *selv* would not be straightforward. Moreover, because *sem* and *selv* are not yet found to co-occur within the same strain to date (Thomas et al., 2006; Argudín et al., 2010), detection issues related to their clustering are not expected if an isolate containing *selv* would have been added to the study (as was attempted with the public WGS data). Nor are *selu* and pseudogene detection issues expected related to their clustering in the database. When adding more variant sequences to the database, as elaborated above to allow subtyping, it is necessary that applied gene detection thresholds and database sequence identity thresholds for clustering are further reviewed to ensure accurate and specific *se* variant detection.

To examine the completeness of WGS-based *se* gene profiling in *S. aureus* isolates, previously assessed *se* gene profiles were extended with *in silico* PCR. Using *in silico* PCR to validate the detection of *se* genes in all in-house sequenced isolates and isolates for which WGS data were publicly available might seem conflicting because both approaches are dependent on the same WGS data. However, although BLAST+ and SRST2 approaches detect complete *se* gene coding sequences, *in silico* PCR only evaluates primer binding sites and obtained amplicon sizes. Sixteen out of 155 primer pairs (i.e., 10.3%) described in literature did not result in a specific *in silico* product. Moreover, from the 139 primer sets used for the *in silico* PCR, another seven pairs (i.e., 5.0%) were identified to anneal in non-conserved gene regions leading to the missed detection of its corresponding target gene in some isolates tested within this study (see **Supplementary Data**). Increasing the number of *S. aureus* isolates, likely also supplements the gene sequence variability and potentially leads to the identification of extra primers that are not developed in sufficiently conserved regions. Although this remains an interesting research question for further research, with the limited number of isolates present, we highlight the added value of incorporating genomic data for more in-depth primer optimization (Vanneste et al., 2018) and show the importance of

[5]https://galaxy.sciensano.be/tool_runner?tool_id=pipeline_staphylococcus

profound sensitivity and specificity studies for developed primer sets before publication. Indeed, if these primer pairs were used in a conventional PCR, a FN result might have been obtained, impeding a full correct characterization of the *se* profile. By applying the *in silico* PCR approach with 139 PCR primer pairs (i.e., 89.7%) selected based on their specificity, something that would be less feasible in a conventional approach, this risk was circumvented and even showed an added value. For the 11 conventionally PCR tested *se* genes, all results were completely congruent with the *in silico* PCR results, thereby validating our approach. Moreover, the *in silico* PCR indicated an inconsistency with the reported *se* profiles in a previous study, which was further confirmed using manual alignments and explained by the applied computational methods. Our review of published primer pairs and *in silico* investigation for their specificity can, therefore, also be valuable information for laboratories not having the resources to perform WGS.

Besides the fact that complete *se* gene profiling can provide an initial insight into the potential causality of strains toward outbreaks, WGS also allows inferring more refined phylogenetic relationships between *S. aureus* isolates based on cgMLST-typing. Because 91.2 ± 5.7% of the 2208 core genome loci could be typed, a relatively high number of genomic markers was available to reliably resolve isolate relationships in this study. cgMLST-typing has gained interest with regard to standardization and harmonization because of the transferability of the applied scheme, enabling the use of the same database (Sabat et al., 2017), which allows SFP outbreaks to be investigated across laboratories. Although other cgMLST schemes have been developed by multiple instances (Leopold et al., 2014), we adopted the PubMLST scheme to infer *S. aureus* isolate relationships in this study because it is the most frequently used scheme, that is, moreover, also applied by the EURL-CPS (Merda et al., 2020). The ultimate level of discriminative power is provided by SNP analysis. However, this can only be used to further fine-tune relationships between closely related strains. Because the overall distance based on cgMLST loci differences between the isolates in this study is large, except for the outbreak strains and two identical strains TIAC1993 and TIAC1994 isolated from one human case, picking a suitable reference genome for SNP analysis would be impossible. Increasing the number of outbreak and non-outbreak isolates can be used to illustrate the discriminative power of WGS even more than in this study. However, the rapid course of disease and the complexity of *S. aureus* isolation because of potentially affected strain viability, often complicate the collection of strains during outbreak investigation. Moreover, not many studies were found to analyze *S. aureus* isolates with WGS in the scope of SFP outbreak investigation. Our study, by demonstrating its added value, might contribute in stimulating the use of WGS in SFP outbreak investigations so that its benefits can be fully exploited.

Data sharing between laboratories is crucial in the investigation of SFP outbreaks. However, the comparability of WGS data can potentially be affected by impaired plasmid extraction performances of commercial DNA preparation kits (Becker et al., 2016; Nouws et al., 2020b), especially when harboring important virulence genes, such as *se* genes. Therefore, we tested a frequently used kit (DNeasy Blood & Tissue kit), the kit recommended by the EURL-CPS for DNA preparation of *S. aureus* isolates (Wizard gDNA Purification kit) and the kit used at the Belgian NRL-FBO/NRL-CPS (GenElute Bacterial gDNA kit) for their influence on WGS data analyses of *S. aureus*. Although the DNeasy Blood & Tissue and Wizard gDNA Purification kits were previously described to have impaired plasmid extraction performances in Gram-negative species, possibly leading to missed WGS-based detection of plasmid-encoded genes (Becker et al., 2016; Pasquali et al., 2019; Nouws et al., 2020b), this assumption could not be extrapolated to *S. aureus* isolates based on the results in our study. Indeed, identical profiles of detected *se* genes, whether or not plasmid-encoded, were obtained between the samples per isolate, irrespective of the applied DNA extraction kit. Moreover, the use of different kits had no influence on inferred cgMLST-based isolate relationships. Kit choice to prepare DNA of *S. aureus* for WGS can, thus, be based on other factors. Because of its earlier communication to be appropriate for WGS data analysis of Gram-negative and -positive bacteria (Nouws et al., 2020b), the GenElute Bacterial gDNA kit can be of interest for usage in laboratories, such as NRLs, investigating multiple foodborne pathogens. We moreover examined the necessity of employing the expensive lysostaphin enzyme within its protocol as specified for Staphylococcal species (Zhao et al., 2012). Because sufficient amounts of DNA were obtained without using lysostaphin, and its use did not influence the outcome of WGS analyses, the GenElute Bacterial gDNA protocol for Gram-positive bacteria not using lysostaphin is especially beneficial for routine application because discarding lysostaphin nearly halves the cost per sample (**Table 2**).

Although a limited number of isolates was used in this study, it is clear that WGS has benefits in the investigation of SFP outbreaks, yielding information on the complete *se* profile and relatedness between strains, all within one single test. Although WGS can provide, besides *se* gene detection, complete isolate characterization by detecting other virulence and antimicrobial resistance (AMR) genes, such as, among others, the virulence *tsst* gene (Kulhankova et al., 2014) responsible for, e.g., menstrual toxic shock syndrome, or the AMR *mecA* gene causing methicillin resistance (Peacock and Paterson, 2015), these characteristics were not addressed in this study because of their irrelevance in SFP or its treatment (Fisher et al., 2018), and because no conventional metadata concerning these characteristics were available with which to compare the performance of WGS-based detection. Performing *in silico* PCR to obtain information on the presence/absence of these genes would be virtually impossible because of the large numbers of AMR and virulence genes that exist. However, other studies have already demonstrated that WGS scores very well in predicting the AMR phenotype within *S. aureus* (Gordon et al., 2014; Cunningham et al., 2020). For WGS to yield all benefits in SFP investigation, it is important that *S. aureus* strain isolation is successful. Because the viability of *S. aureus* can be affected by food processing through, e.g., heating, while integrally preserving the emetic properties of the produced enterotoxins (Fisher et al., 2018), a proportion of the suspected foods might be left unable to be investigated with the

proposed method. Metagenomics sequencing of the complete sample without strain isolation might offer a solution relevant for further research (Boers et al., 2019). However, both WGS and metagenomics sequencing only deliver information at the genotypic level. Ideally, sequencing-based methods should be used in first line to screen for *se* gene presence and can then be combined with a method that allows detection of all produced SEs at the protein level. Nevertheless, through complete *se* gene detection, WGS enables broadening the insight in *S. aureus* and SFP and examines the potential of isolates to produce SEs and to play a role in SFP. Moreover, thanks to its ultimate discriminatory power, WGS can simultaneously more accurately pinpoint strains as the cause for a SFP outbreak and, thus, accelerate its management. Conclusively, this study shows the added value of using WGS in SFP outbreak investigation and encourages its use in a routine setting.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI SRA repository under the accession number PRJNA750393 (http://www.ncbi.nlm.nih.gov/bioproject/750393). Corresponding accession numbers are listed in **Supplementary Table 7**.

AUTHOR CONTRIBUTIONS

SN, NR, and SDK conceived and designed the study. SDK supervised the project. SN performed the wet lab experiments. BB and KV provided the bioinformatic tools in Galaxy. SN and BB performed the bioinformatic analysis. SD and BV were responsible for the former generation of all results obtained with the conventional methods. SN and SDK participated in the interpretation of the results and wrote the manuscript. KV, BB, LL, NR, and KM provided specialized feedback on the obtained results. All co-authors commented and approved the submitted version.

FUNDING

This research that yielded these results was funded by Sciensano (contract RP.Be READY).

ACKNOWLEDGMENTS

We thank the technicians of the service transversal activities in Applied Genomics at Sciensano, Belgium for conducting all Next Generation Sequencing runs. We also express our gratitude to the NRL-FBO/NRL-CPS and the EURL-CPS for sharing *S. aureus* isolates and their metadata concerning the SE/*se* profiles (determined with VIDAS/SET-RPLA and the multiplex PCR) and mutual relationships between isolates (based on PFGE) using conventional methods.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.750278/full#supplementary-material

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Anderson, D. J., Harris, S. R., Godofsky, E., Toriscelli, T., Rude, T. H., Elder, K., et al. (2014). Whole Genome Sequencing of a Methicillin-Resistant *Staphylococcus aureus* Pseudo-Outbreak in a Professional Football Team. *Open Forum Infect. Dis.* 1:2633851. doi: 10.1093/ofid/ofu096

Argudín, M. A., Mendoza, M. C., and Rodicio, M. R. (2010). Food Poisoning and *Staphylococcus aureus* Enterotoxins. *Toxins* 2, 1751–1773. doi: 10.3390/toxins2071751

Aung, M. S., Urushibara, N., Kawaguchiya, M., Ito, M., Habadera, S., and Kobayashi, N. (2020). Prevalence and genetic diversity of staphylococcal enterotoxin (-Like) genes *sey*, *selw*, *selx*, *selz*, *sel26* and *sel27* in community-acquired methicillin-resistant *Staphylococcus aureus*. *Toxins* 12, 1–13. doi: 10.3390/toxins12050347

Aziz, F., Hisatsune, J., Yu, L., Kajimura, J., Sato'o, Y., Ono, H. K., et al. (2020). *Staphylococcus aureus* Isolated from Skin from Atopic-Dermatitis Patients Produces Staphylococcal Enterotoxin Y, Which Predominantly Induces T-Cell Receptor Vα-Specific Expansion of T Cells. *Infect. Immun.* 88:19. doi: 10.1128/IAI.00360-19

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bayles, K. W., and Iandolo, J. J. (1989). Genetic and molecular analyses of the gene encoding staphylococcal enterotoxin D. *J. Bacteriol.* 171, 4799–4806. doi: 10.1128/jb.171.9.4799-4806.1989

Becker, K., Roth, R., and Peters, G. (1998). Rapid and Specific Detection of Toxigenic *Staphylococcus aureus*: Use of Two Multiplex PCR Enzyme Immunoassays for Amplification and Hybridization of Staphylococcal Enterotoxin Genes, Exfoliative Toxin Genes, and Toxic Shock Syndrome Toxin 1 Gene. *J. Clin. Microbiol.* 36, 2548–2553. doi: 10.1128/JCM.36.9.2548-2553.1998

Becker, L., Steglich, M., Fuchs, S., Werner, G., and Nübel, U. (2016). Comparison of six commercial kits to extract bacterial chromosome and plasmid DNA for MiSeq sequencing. *Sci. Rep.* 6:28063. doi: 10.1038/srep28063

Bergdoll, M. S., Borja, C. R., Robbins, R. N., and Weiss, K. F. (1971). Identification of enterotoxin E. *Infect. Immun.* 4, 593–595. doi: 10.1128/iai.4.5.593-595.1971

Blaiotta, G., Ercolini, D., Pennacchia, C., Fusco, V., Casaburi, A., Pepe, O., et al. (2004). PCR detection of staphylococcal enterotoxin genes in *Staphylococcus* spp. strains isolated from meat and dairy products. Evidence for new variants of *seG* and *seI* in *S. aureus* AB-8802. *J. Appl. Microbiol.* 97, 719–730. doi: 10.1111/j.1365-2672.2004.02349.x

Blair, J. E., and Williams, R. E. (1961). Phage typing of staphylococci. *Bull. World Health Organ.* 24, 771–784.

Boers, S. A., Jansen, R., and Hays, J. P. (2019). Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the

routine clinical microbiological diagnostic laboratory. *Eur. J. Clin. Microbiol. Infect. Dis.* 38, 1059–1070. doi: 10.1007/s10096-019-03520-3

Bogaerts, B., Nouws, S., Verhaegen, B., Denayer, S., Van Braekel, J., Winand, R., et al. (2021). Validation strategy of a bioinformatics whole genome sequencing workflow for Shiga toxin-producing *Escherichia coli* using a reference collection extensively characterized with conventional methods. *Microb. Genomics* 7:531. doi: 10.1099/mgen.0.000531

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics* 10, 1–9. doi: 10.1186/1471-2105-10-421

Collery, M. M., and Smyth, C. J. (2007). Rapid differentiation of Staphylococcus aureus isolates harbouring *egc* loci with pseudogenes ψ*ent1* and ψ*ent2* and the *selu* or *selu v* gene using PCR-RFLP. *J. Med. Microbiol.* 56, 208–216. doi: 10.1099/jmm.0.46948-0

Cookson, B. D., Aparicio, P., Deplano, A., Struelens, M., Goering, R., and Marples, R. (1996). Inter-centre comparison of pulsed-field gel electrophoresis for the typing of methicillin-resistant *Staphylococcus aureus*. *J. Med. Microbiol.* 44, 179–184. doi: 10.1099/00222615-44-3-179

Cremers, A. J. H., Coolen, J. P. M., Bleeker-Rovers, C. P., van der Geest-Blankert, A. D. J., Haverkate, D., Hendriks, H., et al. (2020). Surveillance-embedded genomic outbreak resolution of methicillin-susceptible *Staphylococcus aureus* in a neonatal intensive care unit. *Sci. Rep.* 10:2619. doi: 10.1038/s41598-020-59015-1

Cunningham, S. A., Chia, N., Jeraldo, P. R., Quest, D. J., Johnson, J. A., Boxrud, D. J., et al. (2017). Comparison of Whole-Genome Sequencing Methods for Analysis of Three Methicillin-Resistant *Staphylococcus aureus* Outbreaks. *J. Clin. Microbiol.* 55, 1946–1953. doi: 10.1128/JCM.00029-17

Cunningham, S. A., Jeraldo, P. R., Schuetz, A. N., Heitman, A. A., and Patel, R. (2020). *Staphylococcus aureus* whole genome sequence–based susceptibility and resistance prediction using a clinically amenable workflow. *Diagn. Microbiol. Infect. Dis.* 97:115060. doi: 10.1016/j.diagmicrobio.2020.115060

Denayer, S., Delbrassinne, L., Nia, Y., and Botteldoorn, N. (2017). Food-Borne Outbreak Investigation and Molecular Typing: High Diversity of *Staphylococcus aureus* Strains and Importance of Toxin Detection. *Toxins* 9:407. doi: 10.3390/toxins9120407

ECDC (2019). *Proficiency test for Listeria monocytogenes* whole genome assembly – 2018. Europe: European Centre for Disease Prevention and Control. Available online at: https://www.ecdc.europa.eu/en/publications-data/proficiency-test-listeria-monocytogenes-whole-genome-assembly-2018

Ekkelenkamp, M. B., Vervoorn, M. T., Bayjanov, J. R., Fluit, A. C., Benaissa-Trouw, B. J., and Ramjankhan, F. Z. (2018). Therapy and Outcome of *Staphylococcus aureus* Infections of Intracorporeal Ventricular Assist Devices. *Artif. Organs* 42, 983–991. doi: 10.1111/aor.13159

Etter, D., Schelin, J., Schuppler, M., and Johler, S. (2020). Staphylococcal Enterotoxin C—An Update on SEC Variants, Their Structure and Properties, and Their Role in Foodborne Intoxications. *Toxins* 12:584. doi: 10.3390/toxins12090584

European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC) (2021). The European Union One Health 2019 Zoonoses Report. *EFSA J.* 19:6406. doi: 10.2903/j.efsa.2021.6406

Fernández, M. M., De Marzi, M. C., Berguer, P., Burzyn, D., Langley, R. J., Piazzon, I., et al. (2006). Binding of natural variants of staphylococcal superantigens SEG and SEI to TCR and MHC class II molecule. *Mol. Immunol.* 43, 927–938. doi: 10.1016/j.molimm.2005.06.029

Fisher, E. L., Otto, M., and Cheung, G. Y. C. (2018). Basis of Virulence in Enterotoxin-Mediated Staphylococcal Food Poisoning. *Front. Microbiol.* 9, 1–18. doi: 10.3389/fmicb.2018.00436

Fursova, K., Sorokin, A., Sokolov, S., Dzhelyadin, T., Shulcheva, I., Shchannikova, M., et al. (2020). Virulence Factors and Phylogeny of *Staphylococcus aureus* Associated With Bovine Mastitis in Russia Based on Genome Sequences. *Front. Vet. Sci.* 7, 1–10. doi: 10.3389/fvets.2020.00135

Gordon, N. C., Price, J. R., Cole, K., Everitt, R., Morgan, M., Finney, J., et al. (2014). Prediction of *Staphylococcus aureus* Antimicrobial Resistance by Whole-Genome Sequencing. *J. Clin. Microbiol.* 52, 1182–1191. doi: 10.1128/JCM.03117-13

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hennekinne, J.-A., Ostyn, A., Guillier, F., Herbin, S., Prufer, A.-L., and Dragacci, S. (2010). How Should Staphylococcal Food Poisoning Outbreaks Be Characterized? *Toxins* 2, 2106–2116. doi: 10.3390/toxins2082106

Heymans, F., Fischer, A., Stow, N. W., Girard, M., Vourexakis, Z., Courtis, A., et al. (2010). Screening for Staphylococcal Superantigen Genes Shows No Correlation with the Presence or the Severity of Chronic Rhinosinusitis and Nasal Polyposis. *PLoS One* 5:e9525. doi: 10.1371/journal.pone.0009525

Huang, A. D., Luo, C., Pena-Gonzalez, A., Weigand, M. R., Tarr, C. L., and Konstantinidis, K. T. (2017). Metagenomics of Two Severe Foodborne Outbreaks Provides Diagnostic Signatures and Signs of Coinfection Not Attainable by Traditional Methods. *Appl. Environ. Microbiol.* 83, 1–14. doi: 10.1128/AEM.02577-16

Illumina (2018). *Nextera XT DNA library prep kit reference guide #15031942*. Available online at: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-reference-guide-15031942-05.pdf (accessed date May 2019).

Inouye, M., Dashnow, H., Raven, L.-A., Schultz, M. B., Pope, B. J., Tomita, T., et al. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* 6:90. doi: 10.1186/s13073-014-0090-6

Jarraud, S., Peyrat, M. A., Lim, A., Tristan, A., Bes, M., Mougel, C., et al. (2001). A Highly Prevalent Operon of Enterotoxin Gene, Forms a Putative Nursery of Superantigens in *Staphylococcus aureus*. *J. Immunol.* 166, 669–677. doi: 10.4049/jimmunol.166.1.669

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13

Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and Easy *In Silico* Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15

Johler, S., Giannini, P., Jermini, M., Hummerjohann, J., Baumgartner, A., and Stephan, R. (2015). Further Evidence for Staphylococcal Food Poisoning Outbreaks Caused by *egc*-Encoded Enterotoxins. *Toxins* 7, 997–1004. doi: 10.3390/toxins7030997

Johler, S., Sihto, H.-M., Macori, G., and Stephan, R. (2016). Sequence Variability in Staphylococcal Enterotoxin Genes seb, sec, and sed. *Toxins* 8:169. doi: 10.3390/toxins8060169

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1

Karki, A. B., Neyaz, L., and Fakhr, M. K. (2020). Comparative Genomics of Plasmid-Bearing *Staphylococcus aureus* Strains Isolated From Various Retail Meats. *Front. Microbiol.* 11:574923. doi: 10.3389/fmicb.2020.574923

Kohler, P. L., Greenwood, S. D., Nookala, S., Kotb, M., Kranz, D. M., and Schlievert, P. M. (2012). *Staphylococcus aureus* Isolates Encode Variant Staphylococcal Enterotoxin B Proteins That Are Diverse in Superantigenicity and Lethality. *PLoS One* 7:e41157. doi: 10.1371/journal.pone.0041157

Kulhankova, K., King, J., and Salgado-Pabón, W. (2014). Staphylococcal toxic shock syndrome: superantigen-mediated enhancement of endotoxin shock and adaptive immune suppression. *Immunol. Res.* 59, 182–187. doi: 10.1007/s12026-014-8538-8

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/JCM.06094-11

Lee, G. C., Long, S. W., Musser, J. M., Beres, S. B., Olsen, R. J., Dallas, S. D., et al. (2015). Comparative Whole Genome Sequencing of Community-Associated Methicillin-Resistant *Staphylococcus aureus* Sequence Type 8 from Primary Care Clinics in a Texas Community. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* 35, 220–228. doi: 10.1002/phar.1536

Leopold, S. R., Goering, R. V., Witten, A., Harmsen, D., and Mellmann, A. (2014). Bacterial Whole-Genome Sequencing Revisited: Portable, Scalable, and Standardized Analysis for Typing and Detection of Virulence and Antibiotic Resistance Genes. *J. Clin. Microbiol.* 52, 2365–2370. doi: 10.1128/JCM.00262-14

Letertre, C., Perelle, S., Dilasser, F., and Fach, P. (2003b). Identification of a new putative enterotoxin SEU encoded by the *egc* cluster of *Staphylococcus aureus*. *J. Appl. Microbiol.* 95, 38–43. doi: 10.1046/j.1365-2672.2003.01957.x

Letertre, C., Perelle, S., Dilasser, F., and Fach, P. (2003a). A strategy based on 5′ nuclease multiplex PCR to detect enterotoxin genes *sea* to *sej* of *Staphylococcus aureus*. *Mol. Cell. Probes* 17, 227–235. doi: 10.1016/S0890-8508(03)00058-6

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Liang, C., Schaack, D., Srivastava, M., Gupta, S., Sarukhanyan, E., Giese, A., et al. (2016). A *Staphylococcus aureus* Proteome Overview: Shared and Specific Proteins and Protein Complexes from Representative Strains of All Three Clades. *Proteomes* 4:8. doi: 10.3390/proteomes4010008

Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692. doi: 10.1093/nar/gky1080

Lopes, H. R., Noleto, A. L. S., de Las Heras, M. D., and Bergdoll, M. S. (1993). Selective Enterotoxin Production in Foods by *Staphylococcus aureus* Strains That Produce More Than One Enterotoxin. *J. Food Prot.* 56, 538–540. doi: 10.4315/0362-028x-56.6.538

Mason, A., Foster, D., Bradley, P., Golubchik, T., Doumith, M., Gordon, N. C., et al. (2018). Accuracy of Different Bioinformatics Methods in Detecting Antibiotic Resistance and Virulence Factors from *Staphylococcus aureus* Whole-Genome Sequences. *J. Clin. Microbiol.* 56, 1–12. doi: 10.1128/JCM.01815-17

Mehrotra, M., Wang, G., and Johnson, W. M. (2000). Multiplex PCR for Detection of Genes for *Staphylococcus aureus* Enterotoxins, Exfoliative Toxins, Toxic Shock Syndrome Toxin 1, and Methicillin Resistance. *J. Clin. Microbiol.* 38, 1032–1035. doi: 10.1128/JCM.38.3.1032-1035.2000

Merda, D., Felten, A., Vingadassalon, N., Denayer, S., Titouche, Y., Decastelli, L., et al. (2020). NAuRA: Genomic Tool to Identify Staphylococcal Enterotoxins in *Staphylococcus aureus* Strains Responsible for FoodBorne Outbreaks. *Front. Microbiol.* 11, 1–12. doi: 10.3389/fmicb.2020.01483

Monday, S. R., and Bohach, G. A. (1999). Use of Multiplex PCR To Detect Classical and Newly Described Pyrogenic Toxin Genes in Staphylococcal Isolates. *J. Clin. Microbiol.* 37, 3411–3414. doi: 10.1128/JCM.37.10.3411-3414.1999

Moore, G., Cookson, B., Gordon, N. C., Jackson, R., Kearns, A., Singleton, J., et al. (2015). Whole-genome sequencing in hierarchy with pulsed-field gel electrophoresis: the utility of this approach to establish possible sources of MRSA cross-transmission. *J. Hosp. Infect.* 90, 38–45. doi: 10.1016/j.jhin.2014.12.014

Murchan, S., Kaufmann, M. E., Deplano, A., de Ryck, R., Struelens, M., Zinn, C. E., et al. (2003). Harmonization of Pulsed-Field Gel Electrophoresis Protocols for Epidemiological Typing of Strains of Methicillin-Resistant *Staphylococcus aureus*: a Single Approach Developed by Consensus in 10 European Laboratories and Its Application for Tracing the Spr. *J. Clin. Microbiol.* 41, 1574–1585. doi: 10.1128/JCM.41.4.1574-1585.2003

Nagaraj, S., Ramlal, S., Sripathy, M. H., and Batra, H. V. (2014). Development and evaluation of a novel combinatorial selective enrichment and multiplex PCR technique for molecular detection of major virulence-associated genes of enterotoxigenic *Staphylococcus aureus* in food samples. *J. Appl. Microbiol.* 116, 435–446. doi: 10.1111/jam.12364

Nouws, S., Bogaerts, B., Verhaegen, B., Denayer, S., Crombé, Fl, De Rauw, K., et al. (2020a). The Benefits of Whole Genome Sequencing for Foodborne Outbreak Investigation from the Perspective of a National Reference Laboratory in a Smaller Country. *Foods* 9:1030. doi: 10.3390/foods9081030

Nouws, S., Bogaerts, B., Verhaegen, B., Denayer, S., Piérard, D., Marchal, K., et al. (2020b). Impact of DNA extraction on whole genome sequencing analysis for characterization and relatedness of Shiga toxin-producing *Escherichia coli* isolates. *Sci. Rep.* 10:14649. doi: 10.1038/s41598-020-71207-3

Omoe, K., Hu, D.-L., Takahashi-Omoe, H., Nakane, A., and Shinagawa, K. (2003). Identification and Characterization of a New Staphylococcal Enterotoxin-Related Putative Toxin Encoded by Two Kinds of Plasmids. *Infect. Immun.* 71, 6088–6094. doi: 10.1128/IAI.71.10.6088-6094.2003

Ono, H. K., Omoe, K., Imanishi, K., Iwakabe, Y., Hu, D. L., Kato, H., et al. (2008). Identification and characterization of two novel staphylococcal enterotoxins, types S and T. *Infect. Immun.* 76, 4999–5005. doi: 10.1128/IAI.00045-08

Pasquali, F., Do Valle, I., Palma, F., Remondini, D., Manfreda, G., Castellani, G., et al. (2019). Application of different DNA extraction procedures, library preparation protocols and sequencing platforms: impact on sequencing results. *Heliyon* 5:e02745. doi: 10.1016/j.heliyon.2019.e02745

Peacock, S. J., and Paterson, G. K. (2015). Mechanisms of Methicillin Resistance in *Staphylococcus aureus*. *Annu. Rev. Biochem.* 84, 577–601. doi: 10.1146/annurev-biochem-060614-034516

Petit, R. A., and Read, T. D. (2018). *Staphylococcus aureus* viewed from the perspective of 40,000+ genomes. *PeerJ* 6:e5261. doi: 10.7717/peerj.5261

Price, J. R., Didelot, X., Crook, D. W., Llewelyn, M. J., and Paul, J. (2013). Whole genome sequencing in the prevention and control of *Staphylococcus aureus* infection. *J. Hosp. Infect.* 83, 14–21. doi: 10.1016/j.jhin.2012.10.003

Rambaut, A. (2007). *FigTree*. Available online at: http://tree.bio.ed.ac.uk/software/figtree/ (accessed date 2018-November-25)

Roussel, S., Felix, B., Vingadassalon, N., Grout, J., Hennekinne, J.-A., Guillier, L., et al. (2015). *Staphylococcus aureus* strains associated with food poisoning outbreaks in France: comparison of different molecular typing methods, including MLVA. *Front. Microbiol.* 6, 1–12. doi: 10.3389/fmicb.2015.00882

Sabat, A. J., Hermelijn, S. M., Akkerboom, V., Juliana, A., Degener, J. E., Grundmann, H., et al. (2017). Complete-genome sequencing elucidates outbreak dynamics of CA-MRSA USA300 (ST8-spa t008) in an academic hospital of Paramaribo, Republic of Suriname. *Sci. Rep.* 7:41050. doi: 10.1038/srep41050

Sayers, S., Li, L., Ong, E., Deng, S., Fu, G., Lin, Y., et al. (2019). Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic Acids Res.* 47, D693–D700. doi: 10.1093/nar/gky999

Schmitz, F.-J., Steiert, M., Hofmann, B., Verhoef, J., Hadding, U., Heinz, H.-P., et al. (1998). Development of a multiplex-PCR for direct detection of the genes for enterotoxin B and C, and toxic shock syndrome toxin-1 in *Staphylococcus aureus* isolates. *J. Med. Microbiol.* 47, 335–340. doi: 10.1099/00222615-47-4-335

Schouls, L. M., Spalburg, E. C., van Luit, M., Huijsdens, X. W., Pluister, G. N., van Santen-Verheuvel, M. G., et al. (2009). Multiple-Locus Variable Number Tandem Repeat Analysis of *Staphylococcus Aureus*: Comparison with Pulsed-Field Gel Electrophoresis and *spa*-Typing. *PLoS One* 4:e5082. doi: 10.1371/journal.pone.0005082

Schwendimann, L., Merda, D., Berger, T., Denayer, S., Feraudet-Tarisse, C., Kläui, A. J., et al. (2021). Staphylococcal Enterotoxin Gene Cluster: Prediction of Enterotoxin (SEG and SEI) Production and of the Source of Food Poisoning on the Basis of v Saβ Typing. *Appl. Environ. Microbiol.* 87, 2662–2620. doi: 10.1128/AEM.02662-20

Sharma, N. K., Rees, C. E. D., and Dodd, C. E. R. (2000). Development of a Single-Reaction Multiplex PCR Toxin Typing Assay for *Staphylococcus aureus* Strains. *Appl. Environ. Microbiol.* 66, 1347–1353. doi: 10.1128/AEM.66.4.1347-1353.2000

Shylaja, R., Murali, H. S., Batra, H. V., and Bawa, A. S. (2010). A novel multiplex PCR system for the detection of Staphylococcal enterotoxin B, *tsst*, *nuc* and *fem* genes of *Staphylococcus aureus* in food system. *J. Food Saf.* 30, 443–454. doi: 10.1111/j.1745-4565.2010.00218.x

Spaulding, A. R., Salgado-Pabón, W., Kohler, P. L., Horswill, A. R., Leung, D. Y. M., and Schlievert, P. M. (2013). Staphylococcal and Streptococcal Superantigen Exotoxins. *Clin. Microbiol. Rev.* 26, 422–447. doi: 10.1128/CMR.00104-12

Thomas, D. Y., Jarraud, S., Lemercier, B., Cozon, G., Echasserieau, K., Etienne, J., et al. (2006). Staphylococcal Enterotoxin-Like Toxins U2 and V, Two New Staphylococcal Superantigens Arising from Recombination within the Enterotoxin Gene Cluster. *Infect. Immun.* 74, 4724–4734. doi: 10.1128/IAI.00132-06

Umeda, K., Nakamura, H., Yamamoto, K., Nishina, N., Yasufuku, K., Hirai, Y., et al. (2017). Molecular and epidemiological characterization of staphylococcal foodborne outbreak of *Staphylococcus aureus* harboring *seg*, *sei*, *sem*, *sen*, *seo*, and *selu* genes without production of classical enterotoxins. *Int. J. Food Microbiol.* 256, 30–35. doi: 10.1016/j.ijfoodmicro.2017.05.023

Vanneste, K., Garlant, L., Broeders, S., Van Gucht, S., and Roosens, N. H. (2018). Application of whole genome data for *in silico* evaluation of primers and probes

routinely employed for the detection of viral species by RT-qPCR using dengue virus as a case study. *BMC Bioinformatics* 19:312. doi: 10.1186/s12859-018-2313-0

Wildemauwe, C., Godard, C., Verschraegen, G., Claeys, G., Duyck, M.-C., De Beenhouwer, H., et al. (2004). Ten years phage-typing of Belgian clinical methicillin-resistant *Staphylococcus aureus* isolates (1992–2001). *J. Hosp. Infect.* 56, 16–21. doi: 10.1016/j.jhin.2003.09.014

Wilson, G. J., Tuffs, S. W., Wee, B. A., Seo, K. S., Park, N., Connelley, T., et al. (2018). Bovine *Staphylococcus aureus* Superantigens Stimulate the Entire T Cell Repertoire of Cattle. *Infect. Immun.* 86, 1–16. doi: 10.1128/IAI.00505-18

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134. doi: 10.1186/1471-2105-13-134

Zhang, D.-F., Yang, X.-Y., Zhang, J., Qin, X., Huang, X., Cui, Y., et al. (2018). Identification and characterization of two novel superantigens among *Staphylococcus aureus* complex. *Int. J. Med. Microbiol.* 308, 438–446. doi: 10.1016/j.ijmm.2018.03.002

Zhang, S., Iandolo, J. J., and Stewart, G. C. (1998). The enterotoxin D plasmid of Staphylococcus aureus encodes a second enterotoxin determinant (sej). *FEMS Microbiol. Lett.* 168, 227–233. doi: 10.1016/S0378-1097(98)00423-6

Zhao, J., Carmody, L. A., Kalikin, L. M., Li, J., Petrosino, J. F., Schloss, P. D., et al. (2012). Impact of Enhanced *Staphylococcus* DNA Extraction on Microbial Community Measures in Cystic Fibrosis Sputum. *PLoS One* 7:e33127. doi: 10.1371/journal.pone.0033127

Zhou, Z., Alikhan, N.-F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., et al. (2018). GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* 28, 1395–1404. doi: 10.1101/gr.232397.117

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fmicb.2021.738284

Towards Real-Time and Affordable Strain-Level Metagenomics-Based Foodborne Outbreak Investigations Using Oxford Nanopore Sequencing Technologies

*Florence E. Buytaers[1,2†], Assia Saltykova[1,2†], Sarah Denayer[3], Bavo Verhaegen[3], Kevin Vanneste[1], Nancy H. C. Roosens[1], Denis Piérard[4], Kathleen Marchal[2,5] and Sigrid C. J. De Keersmaecker[1]**

[1] Transversal Activities in Applied Genomics, Sciensano, Brussels, Belgium, [2] Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, [3] National Reference Laboratory for Shiga Toxin-Producing Escherichia coli (NRL STEC), Foodborne Pathogens, Sciensano, Brussels, Belgium, [4] National Reference Center for Shiga Toxin-Producing Escherichia coli (NRC STEC), Department of Microbiology and Infection Control, Universitair Ziekenhuis Brussel (UZ Brussel), Vrije Universiteit Brussel (VUB), Brussels, Belgium, [5] Department of Information Technology, IDlab, IMEC, Ghent University, Ghent, Belgium

OPEN ACCESS

Edited by:
Nicholas Bokulich,
ETH Zürich, Switzerland

Reviewed by:
Francisco Diez-Gonzalez,
University of Georgia, United States
Alfonso Benítez-Páez,
Principe Felipe Research Center (CIPF), Spain

***Correspondence:**
Sigrid C. J. De Keersmaecker
Sigrid.DeKeersmaecker@sciensano.be

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: 08 July 2021
Accepted: 13 October 2021
Published: 05 November 2021

Citation:
Buytaers FE, Saltykova A, Denayer S, Verhaegen B, Vanneste K, Roosens NHC, Piérard D, Marchal K and De Keersmaecker SCJ (2021) Towards Real-Time and Affordable Strain-Level Metagenomics-Based Foodborne Outbreak Investigations Using Oxford Nanopore Sequencing Technologies. Front. Microbiol. 12:738284. doi: 10.3389/fmicb.2021.738284

The current routine laboratory practices to investigate food samples in case of foodborne outbreaks still rely on attempts to isolate the pathogen in order to characterize it. We present in this study a proof of concept using Shiga toxin-producing *Escherichia coli* spiked food samples for a strain-level metagenomics foodborne outbreak investigation method using the MinION and Flongle flow cells from Oxford Nanopore Technologies, and we compared this to Illumina short-read-based metagenomics. After 12 h of MinION sequencing, strain-level characterization could be achieved, linking the food containing a pathogen to the related human isolate of the affected patient, by means of a single-nucleotide polymorphism (SNP)-based phylogeny. The inferred strain harbored the same virulence genes as the spiked isolate and could be serotyped. This was achieved by applying a bioinformatics method on the long reads using reference-based classification. The same result could be obtained after 24-h sequencing on the more recent lower output Flongle flow cell, on an extract treated with eukaryotic host DNA removal. Moreover, an alternative approach based on *in silico* DNA walking allowed to obtain rapid confirmation of the presence of a putative pathogen in the food sample. The DNA fragment harboring characteristic virulence genes could be matched to the *E. coli* genus after sequencing only 1 h with the MinION, 1 h with the Flongle if using a host DNA removal extraction, or 5 h with the Flongle with a classical DNA extraction. This paves the way towards the use of metagenomics as a rapid, simple, one-step method for foodborne pathogen detection and for fast outbreak investigation that can be implemented in routine laboratories on samples prepared with the current standard practices.

Keywords: metagenomics, nanopore, Flongle, strain-level, outbreak, food surveillance, SNP analysis, STEC

INTRODUCTION

Foodborne diseases represent a major burden worldwide (WHO, 2015). Foodborne pathogens can cause large outbreaks affecting multiple people sometimes in different regions. In case of an outbreak, the common practice of public health institutions is to investigate human cases and try to relate them to the contaminated food, in order to remove it from the food chain and prevent further contaminations. This process is called source attribution (EFSA, 2019a). This investigation consists of a microbiological and epidemiological part. In many countries, a surveillance system is also in place, screening the food chain in order to remove contaminated foodstuffs before they reach the consumer. In that case, microbial risk assessment and hazard identification are conducted, and the pathogen does not need to be linked to patient's data, but its characteristics could be added to a database in order to conduct retrospective studies and link related cases or serve as background to detect clusters and thus putative outbreaks (ECDC and EFSA, 2019).

In both circumstances (i.e., surveillance or the microbiological part of the outbreak investigation), conventional microbiology methods based on sequential culture steps have been the standard for many years to obtain information on the bacterial contaminant(s) present in food. However, this depends on a series of steps that should be conducted on the samples, therefore requiring larger quantities of the sample that is not always easy to obtain, and most importantly, it requires obtaining an isolate, which is often time-consuming and not always successful. The heterogeneous contamination of food products, the complexity of the matrix, and the difficulty to culture certain organisms might not allow to detect a pathogen at levels as low as the infectious dose reported for human (European Food Safety Authority (EFSA), and European Centre for Disease Prevention and Control (ECDC), 2018). When an isolate is obtained, it is characterized with several (real-time) polymerase chain reactions [(q)PCRs] to detect pathogenicity markers and/or multiple locus sequencing typing (MLST), pulsed-field gel electrophoresis (PFGE), multiple locus variable-number tandem repeat analysis (MLVA), or other typing methods to relate cases of an outbreak, depending on the pathogen. This workflow does not always offer optimal resolution to discriminate the pathogenic agents at a desired level (Nouws et al., 2020a) and requires sequential tests to be conducted in the laboratory (Nouws et al., 2020a), which adds to the total cost and turnaround time of the analysis.

As an alternative, whole-genome sequencing (WGS) offers the ultimate resolution to the single-nucleotide polymorphism (SNP) level of the bacterial genome, allowing the simultaneous detection of all genes present in the bacteria as well as relatedness inference with phylogenetics (Sandora et al., 2014; Bogaerts et al., 2019), and has been recommended by the European Food Safety Authority (EFSA) for use on a list of pathogens in European laboratories (EFSA, 2014). However, circumventing the need for isolation can accelerate the collection of results even more, as well as allow the resolution of cases for which no isolate could be obtained following the detection protocol. Strain-level shotgun metagenomics approaches offer the possibility to obtain the same resolution as WGS, without the need for isolation (Forbes et al., 2017). A recent publication of the EFSA highlighted the need for demonstrating the ability of metagenomics to be used as a new alternative for risk assessment, source attribution, and outbreak investigation (EFSA, 2019b).

In our previous work, we have presented a metagenomics approach to obtain the same level of precision as the conventional bacterial detection methods and isolate's WGS, through direct sequencing of all DNA in the sample after enrichment in a non-selective medium following the ISO standard ISO 13136:2012 (ISO: International Organization for standardization, 2012; Buytaers et al., 2020, 2021b). After short-read sequencing of 12 DNA extracts with or without removal of host DNA in a 48-h Illumina MiSeq run, we were able to link the pathogenic strains derived from metagenomics sequencing of samples containing multiple strains of the same species (*Escherichia coli*) to human isolates from the same outbreak (Buytaers et al., 2020). This was possible using a bioinformatics workflow classifying short reads to a reference genome database (Saltykova et al., 2020). Although Illumina is a widely used sequencing technology generating short reads with high accuracy, it still comes at a high cost for metagenomics, impeding a real implementation in routine. Moreover, the rather long library preparation time for multiple samples that have to be multiplexed to make the run cost-effective, as well as the 48-h sequencing run time, is not ideal for a fast response in case of an ongoing outbreak. Real-time long-read sequencing is now offered by Oxford Nanopore Technologies (ONT) with faster library preparation protocols coupled with the flexibility to cost-efficiently sequence one sample at a time on the flow cells. This could speed up the analysis of samples in an outbreak investigation and help to decrease the cost, which remains important, of metagenomics if using more cost-effective consumables for lower amounts of samples such as the MinION flow cell or the new lower output Flongle flow cell. Furthermore, long-read sequencing offers the possibility to investigate larger genome fragments without the possible bias of short-read metagenomics assembly, which could offer an added value in the context of metagenomics-based outbreak investigation.

Sequencing using Oxford Nanopore Technologies has been previously validated for the characterization of foodborne pathogenic isolates, even during the course of an outbreak (Loman et al., 2015; Quick et al., 2015; Greig et al., 2019), and has since then been tested in some metagenomics studies for pathogen identification by species and gene detection in the mixed reads (Schmidt et al., 2017; Charalampous et al., 2019). It was shown to allow attribution of potentially pathogenic taxa to the corresponding antimicrobial resistance genes they harbored by gene walkout (Leggett et al., 2020). However, strain-level characterization is necessary for the precise resolution of an outbreak, which remains a challenge for ONT metagenomics data partly due to the higher error rate of the technology (Forbes et al., 2018; Gardy and Loman, 2018). In a previous study, Hyeon et al. (2018) used an enriched food sample that was artificially contaminated with *Salmonella*, treated with immunomagnetic separation to concentrate the target bacteria, and whole-genome amplification before it was sequenced using the MinION technology. They obtained 65 and 70 SNP difference

to the WGS isolate reference of the spiked bacteria after 1.5 and 48.5 h of sequencing, respectively (Hyeon et al., 2018). A similar quasimetagenomics method was used to target Shiga toxin-producing *E. coli* (STEC) and *Salmonella* in contaminated flour samples (Forghani et al., 2020). The method proved successful to cluster (without specifying the SNP differences) the metagenomics-obtained strain to the spiked isolate, for multiple single-spiked strains of each pathogen and also on samples co-spiked with one strain of each of the two pathogen species. However, this approach is still rather new, and new proofs of concept are necessary to demonstrate that it can be effectively used, possibly with a lower amount of SNP differences, for more reliable cluster definition in daily outbreak investigation. Indeed, it has not yet been tested with a non-selective enrichment method, a procedure closer to the ones currently followed by the reference laboratories (UE, 2005). Moreover, it has not yet shown its efficiency not only in samples possibly presenting multiple strains of the same species but also to cluster the metagenomics-derived strain to related human cases from the same foodborne outbreak. Finally, sequencing not only on the lower cost but also lower output, Flongle flow cell device still remains to be evaluated for such an application.

We present in this study a proof of concept of shotgun metagenomics outbreak investigation performed after ONT sequencing, combined with a new bioinformatics workflow adapted to long reads, to obtain the characterization of the foodborne pathogen at strain level in samples with various strains of the same pathogen (STEC). The spiked food samples were previously sequenced on Illumina and reported in former studies (Buytaers et al., 2020; Saltykova et al., 2020). A comparison between the results obtained with the two sequencing technologies was made. Moreover, a new approach, *in silico* DNA walking, offering the screening of food samples for pathogens at low cost based on long reads after Flongle sequencing, was evaluated after DNA extraction with or without host DNA removal. Finally, a strategy to integrate metagenomics in the current screening and pathogen characterization at the routine laboratories was proposed based on the results obtained after Flongle, MinION, and Illumina sequencing and their respective cost-effectiveness and execution time.

MATERIALS AND METHODS

Selection of the Sample

Minced beef meat harboring a natural population of commensal *E. coli* bacteria and artificially contaminated with a low infection dose of STEC from a previous study (Buytaers et al., 2020) was used to evaluate the performance of MinION and Flongle sequencing compared to Illumina MiSeq sequencing on the same sample. Briefly, 25 g of the food matrix spiked with 5 colony-forming units (CFU) of STEC was enriched in buffered peptone water for 24 h at 37°C, following the culture described in ISO 13136:2012 for STEC detection in food (ISO: International Organization for standardization, 2012) in order to be representative of the procedures followed by the reference laboratories and therefore the samples they could get to analyze. One milliliter of the mix was used for DNA extraction using the NucleoSpin Food kit (Macherey-Nagel, Düren, Germany) or HostZERO Microbial DNA kit (Zymo Research, Irvine, CA, United States). The latter is advertised as able to remove host DNA. The strain that was chosen to artificially contaminate the food matrix was a STEC O157:H7 *eae*+, *stx1*+, *stx2*+, isolated during an outbreak in Limburg, Belgium, in 2012 (Braeye et al., 2014), and previously characterized through WGS (Nouws et al., 2020b). A negative control, a blank of the enriched food matrix, was previously sequenced on Illumina MiSeq and characterized to pinpoint the presence of commensal *E. coli* bacteria and the absence of STEC virulence genes in the meat prior to spiking (Buytaers et al., 2020; Saltykova et al., 2020).

Oxford Nanopore MinION Sequencing

The DNA library was prepared with the Genomic DNA by Ligation protocol (SQK-LSK109; Oxford Nanopore Technologies, Oxford, United Kingdom) on the DNA extracted with the NucleoSpin kit. It was performed according to the recommendations for MinION sequencing on a MinION flow cell (R9.4.1). The prepared library was then loaded on a primed flow cell (R9.4.1), and a 48-hsequencing run was started, generating 1.2 million reads with a median length of 1,991 bp. The resulting fast5 files obtained at various sequencing time checkpoints were basecalled using Guppy version 4.2.3 (Oxford Nanopore Technologies).

Oxford Nanopore Flongle Sequencing

Two DNA libraries were prepared, respectively, for the DNA extracted with the NucleoSpin and the HostZERO kits with the Genomic DNA by Ligation protocol (SQK-LSK109; Oxford Nanopore Technologies, Oxford, United Kingdom), following recommendations for Flongle sequencing. Each library was then loaded separately on a primed Flongle flow cell (R9.4.1), and a 24-hsequencing run was started, generating 244,019 and 187,966 reads with a median length of 686 and 3,393 bp, respectively, for the NucleoSpin and HostZERO DNA extracts. The basecalling was performed at various sequencing time checkpoints as in the "Oxford Nanopore MinION Sequencing" section.

Long-Read Strain-Level Metagenomics Data Analysis

First, a taxonomic classification with Kraken2 (Wood et al., 2019), using the same databases (in-house database of mammals, archaea, bacteria, fungi, human, protozoa, and viruses) as used for the Illumina analysis of the same samples (Buytaers et al., 2020), was performed on the basecalled reads of MinION and Flongle sequencing, including after specific time check-points. Graphs were created on the classification results using ggplot2 in R.

Second, the presence of virulence genes in the sequenced reads and the genomic context (taxon) of the same sequencing fragment were determined using an *in silico* DNA walking method, previously described for the detection of genetically modified microorganisms using a metagenomics approach (Buytaers et al., 2021a). Briefly, a Basic Local Alignment Search

Tool (BLAST) analysis was performed on all reads using BLASTn version 2.7.1 with default parameters (Camacho et al., 2009) to the databases VirulenceFinder *E. coli* (Joensen et al., 2014) and nucleotide from NCBI [Bethesda (MD): National Library of Medicine (US), 1988]. The hit to the NCBI database of each fragment presenting a virulence gene was used to obtain the genomic origin of the read harboring the virulence factors. The results were finally filtered to retain only the results for the virulence genes *stx1*, *stx2*, *eae*, and *ehxA*. For the goal of this study, focusing on a fast response to the detection of a foodborne pathogen, we presented the results obtained in the shortest timeframe necessary to obtain at least one read confirming the presence of a STEC in the sample. The results were visualized using sunburst charts.

Finally, the *E. coli* strains were inferred using Metamaps v 0.1 (Dilthey et al., 2019). Thereby, ONT reads from MinION and Flongle sequencing, including after specific time checkpoints, were classified against a database containing 2,831 reference sequences corresponding to the 976 complete *E. coli* genomes and complete 1,885 *E. coli* plasmids available from RefSeq on August 11, 2019 (O'Leary et al., 2016). Reads assigned to sub-species-level taxa were extracted.

A gene detection was conducted on the clustered reads of the inferred strains using BLAST version 2.7.1 (Camacho et al., 2009) on the VirulenceFinder *E. coli* database (Joensen et al., 2014) and SerotypeFinder O type and H type (Joensen et al., 2015) with default parameters. The strains containing *stx* genes were considered as STEC strains.

For the phylogenetic analysis, extracted reads of the STEC strain sequenced with ONT devices were mapped to a common STEC reference genome (BA000007.3) using bwa mem v 0.7.17 with the ont2d parameter set. Illumina sequences were previously analyzed through a similar workflow (Buytaers et al., 2020; Saltykova et al., 2020). Bcftools v 1.9 was used for the initial identification of potential SNPs as positions at which at least five reads contained an alternative allele, followed by filtering whereby positions with a minimal depth of 10 reads, a minimal allele frequency of 0.85, and a minimal mapping quality of 50 were retained (**Supplementary Material 1**). Genomic positions that did not meet the minimal sequencing depth and the minimal mapping quality criteria and potential SNPs that did not meet the minimal allele frequency were masked in the consensus sequence. Maximum likelihood substitution model selection and phylogenetic tree inference were performed using MEGA (Kumar et al., 2018), applying the nearest-neighbor-interchange (NNI) heuristic method, keeping all informative sites and using the bootstrap method with 100 replicates as a phylogeny test. The model selected was the Kimura two-parameter model with uniform rates among sites. Strains inferred from the Illumina sequencing of the same metagenomics samples [NucleoSpin extract and HostZERO extract (Buytaers et al., 2020)] and isolates from human (TIAC 1165 and TIAC 1169) and food (TIAC 1151 and TIAC 1152) originating from the same outbreak (Braeye et al., 2014), as well as some sporadic cases from the same serotype O157:H7 (TIAC 1638 and TIAC 1153), were used as background for the phylogenetic tree construction. All isolates were sequenced for a previous study (Nouws et al., 2020b). All workflows of command lines used for bioinformatics analyses in this work are presented in **Supplementary Material 2**.

RESULTS

Long-Read Sequencing on a MinION Flow Cell for Strain-Level Metagenomics Outbreak Investigation

DNA extracted from beef meat spiked with STEC at the lowest infection dose was sequenced on a MinION flow cell. A data analysis workflow was developed in order to produce similar results as those generated with Illumina sequencing (Buytaers et al., 2020) and WGS of isolates, i.e., obtaining and characterizing the reads corresponding to the pathogenic strain in the sample and performing SNP-level phylogeny with this strain.

Taxonomic Classification of All Sequenced Reads

The identification of the taxa present in a metagenomics sample is an important step towards the detection of potential pathogens in a sample. In this case, a taxonomic classification was performed to the genus level on the MinION sequencing output and compared to the results previously obtained with Illumina sequencing of the same sample and of the enriched blank (unspiked) meat (**Figure 1**).

Beef ("*Bos*," blue) was the main species detected in the sample after both sequencing runs (**Figure 1**). This was to be expected as the sample consisted of beef meat. *Ovis* (a genus that includes sheep, olive green) was classified for a small part (2%) of the reads after MinION sequencing. The bacterial genera detected were identical between the two sequencing technologies. *Escherichia*, the pathogen not only artificially spiked in the sample but also endogenously present in the beef before spiking [Blank_Illumina (Buytaers et al., 2020)], was identified for 8 and 6% of the reads after Illumina and MinION sequencing, respectively. All species were detected after 30 min of sequencing on the MinION.

Confirmation of the Presence of a Pathogen in the Sequenced Metagenomics Sample Using *in silico* DNA Walking on Long Reads

In order to indicate the presence of a pathogen in the sample after Illumina sequencing, a virulence gene detection was conducted on all reads (Buytaers et al., 2020). However, with that information, the virulence gene cannot be linked to the pathogen's genome, which would be proof of the presence of the pathogen in the sample. Long-read sequencing offers the possibility to investigate the DNA fragment on which a virulence gene is detected in order to attribute it to a taxon (genomic context). This analysis is also known as *in silico* DNA walking.

As the sample was artificially spiked by a known STEC isolate, our approach was targeted at this pathogen specifically. Therefore, *in silico* DNA walking was applied to all long-read sequences with BLAST on the databases of *E. coli* virulence genes and nucleotides from NCBI, to determine if the *Escherichia*-related virulence genes, in particular *stx* genes defining an *E. coli* as a STEC pathogen, were found on *Escherichia* genome

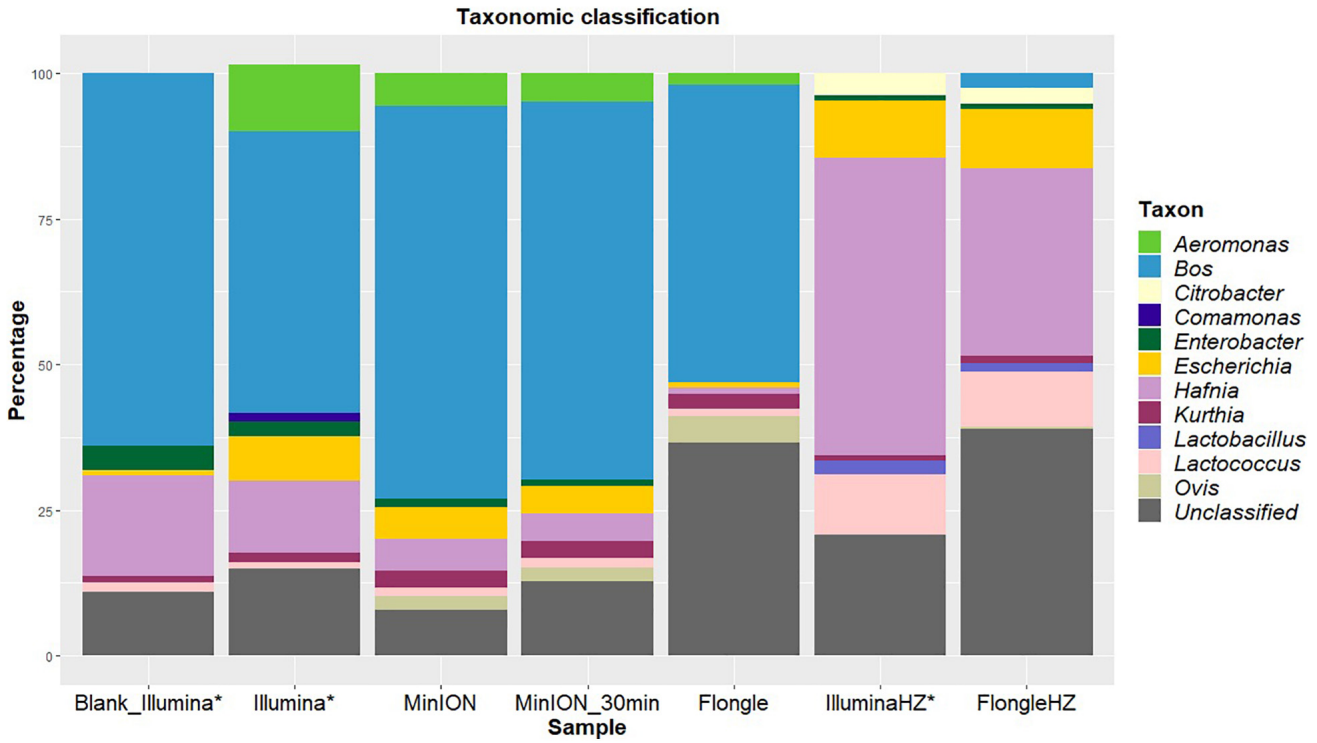


FIGURE 1 | Percentages of reads classified to the genus level using Kraken2 (taxonomic classification tool) from blank and spiked beef samples extracted with two DNA extraction kits (one involving host removal, HZ) and sequenced on Illumina (MiSeq), MinION, or Flongle, with in-house databases of mammals, archaea, bacteria, fungi, human, protozoa, and viruses. The data for Illumina sequencing (*) was published in Buytaers et al. (2020). Light blue represents the proportion of "Bos" corresponding to beef reads. Yellow indicates the presence of "Escherichia" in the sample. The reads that could not be classified to the genus level for mammals, archaea, bacteria, fungi, human, protozoa, or viruses are represented in gray.

sequences, proving the presence of a pathogenic strain in the sample. This approach was tested as a fast alternative to obtain minimal characterization information on the pathogen in the sample before the inference of the strains from the metagenomics reads.

The results, presented in **Figure 2**, show that the virulence genes characteristic of the spiked STEC pathogen (*stx*, *eae*, and *ehxA*) could be linked to *Escherichia* fragments after already 1 h of MinION sequencing. This demonstrated that an *Escherichia* strain carried these genes, therefore indicating that STEC DNA was present in the samples. Moreover, as the enriched blank meat was previously sequenced and characterized (Buytaers et al., 2020), we can rule out the presence of STEC, *E. coli* virulence genes, or *stx* phages in the meat prior to the artificial contamination.

The *stx* genes (*stx1* and *stx2*) were also linked to genomic regions of *Enterobacteriaceae* and to bacteriophages. The reads assigned to *Enterobacteriaceae* could also correspond to STEC bacteria, as *Enterobacteriaceae* is the family of the *Escherichia* genus. Shorter reads may not cover any species- or genus-specific genomic features, preventing their univocal assignment to a single higher level taxon. Such reads are attributed by BLAST to a common ancestor of higher taxa from which the read could potentially be derived, e.g., the family *Enterobacteriaceae*. The same could apply to reads classified as phages, as the *stx* genes present in the STEC genome derive from the integration of these phages, but these could also be present in their mobile form in the environment.

Outbreak Resolution and Strain Characterization From Long-Read Sequences by Strain-Level Inference, Gene Detection, and Single-Nucleotide Polymorphism Phylogeny

Finally, as an equivalent to the characterization of an isolate obtained in routine, a strain-level analysis was performed on all sequenced metagenomics reads to obtain clusters of reads corresponding to the different *E. coli* strains present in the sample. The presence of the STEC strain was confirmed based on the detection of *stx* genes in the clustered reads. It corresponded to a strain mapped to the taxon 741093 from the Metamaps analysis (RefSeq NC_017906.1, NCBI:txid741093). Two other non-pathogenic strains were detected in the samples and mapped to the Metamaps proprietary taxa x494 (RefSeq NZ_CP019271.1, NCBI:txid562) and 745156 (RefSeq NZ_CP009166.1, NCBI:txid745156) (Dilthey et al., 2019). Metamaps uses an extended database taxonomy where some NCBI taxonomic nodes are further subdivided to ensure higher resolution of taxonomic assignment. The same strains were detected after Illumina sequencing (Saltykova et al., 2020). The STEC strain was further investigated for SNP phylogeny to

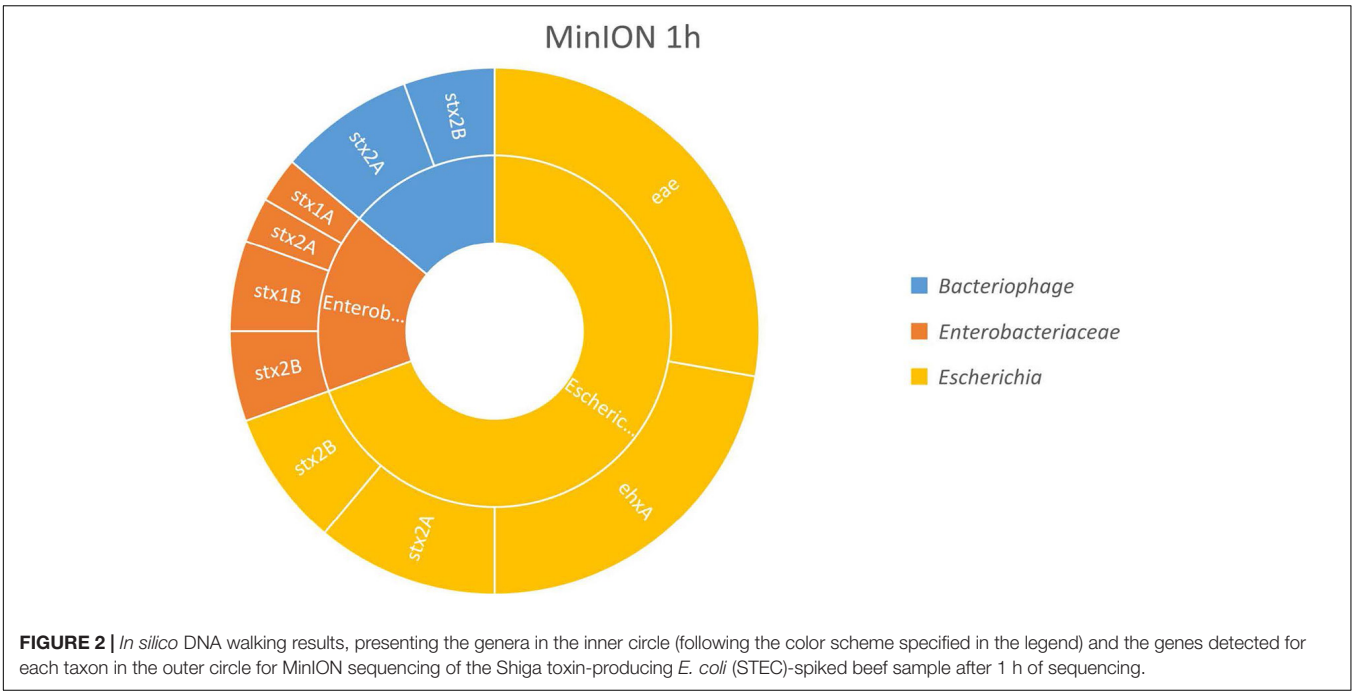


FIGURE 2 | *In silico* DNA walking results, presenting the genera in the inner circle (following the color scheme specified in the legend) and the genes detected for each taxon in the outer circle for MinION sequencing of the Shiga toxin-producing *E. coli* (STEC)-spiked beef sample after 1 h of sequencing.

relate it to other cases (i.e., isolates from food and human origin related to the same outbreak as the spiked isolate and sporadic cases). Strains inferred from Illumina sequencing of the same sample (Buytaers et al., 2020) were also included in the tree (**Figure 3**). The inferred STEC strain obtained after 12, 24, and 48 h of MinION sequencing clustered with the corresponding isolates and metagenomics strain obtained from Illumina sequencing, with 0 SNPs distance (**Supplementary Material 3**), and separated from the sporadic cases. The presence of three virulence genes of importance for STEC characterization (*eae*, *stx1*, and *stx2*), as well as the serotyping genes (O-type and H-type), was also confirmed in the genome of the inferred STEC strain. The serotype and virulence genes in the inferred STEC strain correspond to the genes present in the strain that was spiked. The reference coverage from the MinION run starting from 12 h of sequencing was comparable to the coverage obtained from isolates of the same outbreak and strain inferred from Illumina metagenomic sequencing and therefore considered as sufficient for a phylogenetic analysis. Shorter sequencing time on the MinION did not offer sufficient coverage to conduct the phylogenetic analysis (**Supplementary Material 3**).

Investigation of Long-Read Flongle Sequencing as a Less Expensive Alternative for Strain-Level Metagenomics Outbreak Investigation

The same sample of beef meat containing an endogenous population of non-pathogenic *E. coli* and spiked with a STEC pathogen, previously characterized to the strain level after Illumina sequencing (Buytaers et al., 2020), was sequenced on a Flongle flow cell to investigate a less-expensive alternative. However, as the output of the Flongle is approximatively 10 times lower than the MinION, we also sequenced on the Flongle DNA for which the extraction involved host removal, previously sequenced on Illumina (Buytaers et al., 2020), in an attempt to increase the amount of reads linked to the microbial pathogen. The data analysis on the sequenced long reads was the same as the data analysis presented for the long reads sequenced on the MinION. The analysis was also conducted at different time points of the Flongle sequencing run to determine the time needed to achieve the expected results.

Taxonomic Classification of All Sequenced Reads

After Flongle sequencing, the main genus detected in the sample without host DNA removal was *Bos* (**Figure 1**). The same bacterial taxa, with the exception of *Enterobacter*, were detected as for the Illumina and MinION sequencing, including *Escherichia*, but with a higher percentage of unclassified reads. As for MinION sequencing, a small portion (5%) of mammal reads were incorrectly classified as *Ovis*.

The DNA extract treated with host DNA removal agent (FlongleHZ) presented 2% of reads classified as *Bos* and 0.5% of reads classified as *Ovis*, although no reads were classified as mammals in the Illumina sequencing of the same DNA extract. However, this is a large decrease compared to the amount of Flongle reads classified as eukaryotes without the host DNA removal step (50%). The bacterial taxa detected were the same for this sample after Illumina or Flongle sequencing and, except for the absence of *Aeromonas* and *Comamonas* and the presence of *Citrobacter* and *Lactobacillus*, were identical to the bacterial taxa detected without host DNA removal. This difference might be explained by the presence of bacterial DNA in the extraction buffer or its presence at very low level in the food sample. *Escherichia* represented 10% of the reads, which is slightly higher than the values obtained without host DNA removal. The sample

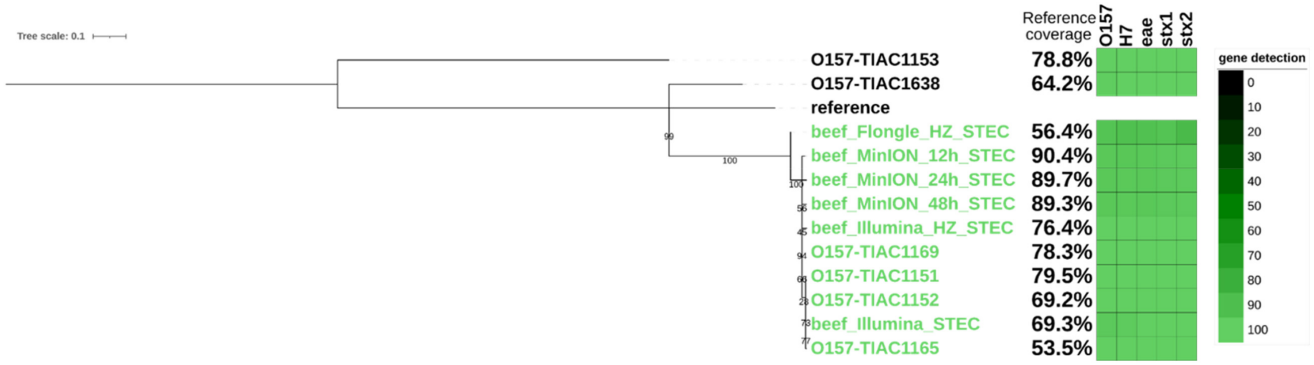


FIGURE 3 | Single-nucleotide polymorphism (SNP)-based phylogenetic tree of STEC strains inferred from metagenomics sequencing (beef) and of sequenced isolates with percentage of the reference genome covered (i.e., percentage of reference genome that is useful for SNP analysis, see section "Materials and Methods") and gene detection (O-type and H-type and genes eae, stx1, and stx2; green shaded blocks representing the query coverage) in each strain represented on the side of the branch. Isolates TIAC 1151, 1152, 1153, and 1638 are from food origin. Isolates TIAC 1165 and 1169 are from human origin. Reference: *E. coli* O157:H7 str. Sakai (BA000007.3). Green: closely related strains from the outbreak cluster. Black: sporadic cases outside the outbreak cluster. The scale bar represents nucleotide substitution per 100 nucleotide sites. Node values represent bootstrap support values.

with host DNA removal sequenced on the Flongle presented the highest percentage of unclassified reads (39%).

Confirmation of the Presence of a Pathogen in the Flongle-Sequenced Metagenomics Sample Using *in silico* DNA Walking on Long Reads

Similar as for MinION sequencing, an *in silico* DNA walking was conducted in order to attribute a genomic context (taxon) to detected virulence genes. This analysis was conducted on all reads generated at different time points during the Flongle sequencing of the two DNA extracts (with or without host DNA removal).

After 1 h of sequencing, the virulence genes characteristic of a STEC (i.e., *stx*, *eae*, and *ehxA*) could be retrieved in the sample treated with host DNA removal (**Figure 4A**) and detected on genome fragments that could be assigned to *Escherichia*. Similarly as with the MinION analysis, the virulence genes were also found associated in smaller proportions to *Enterobacteriaceae*, which correspond to the family of the *Escherichia* genus, or *stx1* phage, the bacteriophage carrying the *stx1* gene that can be inserted in the STEC genome. The classification to a higher level (*Enterobacteriaceae* or phage) might be explained by the short length of the reads.

Without host DNA removal (**Figure 4B**), 5 h of sequencing were sufficient to obtain the required information to determine that the pathogen was present in the sample, i.e., virulence genes *stx*, *eae*, and *ehxA* associated to *Escherichia* genome. Again, the virulence genes could be also assigned to *Enterobacteriaceae*, as well as bacteriophage for some *stx2* genes.

As a STEC is defined as an *E. coli* harboring an *stx* gene, the information presented was sufficient to conclude that a STEC was present in the samples, after 1 h of sequencing with host DNA removal and 5 h of sequencing without host DNA removal. However, a longer sequencing time would be required to obtain all virulence genes characterizing the strain that was spiked. In our workflow, the full characterization of the STEC present in the sample is done at the next step, after strain inference, in order to characterize specifically each potential pathogenic strain present in the sample.

Outbreak Resolution and Strain Characterization From Flongle Long-Read Sequences by Strain-Level Inference, Gene Detection, and Single-Nucleotide Polymorphism Phylogeny

The different *E. coli* strains present in the sample were inferred from the reads of the two Flongle sequencing runs, and the STEC strain was identified among these strains after detection of *stx* genes in the clustered reads (**Supplementary Materials 4, 5**).

The pathogenic strain corresponded to reads that mapped to the Metamaps taxon 741093 (RefSeq NC_017906.1, NCBI:txid741093) for the DNA extract without host DNA removal, i.e., a similar strain as found with MinION sequencing, and to Metamaps taxon x13 (RefSeq NZ_CP012802.1, NCBI:txid83334) for the DNA extract with host DNA removal, which is also a STEC O157:H7. The endogenous strains were mainly mapped as Metamaps taxon 745156 (RefSeq NZ_CP009166.1, NCBI:txid745156), similarly as for the MinION sequencing, as well as Metamaps taxon x311 (RefSeq NZ_CP019267.1, NCBI:txid562) for the extract without host DNA removal (**Supplementary Materials 4, 5**).

After SNP calling, it was observed that the coverage of the reference genome (**Supplementary Material 3**) was insufficient to conduct a SNP-level phylogenetic analysis (less than 1%) for the DNA extract without host DNA removal. Therefore, the inferred STEC strain obtained after Flongle sequencing of the DNA extract without host DNA removal was not included in the phylogenetic tree. However, 24 h of Flongle sequencing of the DNA extract with host DNA removal led to obtaining clustered reads covering 56% of the genome at or above 10× coverage, which was sufficient to cluster the metagenomics-derived strain with the outbreak cases on the phylogenetic tree (**Figure 3**). Serotyping genes (O-type and H-type) as well as virulence genes *eae*, *stx1*, and *stx2* could be detected with high

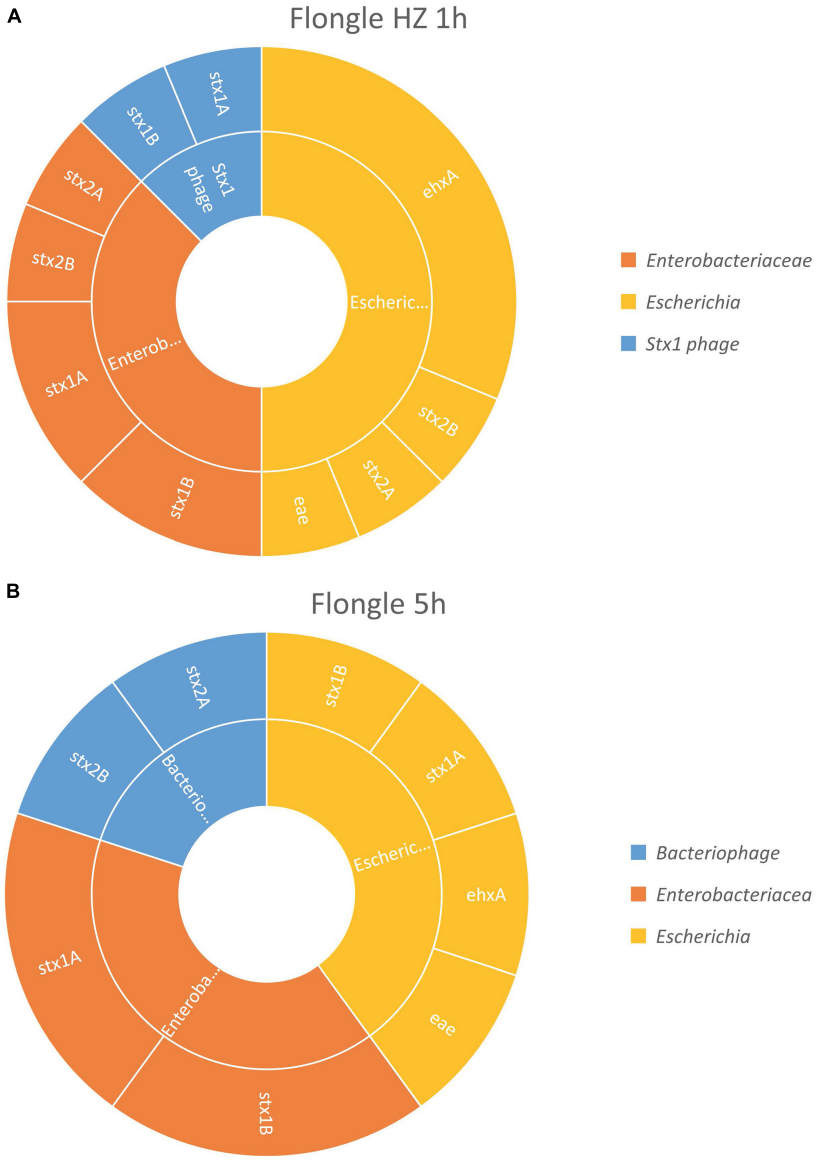


FIGURE 4 | *In silico* DNA walking results, presenting the genera in the inner circle (following the color scheme specified in the legend) and the genes detected for each taxon in the outer circle of the STEC-spiked beef sample after DNA extraction with or without host removal. **(A)** Flongle sequencing after 1 h of sequencing, DNA extract with host removal. **(B)** Flongle sequencing after 5 h of sequencing, DNA extract without host removal.

identity in the strain, confirming that it was similar to the spiked strain. A distance of 0–3 SNPs per million genomic positions (**Supplementary Material 3**) was observed for the other isolates from the same outbreak as well as the metagenomics-derived strains from Illumina or MinION sequencing, which is in the expected range. However, the distances of the outbreak strain to the background isolates (TIAC 1153 and TIAC 1638) were somewhat lower with Flongle sequencing data after host DNA removal than with MinION and Illumina sequencing data (30 SNPs per million of genomic positions for the Flongle sequencing compared to 39–46 SNPs to TIAC 1638 for Illumina and MinION, and 126 SNPs per million of genomic positions for the Flongle sequencing compared to 139–155 SNPs to TIAC 1153 for Illumina and MinION; **Supplementary Material 3**), indicating that not all SNPs could be called at the obtained coverage.

DISCUSSION

The rapid and precise characterization of a pathogen during foodborne outbreak investigation, as well as the tracing back to the food source, is crucial to stop further spreading of the infections. Therefore, a metagenomics approach has been proposed as an alternative to the currently performed microbiological analyses requiring a not-always-straightforward isolation of the pathogenic strain. As previously described

(Buytaers et al., 2020), Illumina sequencing may be used to obtain the full information necessary for outbreak investigation from metagenomics samples, without the need for isolation, and this to the strain level, after about one full week of lab work (Buytaers et al., 2021b). However, not only the need for more proofs of concept but also the high cost of such an analysis impact its potential implementation as a routine practice. To render the analysis somehow more cost-effective, while still taking the required coverage into account, 8–12 samples were pooled into one Illumina MiSeq run in previous studies (Leonard et al., 2016; Buytaers et al., 2020, 2021b). However, it might not always be possible to analyze this number of samples as the number of available food samples during outbreak investigation varies and is not gathered at a single time point. Besides, delaying the sequencing run to gather sufficient samples is not an option when a fast response is required, especially in outbreak investigation. Using a smaller number of samples in the run would however substantially increase the sequencing cost per sample. Moreover, these runs, generating 2×250-bp reads, have a set sequencing duration of 48 h, which is significant during ongoing outbreak investigations. Long-read sequencing and flexibility in sequencing time, which is made possible by ONT, could offer a solution to these drawbacks.

In this study, we first sequenced an artificially contaminated sample (beef containing an endogenous community of non-pathogenic bacteria including *E. coli*, spiked at very low dose with STEC), previously sequenced on Illumina (Buytaers et al., 2020), with a MinION flow cell. The data analysis followed the same flow as the analysis previously described for Illumina sequencing (Buytaers et al., 2020), but with adapted algorithms and tools for taxonomic classification, virulence gene detection, and genome inference of long reads. This allowed to match the contaminated food with human isolates from the same outbreak, after a shorter sequencing time. After only 12 h of sequencing, endogenous and pathogenic *E. coli* strains could be obtained from the sequenced reads, and the clustered reads corresponding to the STEC could be linked to outbreak isolates from food and human origin. The virulent strain-related reads harbored all virulence genes expected from the spiked bacteria and could be placed accurately in a phylogenetic tree with 0 SNP difference to the outbreak cluster, which is much lower than the SNP distance previously obtained after metagenomics ONT sequencing (Hyeon et al., 2018). The high number of SNPs observed by Hyeon et al. (2018) could be due to the specificity of the SNP calling procedure that was used. The authors applied a workflow based on the CFSAN pipeline, with a relatively low minimal allele frequency for SNP calling (0.6). These settings have, however, shown to exhibit a lower SNP calling accuracy even with the more accurate Illumina sequencing data, with higher SNP distances between the outbreak isolates as a result (Saltykova et al., 2018). Moreover, the long reads sequenced with ONT also offer the possibility to investigate at the same time the genomic context of reads carrying these virulence genes, in an *in silico* DNA walking approach. Recently, the European authorities proposed the attribution of the virulence genes to their respective bacterial host as an isolate-free alternative to confirm the presence of foodborne pathogens after metagenomics sequencing (EFSA, 2019b). We were able to detect the expected virulence genes (*stx*, *eae*, and *ehxA*) on genome fragments that could be linked to an *Escherichia* genome, therefore confirming the presence of the pathogen in the sample after only 1 h of sequencing on the MinION flow cell. This approach could eventually be implemented in real time while receiving data from the sequencer, as it has been shown previously for AMR genes (Leggett et al., 2020).

MinION sequencing offers the opportunity to work with long reads, allowing access to the genomic context of the reads sequenced, as well as the flexibility of real-time sequencing and sequencing one sample at a time. However, it remains an expensive consumable, and therefore, the lower cost Flongle flow cell was also tested. The Flongle flow cell was ideal to rapidly obtain a confirmation of the presence of a pathogen in the sample at the lowest cost after taxonomic classification and *in silico* DNA walking. Indeed, it allowed to confirm the presence of the STEC strain (detection of *stx* gene in *Escherichia* genome) after 1 h of sequencing if host DNA removal was conducted or 5 h with traditional DNA extraction. As the output of the Flongle flow cell is substantially lower compared to the output of the MinION flow cell, retrieval of information for strain comparison was only possible when host DNA removal was conducted during the DNA extraction. The coverage of the reference genome by the clustered reads corresponding to the STEC strain obtained from the extract without host DNA removal was not sufficient to establish phylogenetic links. The threshold to determine if a strain contains sufficient reads using metagenomics to perform further characterization or SNP phylogeny is hard to define strictly as lower coverages are also observed for genomics on isolated strains (e.g., TIAC 1165 covering 54% of the reference genome, **Figure 3**). More analyses such as the one within this study, including for other pathogens, are necessary to pinpoint such limits. We could also observe that the reference genome to which the reads were assigned after Flongle sequencing was different from the reference genomes mostly covered after MinION or Illumina analyses. However, the different references to which the STEC reads clustered were all STEC O157, and the interpretation of the results was not impacted by the reference (SNP-level phylogeny obtained for the different strains). A future alternative could be to pool reads assigned to groups of similar references instead of working with individual references, as already proposed in the work of Saltykova et al. (2020) for short-read sequences.

In the present work, a three-step analysis has been applied on food samples sequenced with different flow cells. For each step, the minimal time required to obtain results was assessed. First, a taxonomic classification to obtain an overview of the genetic content of the food sample, followed by virulence gene detection coupled to an *in silico* DNA walking method for hazard identification. These tests could be performed in a very fast timeframe of a few hours, depending on the treatment of the DNA extract and the selected sequencing flow cell, and could even be implemented in real time in the future. This could potentially solve partially, i.e., when food leftovers are available and it concerns a bacterial origin (for which isolation is currently the routine approach), the issue of foodborne outbreaks

for which the food source cannot be determined, accounting for 60% of all (i.e., also including other agents as source) outbreaks notified in the EU (EFSA, 2021). Finally, as a third step, strain-level phylogeny in order to relate human cases of an outbreak to its food source can be achieved after 12 h of sequencing on a MinION flow cell or 24 h on a Flongle flow cell if host DNA removal was applied during the DNA extraction. Notably, for *in silico* DNA walking, a threshold of one read harboring an *stx* gene and traced back to the *Escherichia* genus was considered as sufficient to determine the minimal time to suspect the presence of a pathogen in the sample. However, a discussion within the international scientific community is necessary to determine such threshold, and we recommend to continue the sequencing after this minimal time to collect more information. Moreover, obtaining and characterizing the pathogenic strain (third step) are still necessary to confirm the suspicion. Based on this work, a new strategy for detection of bacterial pathogens in food, using shotgun metagenomics, could be proposed to the reference laboratories (**Figure 5**): a screening of all food samples that might be related to a foodborne outbreak, including those for which the contaminant is unknown, for pathogens using the Flongle and taxonomic classification followed by virulence gene detection and *in silico* DNA walking, potentially in real time. This might involve additional enrichment media and/or conditions to be able to cover all bacterial foodborne pathogens, depending on the specific outbreak based on patient's symptoms, to fully replace the conventional way of working. Once the presence of a bacterial pathogen is confirmed in a food sample, this analysis can be followed by a strain-level read classification and phylogeny that can be attempted on the Flongle sequencing data or, if not possible, based on further Illumina or MinION sequencing. Also, the choice of the sequencing technology will depend on a cost analysis based on the amount of samples to sequence as well as the timeframe to obtain results and the capacities of the laboratory. This strategy could be run in parallel with attempting to obtain a bacterial isolate from the same food samples, which would ideally be sequenced with WGS in order to populate the still necessary databases that are required to perform the metagenomics-based bioinformatics analyses. The perspective of such a strategy consolidates the new perception that metagenomics has the ability to be used as a new alternative for outbreak investigation, source attribution, and risk assessment of foodborne microorganisms (EFSA, 2019b). In order to implement the same data analysis applied to artificially STEC-contaminated samples in this study to other bacterial pathogens, the same workflow can be followed, and only the databases for gene detection and read mapping have

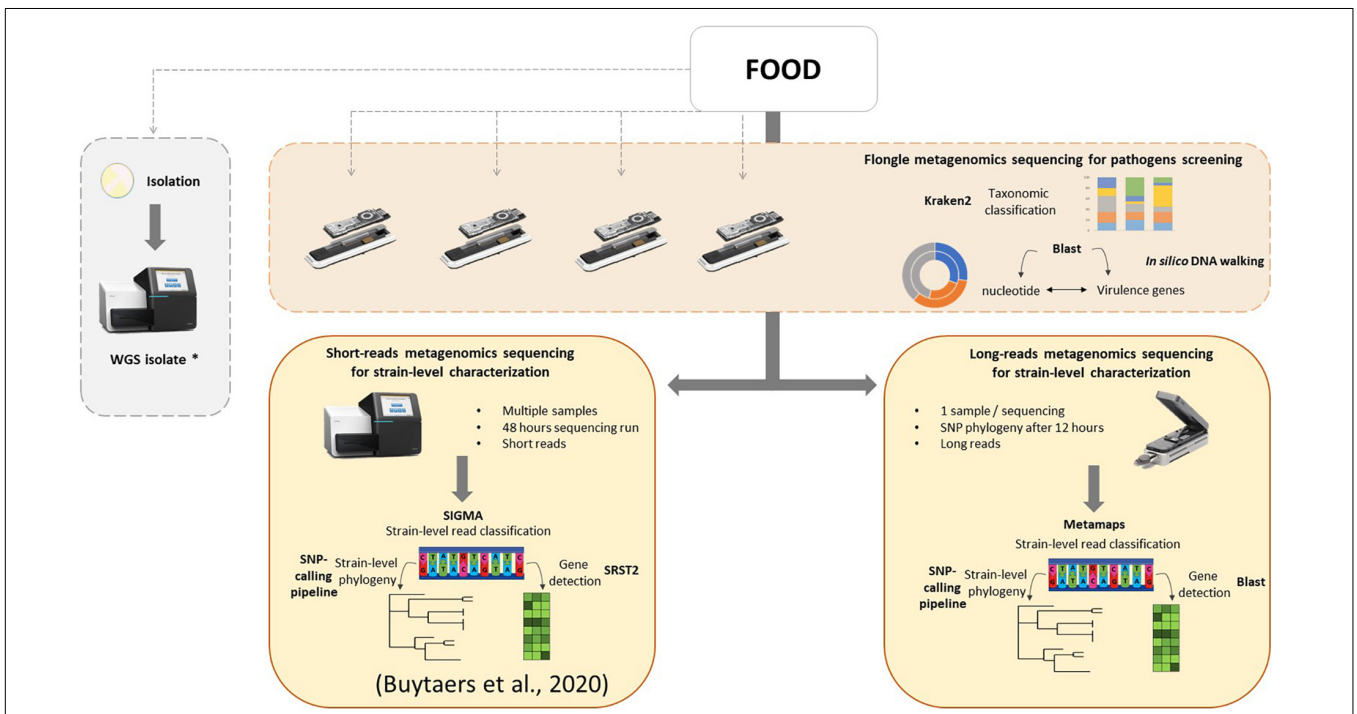


FIGURE 5 | Integrated metagenomics-based strategy for microbiological foodborne outbreak investigation. As first optional steps, food samples can be screened for the presence of bacterial pathogens using metagenomics Flongle sequencing, taxonomic classification, and *in silico* DNA walking (based on BLAST to the nucleotide database and virulence genes database) in parallel with the ongoing attempt to isolate the pathogen, followed by WGS of the obtained isolates. A strain-level characterization can be attempted from the Flongle sequencing or conducted after using an Illumina strategy (more cost-effective for multiple samples) or a MinION strategy (fast response for one sample) in food samples for which the presence of a pathogen is confirmed. The strain-level data analysis for Illumina sequencing was previously presented Buytaers et al. (2020). The strain-level data analysis workflow for MinION sequencing is based on classification using Metamaps, a gene detection with BLAST, and phylogenetics with a SNP-calling pipeline. The asterisk (*) indicates that WGS isolate data is interesting to feed to reference genome databases for the classification of the metagenomics reads for future analyses.

to be adapted according to the contaminant(s) detected through taxonomic classification.

The implementation of such a metagenomics approach in routine, however, still requires overcoming several challenges. First, the data analysis currently requires sufficient informatics hardware, especially performant GPUs for real-time base-calling and analysis. Additionally, trained bioinformaticians are needed, as no automated pipeline has been developed yet for a strain-level pathogen characterization. Benchmarking studies comparing more bioinformatics tools need to be performed to identify tools allowing to obtain similar results in the same, or even faster, timeframe. For this, we believe that studies such as this one offer interesting datasets to be explored further. Second, some consumables like the Flongle flow cells, which have a very short storage life, can be difficult to obtain in a short timeframe when the demand exceeds the production capacities as experienced during the COVID-19 pandemic. The output of Flongle flow cells might also be difficult to predict due to the possible instability of the very low amount of pores not only before loading but also after loading. This might affect low-level contamination samples. The MinION sequencing resulted in a larger output, allowing the potential applicability to other samples, regardless of the quality of the flow cell (number of pores). The work of Forghani et al. (2020) showed that a quasimetagenomics method to the strain level with MinION sequencing can be extended to other strains of STEC and other bacterial pathogens without a problem. However, our study, although only including one serotype, showed the potential of the metagenomics approach for samples presenting a population of several different *E. coli* strains (including non-pathogenic strains). More studies might be necessary to validate the potential of long-read strain-level metagenomics for food safety assessment and foodborne outbreak investigation for other pathogens, including viruses and parasites. The enrichment and extraction methods might have to be adapted depending on the pathogens to investigate. Moreover, while we analyzed samples contaminated with the lowest infectious dose, more studies with different contamination loads might lead to a more precise limit of detection for the method, especially as the number of pathogenic cells is undetermined after enrichment.

In conclusion, this work is a proof of concept of the potential to conduct real-time and affordable strain-level outbreak investigation based on ONT long reads, testing the potential of the MinION as well as the Flongle flow cells. Although a limited amount of samples and only one STEC strain was included in our proof-of-concept study, we demonstrated the ability to obtain the characterization and relatedness of a STEC spiked at a very low dose in a food matrix based on metagenomics sequencing on a MinION flow cell after only 12 h or on a Flongle flow cell after 24 h if host DNA removal was applied during the DNA extraction. Moreover, we also presented a rapid strategy to confirm the presence of a pathogen in a food matrix based on long-read sequencing without the need for isolation (i.e., *in silico* DNA walking). All this was possible on food samples enriched in a non-selective medium following the ISO practice. This makes it particularly interesting for reference laboratories when only limited quantities of the food samples are left, and there is no need for sequential culturing steps on pathogen-specific selective media, with pathogen-specific growth conditions. Moreover, with the method we propose, food that has been enriched at the reference laboratory can be sequenced with a metagenomics workflow in parallel to the isolation protocol, without the need for a different enrichment protocol, which can be particularly interesting when no isolate can be obtained. Finally, these results allowed proposing a more global perspective, as a metagenomics-based strategy to be used by the routine (reference) laboratories, determined by the required level of information required, cost-effectiveness, and timeframe to obtain results. This contributes to the demand of the EFSA asking to demonstrate the ability of metagenomics to be used as a new alternative for risk assessment, source attribution, and outbreak investigation (EFSA, 2019b).

DATA AVAILABILITY STATEMENT

All sequencing data is publicly available at NCBI SRA under BioProject PRJNA736700.

AUTHOR CONTRIBUTIONS

FB, AS, and SDK conceptualized the study, conducted the formal analysis, conducted the investigation, were responsible for the methodology, and wrote the original draft. FB, SD, BV, and DP curated the data. NR and SDK were involved in the funding acquisition and were involved in project administration. FB, SD, BV, KV, NR, and DP provided the resources. FB, AS, and KV were responsible for the software. KM and SDK supervised the study. SD, NR, DP, and SDK performed the validation. FB was responsible for the visualization. All authors reviewed and edited the manuscript, contributed to the article, and approved the submitted version.

FUNDING

This research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RF 17/6316 StEQIDEMIC.be and by Sciensano through contract RP NeXSplorer.iph and Be READY.

ACKNOWLEDGMENTS

We are grateful to Stefan Hoffman from Transversal Activities in Applied Genomics (Sciensano) for the initial set-up of Oxford Nanopore Technologies in the service.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.738284/full#supplementary-material

Supplementary Material 1 | Reference coverage and allele distribution in the inferred strains from long-reads metagenomics sequencing.

Supplementary Material 2 | Workflow of command lines for bioinformatics analyses used in this study.

Supplementary Material 3 | Reference coverage and total and relative SNP distances (per million genomic positions) of the STEC phylogenetic tree.

Supplementary Material 4 | Metamaps result (STEC reference genome assignment of the reads) after Flongle sequencing.

Supplementary Material 5 | Metamaps result (STEC reference genome assignment of the reads) after Flongle sequencing and host DNA removal.

REFERENCES

Bogaerts, B., Winand, R., Fu, Q., Van Braekel, J., Ceyssens, P. J., Mattheus, W., et al. (2019). Validation of a bioinformatics workflow for routine analysis of whole-genome sequencing data and related challenges for pathogen typing in a European National Reference Center: *Neisseria meningitidis* as a proof-of-concept. *Front. Microbiol.* 10:362. doi: 10.3389/fmicb.2019.00362

Braeye, T., Denayer, S., De Rauw, K., Forier, A., Verluyten, J., Fourie, L., et al. (2014). Lessons learned from a textbook outbreak: EHEC-O157:H7 infections associated with the consumption of raw meat products, June 2012, Limburg, Belgium. *Arch. Public Health* 72:44. doi: 10.1186/2049-3258-72-44

Buytaers, F. E., Saltykova, A., Mattheus, W., Verhaegen, B., Roosens, N. H. C., Vanneste, K., et al. (2021b). Application of a strain-level shotgun metagenomics approach on food samples: resolution of the source of a *Salmonella* food-borne outbreak. *Microb. Genom.* 7:000547. doi: 10.1099/mgen.0.000547

Buytaers, F. E., Fraiture, M.-A., Berbers, B., Vandermassen, E., Hoffman, S., Papazova, N., et al. (2021a). A shotgun metagenomics approach to detect and characterize unauthorized genetically modified microorganisms in microbial fermentation products. *Food Chem. Mol. Sci.* 2:100023. doi: 10.1016/j.fochms.2021.100023

Buytaers, F. E., Saltykova, A., Denayer, S., Verhaegen, B., Vanneste, K., Roosens, N. H. C., et al. (2020). A practical method to implement strain-level metagenomics-based foodborne outbreak investigation and source tracking in routine. *Microorganisms* 8:1191. doi: 10.3390/microorganisms8081191

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Charalampous, T., Kay, G. L., Richardson, H., Aydin, A., Baldan, R., Jeanes, C., et al. (2019). Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* 37, 783–792. doi: 10.1038/s41587-019-0156-5

Dilthey, A. T., Jain, C., Koren, S., and Phillippy, A. M. (2019). Strain-level metagenomic assignment and compositional estimation for long reads with MetaMaps. *Nat. Commun.* 10:3066. doi: 10.1038/s41467-019-10934-2

ECDC, and EFSA (2019). *EFSA and ECDC Technical Report on the Collection and Analysis of Whole Genome Sequencing Data From Food-Borne Pathogens and Other Relevant Microorganisms Isolated from Human, Animal, Food, Feed and Food/Feed Environmental Samples in the Joint ECDC-EFSA Molecular Typing Database*, Vol. 16. Parma: EFSA Supporting Publications, doi: 10.2903/sp.efsa.2019.EN-1337

EFSA (2014). *Use of Whole Genome Sequencing (WGS) of Food-Borne Pathogens for Public Health Protection*. Parma: EFSA, doi: 10.2805/66246

EFSA (2019a). *The European Union One Health 2018 Zoonoses Report*. Parma: EFSA. doi: 10.2903/j.efsa.2019.5926

EFSA (2019b). Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. *EFSA J.* 17:e05898. doi: 10.2903/j.efsa.2019.5898

EFSA (2021). The European Union One Health 2019 zoonoses report. *EFSA J.* 19:6406. doi: 10.2903/j.efsa.2021.6406

European Food Safety Authority (EFSA), and European Centre for Disease Prevention and Control (ECDC) (2018). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. *EFSA J.* 16:e05500. doi: 10.2903/j.efsa.2018.5500

Forbes, J. D., Knox, N. C., Peterson, C. L., and Reimer, A. R. (2018). Highlighting clinical metagenomics for enhanced diagnostic decision-making: a step towards wider implementation. *Comput. Struct. Biotechnol. J.* 16, 108–120. doi: 10.1016/j.csbj.2018.02.006

Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F., and Reimer, A. (2017). Metagenomics: the next culture-independent game changer. *Front. Microbiol.* 8:1069. doi: 10.3389/fmicb.2017.01069

Forghani, F., Li, S., Zhang, S., Mann, D. A., Deng, X., den Bakker, H. C., et al. (2020). *Salmonella enterica* and *Escherichia coli* in wheat flour: detection and serotyping by a quasimetagenomic approach assisted by magnetic capture, multiple-displacement amplification, and real-time sequencing. *Appl. Environ. Microbiol.* 86:e00097-20. doi: 10.1128/AEM.00097-20

Gardy, J. L., and Loman, N. J. (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* 19, 9–20. doi: 10.1038/nrg.2017.88

Greig, D. R., Jenkins, C., Gharbia, S., and Dallman, T. J. (2019). Comparison of single-nucleotide variants identified by Illumina and Oxford Nanopore technologies in the context of a potential outbreak of Shiga toxin–producing *Escherichia coli*. *Gigascience* 8:giz104. doi: 10.1093/gigascience/giz104

Hyeon, J., Li, S., Mann, D. A., Zhang, S., Li, Z., Chen, Y., et al. (2018). Quasimetagenomics-based and real-time-sequencing-aided detection and subtyping of *Salmonella enterica* from food samples. *Appl. Environ. Microbiol.* 84:e02340-17. doi: 10.1128/AEM.02340-17

ISO: International Organization for standardization (2012). *ISO/TS 13136:2012 Microbiology of Food and Animal Feed–Real-Time Polymerase Chain Reaction (PCR)-Based Method for the Detetion of Food-Borne Pathogens–Horizontal Method for the Detection of Shiga Toxin-Producing Escherichia coli (STEC) and the Determi*. Geneva: International Organization for standardization.

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13

Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., and Scheutz, F. (2015). Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J. Clin. Microbiol.* 53, 2410–2426. doi: 10.1128/JCM.00008-15

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Leggett, R. M., Alcon-Giner, C., Heavens, D., Caim, S., Brook, T. C., Kujawska, M., et al. (2020). Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat. Microbiol.* 5, 430–442. doi: 10.1038/s41564-019-0626-z

Leonard, S. R., Mammel, M. K., Lacher, D. W., and Elkins, C. A. (2016). Strain-level discrimination of shiga toxin-producing *Escherichia coli* in spinach using metagenomic sequencing. *PLoS One* 11:e0167870. doi: 10.1371/journal.pone.0167870

Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444

National Library of Medicine (US) (1988). *National Center for Biotechnology Information (NCBI)*. Bethesda, MD: National Library of Medicine (US).

Nouws, S., Bogaerts, B., Verhaegen, B., Denayer, S., Crombé, F., De Rauw, K., et al. (2020a). The benefits of whole genome sequencing for foodborne outbreak investigation from the perspective of a National Reference Laboratory in a Smaller Country. *Foods* 9:1030. doi: 10.3390/foods9081030

Nouws, S., Bogaerts, B., Verhaegen, B., Denayer, S., Piérard, D., Marchal, K., et al. (2020b). Impact of DNA extraction on whole genome sequencing analysis for characterization and relatedness of Shiga toxin-producing *Escherichia coli* isolates. *Sci. Rep.* 10:14649. doi: 10.1038/s41598-020-71207-3

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., et al. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* 16:114. doi: 10.1186/s13059-015-0677-2

Saltykova, A., Buytaers, F. E., Denayer, S., Verhaegen, B., Piérard, D., Roosens, N. H. C., et al. (2020). Strain-level metagenomic data analysis of enriched

in vitro and in silico spiked food samples: paving the way towards a culture-free foodborne outbreak investigation using STEC as a case study. *Int. J. Mol. Sci.* 21:5688. doi: 10.3390/ijms21165688

Saltykova, A., Wuyts, V., Mattheus, W., Bertrand, S., Roosens, N. H. C., Marchal, K., et al. (2018). Comparison of SNP-based subtyping workflows for bacterial isolates using WGS data, applied to *Salmonella enterica* serotype Typhimurium and serotype 1,4,[5],12:i:-. *PLoS One* 13:e0192504. doi: 10.1371/journal.pone.0192504

Sandora, T. J., Gerner-Smidt, P., and McAdam, A. J. (2014). What's your subtype? The epidemiologic utility of bacterial whole-genome sequencing. *Clin. Chem.* 60, 586–588. doi: 10.1373/clinchem.2013.217141

Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., et al. (2017). Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* 72, 104–114. doi: 10.1093/jac/dkw397

UE (2005). COMMISSION REGULATION (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. *Off. J. Eur. Union* 32, 1–26.

WHO (2015). *WHO Estimates of the Global Burden of Foodborne Diseases*. Geneva: WHO.

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *bioRxiv* [Preprint] doi: 10.1101/762302

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 04 February 2022
doi: 10.3389/fmicb.2021.798010

Comparative Genome Analysis of Three *Komagataeibacter* Strains Used for Practical Production of Nata-de-Coco

*Koji Ishiya[1], Hideki Kosaka[2], Takashi Inaoka[3], Keitarou Kimura[3] and Nobutaka Nakashima[1]**

[1] Bioproduction Research Institute, National Institute of Advanced Industrial Sciences and Technology (AIST), Sapporo, Japan, [2] Research and Development Department, Fujicco Co., Ltd., Chuo-ku, Japan, [3] Institute of Food Research, National Agriculture and Food Research Organization (NFRI/NARO), Tsukuba, Japan

We determined the whole genome sequences of three bacterial strains, designated as FNDCR1, FNDCF1, and FNDCR2, isolated from a practical nata-de-coco producing bacterial culture. Only FNDCR1 and FNDCR2 strains had the ability to produce cellulose. The 16S rDNA sequence and phylogenetic analysis revealed that all strains belonged to the *Komagataeibacter* genus but belonged to a different clade within the genus. Comparative genomic analysis revealed cross-strain distribution of duplicated sequences in *Komagataeibacter* genomes. It is particularly interesting that FNDCR1 has many duplicated sequences within the genome independently of the phylogenetic clade, suggesting that these duplications might have been obtained specifically for this strain. Analysis of the cellulose biosynthesis operon of the three determined strain genomes indicated that several cellulose synthesis-related genes, which are present in FNDCR1 and FNDCR2, were lost in the FNDCF1 strain. These findings reveal important genetic insights into practical nata de coco-producing bacteria that can be used in food development. Furthermore, our results also shed light on the variation in their cellulose-producing abilities and illustrate why genetic traits are unstable for *Komagataeibacter* and *Komagataeibacter*-related acetic acid bacteria.

Keywords: cellulose, comparative genome analysis, duplicated sequences, *Komagataeibacter*, nata-de-coco production

OPEN ACCESS

Edited by:
Jerome Combrisson, Mars, United States

Reviewed by:
Mario Tello, University of Santiago, Chile
Rodrigo Bentes Kato, Federal University of Minas Gerais, Brazil

****Correspondence:***
Nobutaka Nakashima
n-nakahsima@aist.go.jp

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *19 October 2021*
Accepted: *22 December 2021*
Published: *04 February 2022*

Citation:
Ishiya K, Kosaka H, Inaoka T, Kimura K and Nakashima N (2022) Comparative Genome Analysis of Three Komagataeibacter Strains Used for Practical Production of Nata-de-Coco. Front. Microbiol. 12:798010. doi: 10.3389/fmicb.2021.798010

INTRODUCTION

Numerous bacteria produce cellulose extracellularly. Bacterial cells are often entrapped in the cellulose fiber, forming pellicle at the air–liquid interface. Depending on the bacterial strain and culture conditions, cells consume as much as 10% of the total energy budget for cellulose biosynthesis (Ross et al., 1991; Chawla et al., 2009). Although the reason for forming pellicle remains controversial, bacterial cells in the pellicle might be able to yield oxygen more efficiently than planktonic cells in liquid (Lee et al., 2014). Bacterial cellulose is expected to offer great potential for use in biomedical and food industries because of its high water-holding capacity, ultrafine network structure, biocompatibility, and high tensile strength in a wet state (Londrina et al., 2018).

A well-known example of bacterial cellulose in industry is nata-de-coco, an indigenous jelly-like food of Southeast Asia. The fibrous base of nata-de-coco, which consists of pure bacterial cellulose, is produced industrially through fermentation of coconut water using acetic acid bacteria (AAB). Among AAB, the species of *Komagataeibacter* genus (also known as *Gluconacetobacter* genus), which are kinds of α-proteobacteria and obligately aerobic and acidophilic bacteria, are the most preferably employed practically because the bacteria of this genus have the exceptionally high capability for producing cellulose (Ryngajłło et al., 2019).

However, the productivity of nata-de-coco by these bacteria is highly unstable; AAB are well known to be genetically unstable and to lose cellulose producing ability frequently in the absence of selection pressures (Matsutani et al., 2015; Ryngajłło et al., 2019). The genetic instability of AAB is probably attributable to the complexity of genome organization. Some examples reported earlier in the literature have shown the presence of more numerous mobile genetic elements and of duplicated sequences than in ordinary bacteria (Azuma et al., 2009; Zhang et al., 2017; Ryngajłło et al., 2019).

The *Komagataeibacter medellinensis* NBRC 3288 strain (formerly, *Gluconacetobacter xylinus* NBRC 3288) was isolated from rice vinegar as a cellulose-producing strain in 1954, but complete loss of its production capability was reported in 1999 (Navarro and Komagata, 1999). Then, Matsutani et al. (2015) found crucially important gene truncation in the cellulose biosynthesis operon (*bcs* operon) of this strain in 2015. In *Komagataeibacter* bacteria, the *bcs* operon is typically organized as *bcsZHABCD-bglX* or *bcsABXYC*. The genes in these operons encode cyclic dimeric ($3' \rightarrow 5'$) guanosine monophosphate (c-di-GMP) -regulated cellulose synthase (BcsA and BcsB; inner membrane-associated), an outer membrane porin protein (BcsC), putative periplasmic protein (BcsD), cellulose-complementing protein A (BcsH), probable cellulose deacylase (BcsX), probable cellulose acylase (BcsY), endo-β-1,4-glucanase (BcsZ), and β-glucosidase (bglX). Moreover, the genes play a central role in cellulose biosynthesis (Römling and Galperin, 2015).

To elucidate the genomic characterization of nata-de-coco producing bacteria, we determined the whole genome sequences of three bacterial strains (FNDCR1, FNDCF1, FNDCR2) that had been isolated from a practical culture of nata-de-coco in Japan. This report is the first of a study of whole genome sequences of practical nata-de-coco producing bacteria. All of the strains were bacteria of *Komagataeibacter* genus. These data support better understanding of the physiology of *Komagataeibacter* bacteria and elucidate the efficient and stable production of nata-de-coco and biocellulose.

MATERIALS AND METHODS

Bacterial Strains and General Techniques

The nata-de-coco producing bacteria were isolated from a nata-de-coco production facility in Japan. The FNDCR1 strain was stocked in 2017 as a strain that had shown good nata-de-coco productivity from those stored in a nata-de-coco production facility in Japan. The FNDCF1 strain was isolated in 2016 when the productivity loss occurred in the facility. FNDCR2 is a strain that was isolated in 2017 after 20 passages of culture from a stock solution stored at the facility for producing nata-de-coco. In this study, all strains were single colony isolated in 2020, followed by cellulose productivity experiments and whole genome sequencing. Unless stated otherwise, cells were cultured at 28°C using Hestrin–Schramm (HS) medium: 5 g/L yeast extract, 5 g/L peptone, 2.7 g/L Na_2HPO_4, 1.5 g/L citric acid, 0.1% (v/v) acetic acid, and 10 g/L sucrose or glucose. For preparation of HS-agar plates, agarose was added at 20 g/L.

Cellulose Production and Quantification

After fermentation, the cellulose amount was quantified according to the method described in an earlier report, with modifications (Ryngajłło et al., 2019). In brief, cells were spread onto an agar plate (in a 90-mm-diameter Petri dish) from a glycerol stock. After incubating the plate for 4 days, 2 mL of liquid medium was added onto the surface. The cells were scraped and collected. After the optical density of the cell suspension at 600 nm was adjusted to 0.2, 500 μL of the suspension was added to an Erlenmeyer flask containing 9.5 mL of fresh medium. The culture was incubated for 5 days without shaking. The cellulose pellicle was removed from the flask, treated with 1 M NaOH solution overnight in a Petri dish, and washed extensively with distilled water until the water pH became neutral. The washed cellulose was moved to a new and clean Petri dish, of which the weight was pre-recorded. The cellulose was dried completely, typically for 4 h, at 60°C. The Petri-dish and cellulose total weight was recorded.

Whole-Genome Sequencing

Whole-genome DNA was isolated using the neutral phenol—chloroform—isoamyl alcohol method (Nakashima and Tamura, 2018) with one modification: cellulase (Cellulase, from *Aspergillus niger*; Fujifilm Wako Pure Chemical Corp., Osaka, Japan) was added at 1 mg/mL and incubated for 1 h at 28°C before harvesting. The purified DNA was subjected to sequencing with PacBio RS II and Illumina Hiseq 2,500 systems (Hokkaido System Science Co., Ltd., Sapporo, Japan). The Hiseq reads were obtained with paired-end 100 base pairs (bp) sequencing. The numbers of reads over ≥ Q30 were 47,239,534, 47,458,318, and 44,626,206 bp, respectively, for FNDCR1, FNDCF1, and FNDR2 strains. In the PacBio reads, the numbers of subreads and mean subread lengths were 145,440 and 6,655 bp, 149,926 and 7,350 bp, and 175,664 and 9,362 bp, respectively, for FNDCR1, FNDCF1, and FNDR2 strains. The raw sequencing reads have been deposited in NCBI SRA under their accession numbers SRR15903111-SRR15903116.

The Hiseq reads were trimmed before assembly (TrimGalore! v 0.4.4).[1] Both the trimmed Hiseq and PacBio reads were used to construct the genomes using the Unicycler software (v.

[1]https://github.com/FelixKrueger/TrimGalore

0.4.6; Wick et al., 2017) with default parameters. The assembly genome completeness was assessed using BUSCO (v.5.2.2; rhodospirillales_odb10) (Manni et al., 2021). The assembled genomes were classified into plasmids and chromosomes using PlasFlow (v. 1.1.0; Krawczyk et al., 2018) with default parameters, respectively. Genome annotation was performed using the prokka (v1.13.7) software (Seemann, 2014) and web-based RAST (v.2.0; Aziz et al., 2008)[2] with preset parameters, respectively. The CGview server beta was used to illustrate the genome maps[3] (Grant and Stothard, 2008; Grant et al., 2012). The assembled sequences were deposited in the NCBI Genbank under the BioProject number PRJNA763499.

Phylogenetic Analysis

From the assembled genomes, 16S rRNA gene sequences were extracted (barrnap v.0.9).[4] Phylogenetic classification was then performed from the 16S rRNA gene sequences using the Alignment, Classification and Tree Service[5] of SILVA (Release 138; Quast et al., 2013).

Based on the classification results, the public genome of an assigned genus (*Komagataeibacter* genus) was obtained from NCBI Genbank[6] using NCBI-genome-download v.0.3.0.[7] To ensure the genome assembly levels, the downloaded genomes were limited to "complete" whole genomes.

To identify the core gene sets from three new assembly genomes and downloaded genomes, a pan-genomic analysis was performed (Roary v.3.11.2; Page et al., 2015). Using these core gene sets, a maximum-likelihood phylogenetic tree was constructed with IQ-TREE v.2.0.3 (bootstrap: 1,000 replicates; best-fitting substitution model: GTR + F + R6 model) (Nguyen et al., 2015). The phylogenetic tree was visualized using iTOL (v.6.3)[8] (Letunic and Bork, 2021).

Identification of Synteny Blocks and Duplicated Regions

To detect synteny blocks, the three assembled genomes were subjected to pairwise comparisons to themselves and other genomes (Sibelia v.3.0.7; Minkin and Medvedev, 2020). The detected synteny blocks were visualized as paths between genomes (Circa v.1.2.2; OMGenomics). Duplicate regions within the genome were sought (RepSeek v.6.6; Achaz et al., 2006) with minimum length thresholds of 1,000, 2,000, and 3,000 bp. Moreover, correlation tests and linear regression analyses were conducted using implemented functions in R language (R v.4.1.1; R Development Team, 2021) to assess the relation between the duplicated lengths and the number of mobile genetic elements. The R script are available on https://github.com/omics-tools/natagenome_rscript.

[2]https://rast.nmpdr.org/

[3]http://cgview.ca

[4]https://github.com/tseemann/barrnap

[5]https://www.arb-silva.de/aligner/

[6]https://www.ncbi.nlm.nih.gov/genbank/

[7]https://github.com/kblin/ncbi-genome-download

[8]https://itol.embl.de/

RESULTS

Validation of Cellulose Productivity

Cellulose productivity of three strains was quantified as described in section "Materials and Methods" (**Figure 1**). As carbon sources, glucose or sucrose was fed to the cells. The FNDCR2 strain produced more cellulose than the FNDCR1 strain, although the FNDCF1 strain did not produce at all. This result indicates that the nata-de-coco producing bacterial culture contained both cellulose producing and non-producing cells.

Whole Genome Sequence Determination

To identify the taxonomy and to investigate cellulose biosynthesis mechanisms of the present strains, the genome sequences were obtained with shotgun sequencing. For constructing complete genome sequences, both short-read and long-read sequencing were performed. After sequencing, the whole genome sequences were constructed with hybrid assembly (section "Materials and Methods"). The assembled genomes of FNDCR1, FNDCF1, and FNDCR2, respectively, showed high completeness of 99.5, 99.6, and 99.7%. All the strains were found to harbor multiple circular plasmids in addition to circular chromosomes (**Figures 2–4**). There was no shared plasmid in these three strains (**Supplementary Figure 1**). The relative value of sequencing coverage of each plasmid is expected to reflect the plasmid copy number. Two CRISPR sequences were found on the pKFR1-1 plasmid.

Phylogenetic Relation Analysis

Each strain harbored five 16 rRNA genes on the chromosome. All 16S rRNA sequences of the FNDCR1, FNDCF1, and FNDCR2 strains were assigned to those of *Komagataeibacter* with > 99% similarities (**Supplementary Table 1**). The phylogenetic tree constructed based on the core gene set derived from the pan-genome analysis of *Komagateibacter* (**Figure 5**) shows that the non-cellulose synthesizing type, FNDCF1, is in a clade close to the root in the phylogenetic tree of *Komagateibacter* bacteria. This strain is close to the *Komagataeibacter kakiaceti*

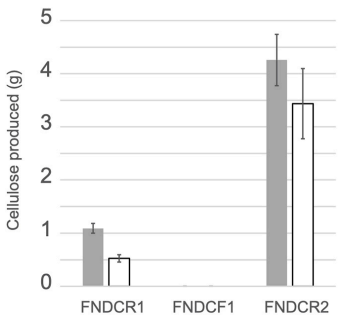


FIGURE 1 | Cellulose productivity of the present strains. Cellulose was produced from glucose (gray bars) or sucrose (open bars) using the designated strains. Quantification results of cellulose from three replicate experiments are shown as mean ± SEM.

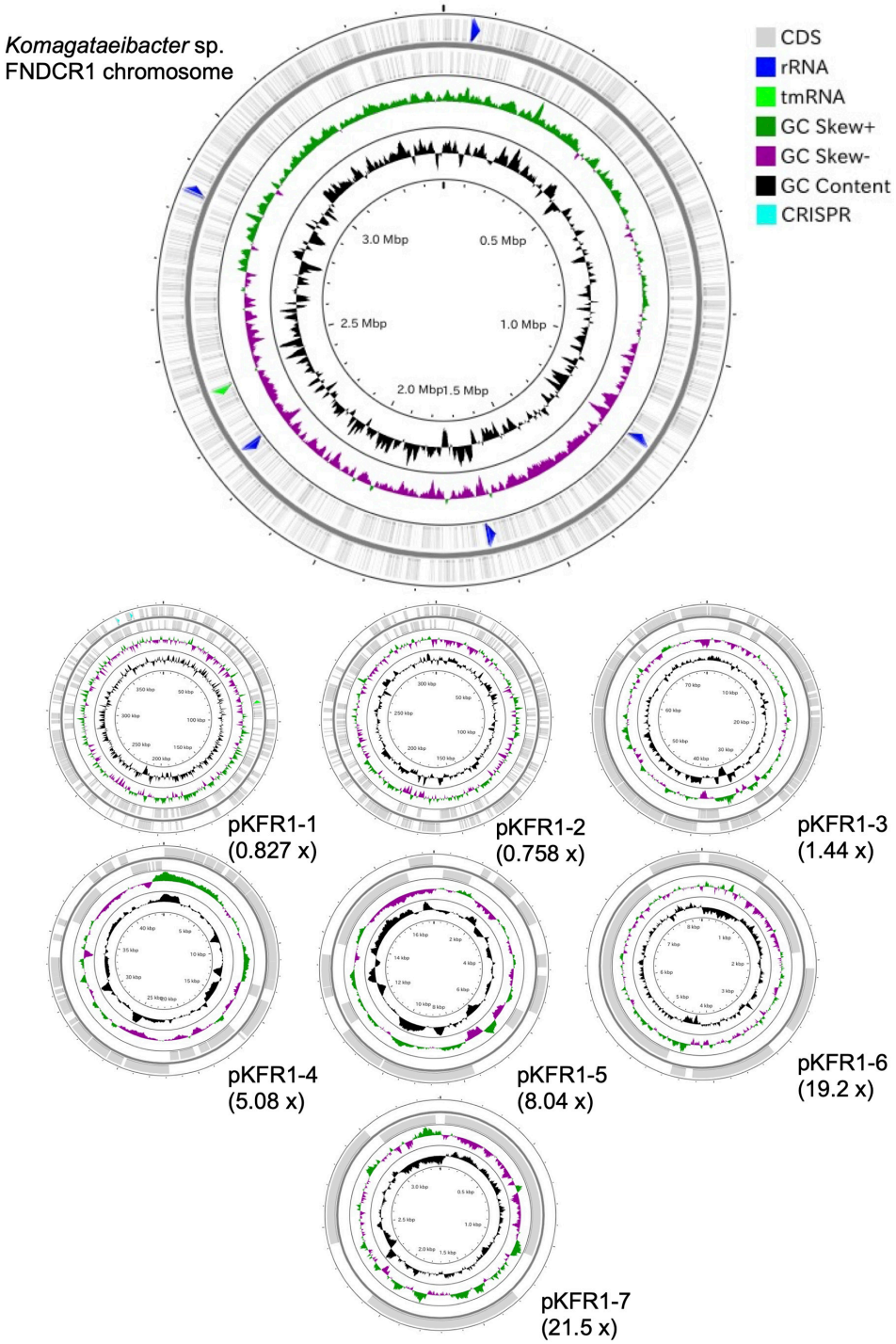


FIGURE 2 | Genome structure of the FNDCF1 strain. One circular chromosome and seven putative plasmids (pKFR1-1, pKFR1-2, pKFR1-3, pKFR1-4, pKFR1-5, pKFR1-6, and pKFR1-7) are shown. The value presented in parentheses represents the sequencing coverage of a plasmid relative to that of chromosome. Smaller contigs below 5,000 bp and non-circularized contigs are not shown in this figure.

JCM25156 strain isolated from traditional Japanese fruit vinegar (Iino et al., 2012). The cellulose-producing type FNDCR1 is located close to the vinegar-producing acetic acid bacterium *Komagataeibacter maltaceti* LMG 1529 (Zhang et al., 2018). The cellulose-producing type FNDCR2 is represented as a single clade, with the larger clade being closer to the *Komagataeibacter rhaeticus* lineage. However, none of the present genomes is the same as known strain genomes. These results indicate that all the

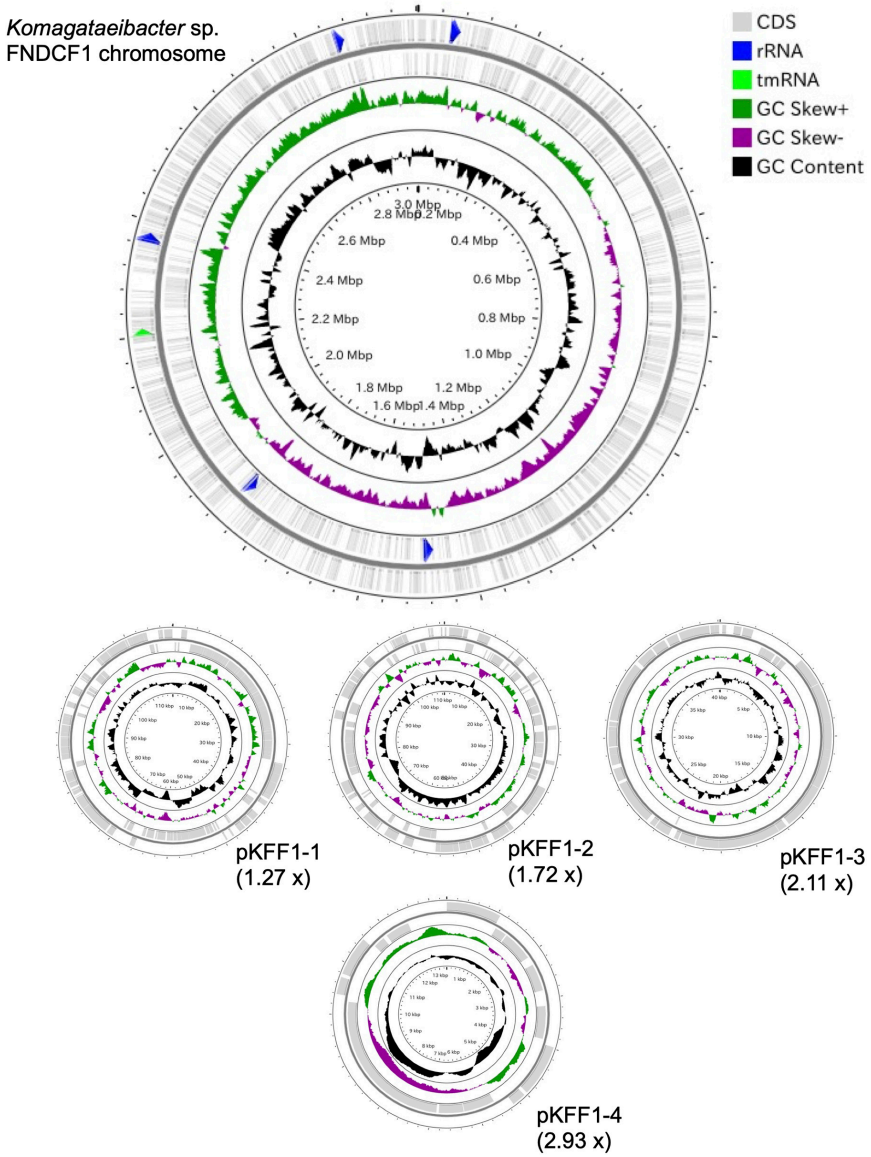


FIGURE 3 | Genome structure of the FNDCR1 strain. One circular chromosome and four putative plasmids (pKFF1-1, pKFF1-2, pKFF1-3, and pKFF1-4 pKFR1-7) are shown. The value given in parentheses represents the sequencing coverage of a plasmid relative to that of chromosome. Smaller contigs below 5,000 bp and non-circularized contigs are not shown in this figure.

present strains belong to *Komagataeibacter* genus, but all might be novel strain genomes.

Analysis of Genome Structure and Genome Complexity in Cellulose Producing Strains

To compare the genome structures of the respective strains, the three assembled genomes were searched for synteny blocks between themselves and other genomes. Although homologous synteny blocks were found between the three genomes, the overall genome structure did not closely match each other (**Figure 6**). Results show that the plasmids in each strain harbor regions with DNA homologous to the chromosome (**Figure 7**). For the plasmids in the FNDCR1 and FNDCR2 strains, many homologous regions were found in their chromosomes, although few such regions were found in the FNDCF1 chromosome.

For the whole genomes in the FNDCR1 and FNDCR2 strains, many duplicated sequences were observed, but these characteristic structures were not found in FNDCF1 strain (**Figures 5**, **8**). Furthermore, to investigate relationships between these duplications and mobile genetic elements, the comparative genome analysis was performed with published other *Komagataeibacter* genomes. Detected duplicates with minimum length of 1,000 bp were significantly and positively correlated with the number of mobile genetic elements in

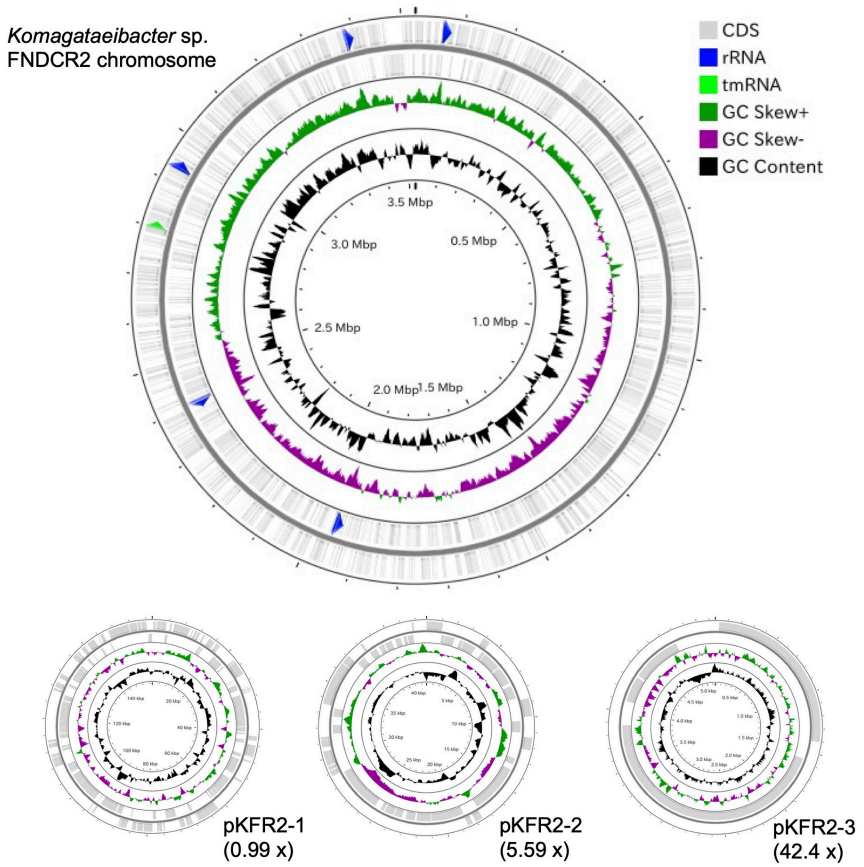


FIGURE 4 | Genome structure of the FNDCR2 strain. One circular chromosome and three putative plasmids (pKFR2-1, pKFR2-2, pKFR2-3) are shown. The value given in parentheses represents the sequencing coverage of a plasmid relative to that of chromosome. Smaller contigs below 5,000 bp and non-circularized contigs are not shown in this figure.

the genome. This relation was more apparent for plasmids than for chromosomes (**Figure 8**). However, for duplicates with 2,000 bp or greater length, no positive correlation was found either for chromosomes or for plasmids. These results suggest that mobile genetic elements can be associated with the transposition of sequences less than 2,000 bp on their genomes. In addition, the assembled genome of FNDCR1 clearly has more duplicated sequences than the other strains, suggesting the existence of a strain-specific evolutionary event related with accelerated sequence duplication. The coding sequences on these repeats and their neighbors have been summarized as **Supplementary Table 2** and **Supplementary Figure 2**. As a result, about 30 and 19% of the repeated coding sequences in chromosomes and plasmids were annotated as genes related to mobile genetic elements, respectively (**Supplementary Figure 2**). This result supports the association between the number of mobile genetic elements in the genome and the number of repeats as shown in **Figures 5**, **8**. Most of the other repeated coding sequences were annotated as hypothetical proteins, and their functions and effects are not well understood. Interestingly, some cellulose synthesis-related genes (cellulose synthase 1, cellulose synthase operon protein C and D) were also observed in the repeated and repeated-neighbor regions in chromosomes (**Supplementary Table 2**). Transposons (ISs) were also observed in the neighboring region of the cellulose synthesis operon of the three strains in this study (**Figure 9**). These results suggest that mobile genetic elements in *Komagataeibacter* genomes may also affect the loss and gains of cellulose synthesis-related genes.

The sequence of each plasmid was compared to those of all other plasmids presented herein, as well as those of plasmids found previously in *K. xylinus* E25 strain (**Supplementary Figure 1**). Actually, pKFR1-2, pKFR2-1, and pGX5 (*K. xylinus* E25) were mutually similar. Also, pKFR1-4 was similar to pKFR2-2 throughout the plasmids. Moreover, pKFF1-4, pKFR1-4, pKFR2-1, pGX4 (*K. xylinus* E25), and pGX3 (*K. xylinus* E25) shared partial homology: an approximately 3 kb part of pKFR1-7 (total length of 3.4 kb) was highly homologous to the part of pKFR1-6 and pGX1 (*K. xylinus* E25). In addition, a consensus sequence of approximately 3.7 kb in pKFR1-1, pKFR1-2, pKFR1-3, pKFR1-4, pKFR1-5, pKFR1-6, pKFR2-1, and pKFR2-2 was identified. On this sequence, only transposase and resolvase were coded. Therefore, the same transposable elements were present on these plasmids.

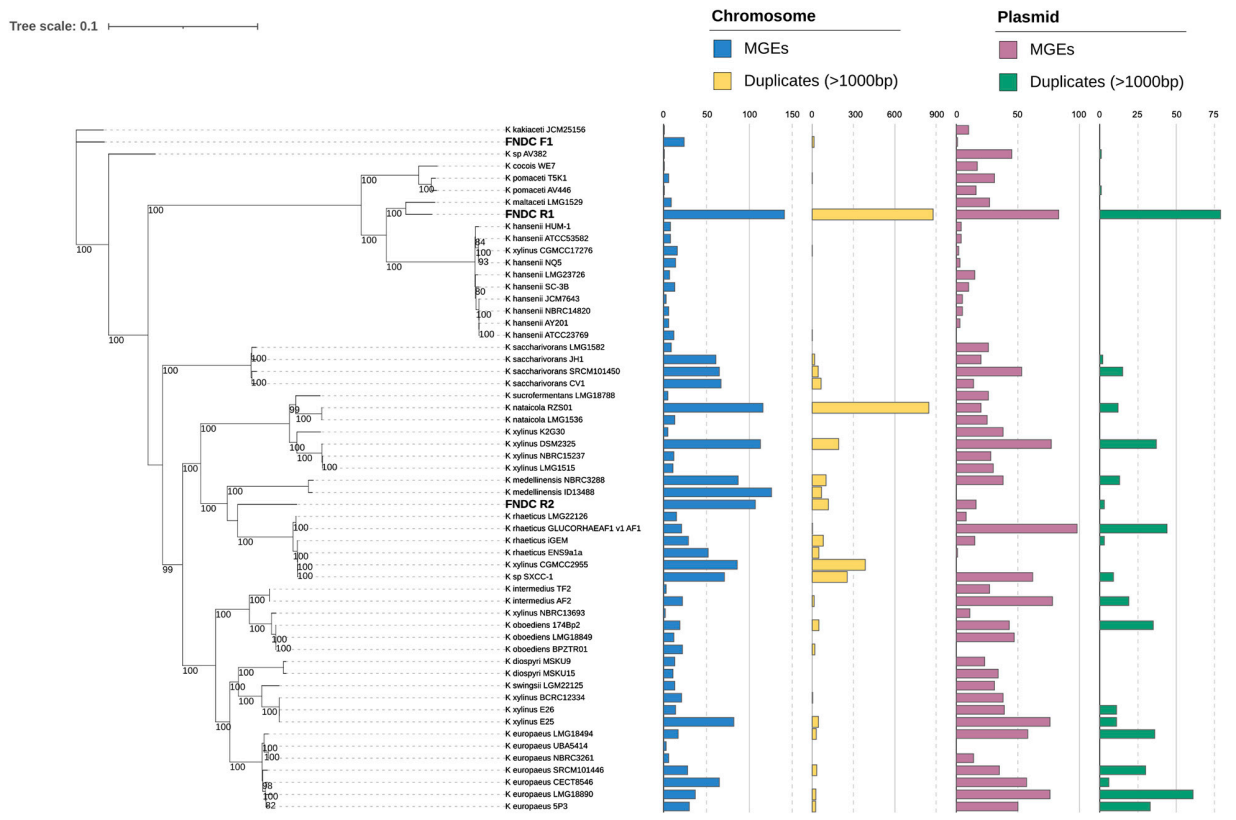


FIGURE 5 | Maximum-likelihood phylogenetic tree in *Komagataeibacter* genus. A maximum-likelihood phylogenetic tree was built using IQ-TREE based on the public *Komagataeibacter* genomes and the present strain genomes (FNDCR1, FNDCF1, and FNDCR2). The numbers at nodes are the bootstrap values > 80% as a percentage of 1,000 replications. The bar plot on the right shows the number of mobile genetic elements (MGEs) and Duplicates > 1,000 bp, respectively, on the chromosome and plasmid. The term "Duplicates" refers to homologous sequence pairs in the genome, as detected by RepSeek. MGEs and Duplicates on chromosomes, respectively, correspond to cyan and yellow, whereas MGEs and Duplicates on plasmids, respectively, correspond to pink and green.

Taking these findings together, one can infer that the genome structure of FNDCR1 and FNDCR2 strains are complex, leading us to predict that the genomes of these strains are genetically unstable.

Structure of the Cellulose Biosynthesis Operon

For cellulose biosynthesis, the *bcsABCD* genes are the most important genes: with other accessory genes, they are involved in efficient biosynthesis (Römling and Galperin, 2015). We manually extracted information related to *bcs* and other genes from the gff3 annotation file created using the prokka and RAST programs (**Figure 9**). It is noteworthy that many strains of *Komagataeibacte*r bacteria harbor two *bcs* operons (type I and II). Typically, the type I and II operons, respectively, comprise *bcsZ-bcsH-bcsAI-bcsBI-bcsCI-bcsDI-bglX*, and *bcsAII-bcsBII-bcsX-bcsY-bcsCII* (the *bcsA* and *bcsB* genes are sometimes found as a fused gene, *bcsAB*) (Römling and Galperin, 2015; Ryngajłło et al., 2019).

The FNDCR1 and FNDCR2 strains harbored operons of both types as intact, however, only one *bcs* operon was found in the FNDCF1 strain. The sole *bcs* operon of the FNDCF1 strain was a fused one of type I and II operons, respectively. Furthermore, no *bcsCI*, *bcsAII*, or *bcsBII* gene was found. A premature stop codon appeared in the C-terminal part of *bcsCII* in the FNDCF1 strain (**Figure 9**). The loss of these genes can have affected the function of guanosine monophosphate (c-di-GMP)-regulated cellulose synthase and outer membrane porin proteins in the FNDCF1 strain. This operon structure explains why this strain does not produce cellulose at all.

We also found that the genes for Tiorf138 protein, hypothetical protein of Cupin superfamily, hydrolase of NUDIX family, and chaperone protein HtpG were located immediately downstream of *bcsDI* in all the presented strains. These four genes were also located similarly on the chromosomes of *Komagataeibacter* spp. E25, NBRC3288, and ATCC23769 strains (**Supplementary Table 3**), implying possible involvement of these genes in cellulose biosynthesis. The "Discussion" section provides additional details.

DISCUSSION

The genetic instability of *Komagataeibacter* bacteria at least partly derives from the presence of the exceptionally huge

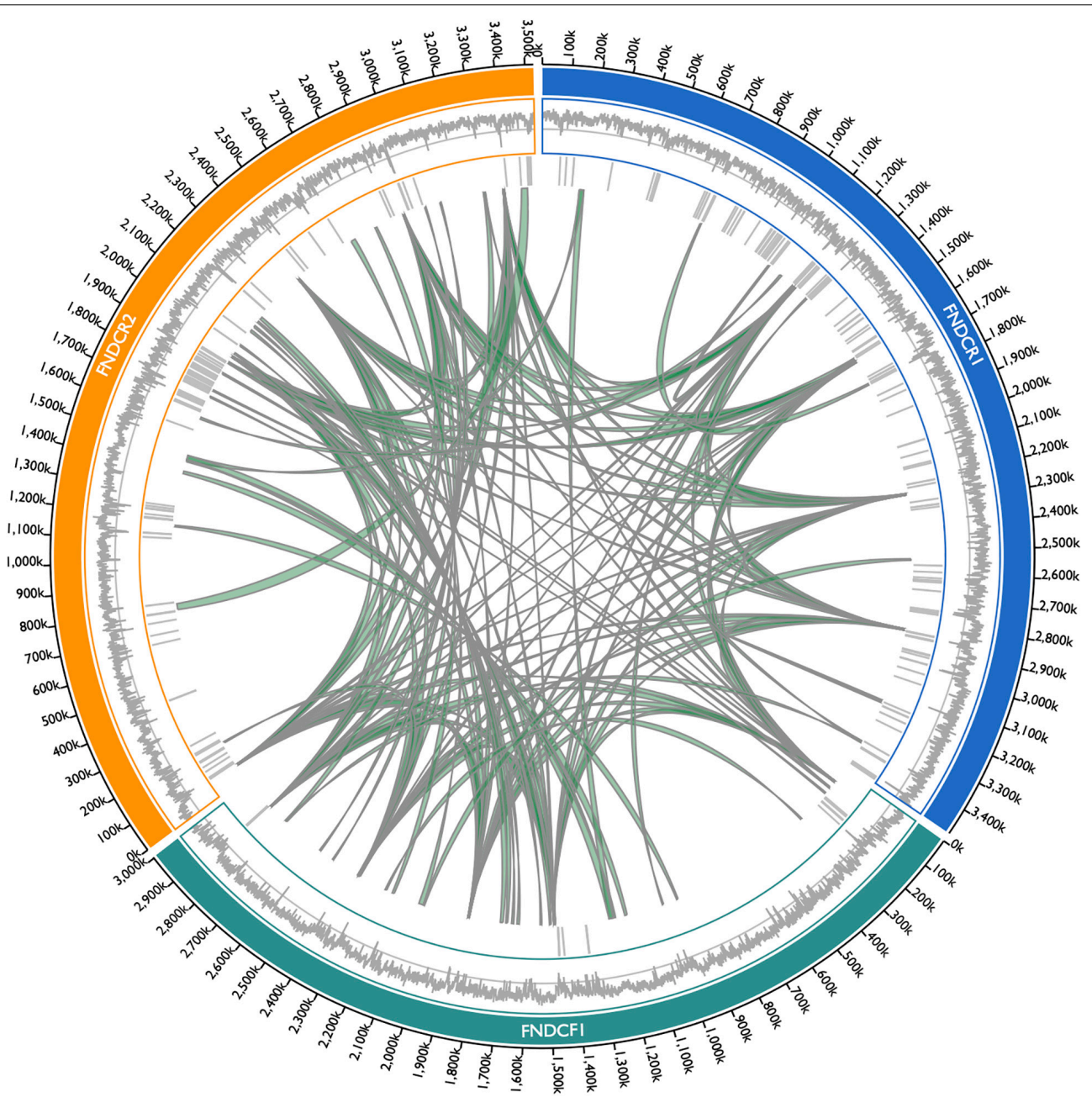


FIGURE 6 | Synteny blocks among the present strain chromosomes. Synteny blocks > 5,000 bp within/among chromosome genomes of FNDCR1 (cyan), FNDCF1 (light green), and FNDCR2 (orange) are shown as gray-colored paths. Synteny blocks were detected using Siberia. The barcode plot outside the synteny block shows where the MGEs are located in the genome. The outer line plot shows the GC content in the genome. The middle line shows the 50% GC content. Numbers on the outside represent the chromosome positions.

number of mobile genetic elements on the genomes (Azuma et al., 2009; Zhang et al., 2017). Furthermore, high degrees of gene duplication, hyper-mutability, and complex plasmid contents have been reported for many *Komagataeibacter* bacteria (Römling and Galperin, 2015). The present study revealed a similar complex genomic structure of nata-de-coco producing *Komagataeibacter* bacteria. These findings show many truncations or insertions of ISs within the *bcs* operon genes of the present and other *Komagataeibacter* strains (**Supplementary Table 3**). These genetic variations are inferred to be the result of genomic structure complexity. Additionally, this report is the first of a positive correlation found between the number of mobile genetic elements and the number of duplicated sequences on the genomes (**Figure 8**) and the first describing that not all genomes of *Komagataeibacter* strains are complex (**Figures 7**, **8**; the FNDCF1 strain had the simple genome).

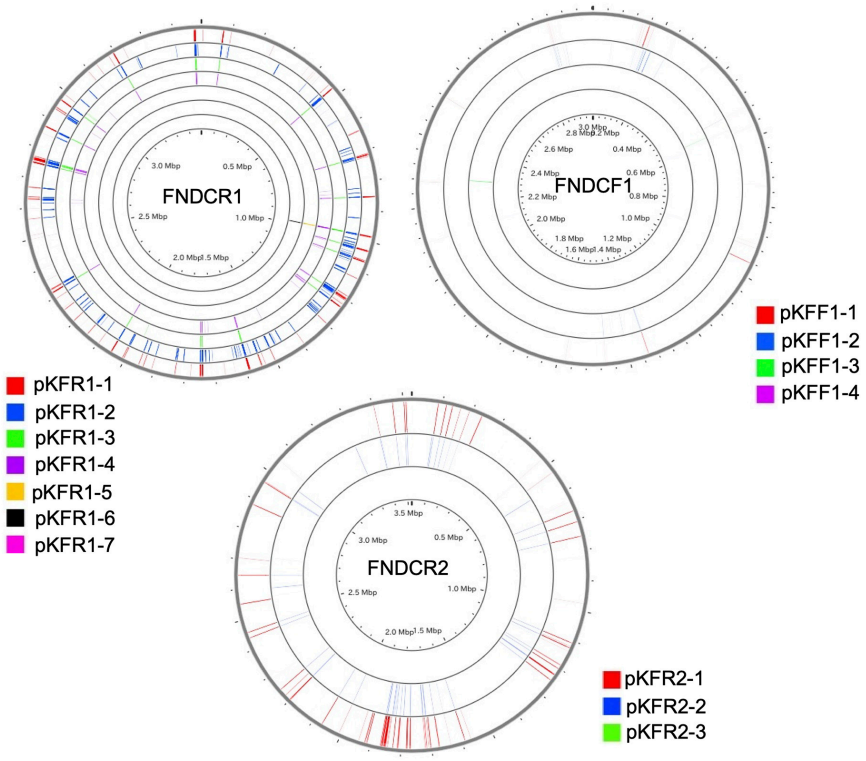


FIGURE 7 | Sequences shared with the chromosome and the plasmids. Homology search (blast algorithm) between each chromosome and plasmids in each strain (other circles) was performed on the CGview server beta. Each plasmid sequence was used as a query. Homologous sequences found on the chromosome are marked with colored lines.

The *bcsCII* gene of the FNDCF1 strain had a premature stop codon. Some *bcs* genes were not found on the genome (**Figure 9**). In the *Komagataeibacter medellibensis* NBRC3288 strain, the loss and restoration of cellulose productivity were observed according to a loss-of-function mutation (frameshift mutation) and intragenic suppressor mutations (reversion of frameshift) in the *bcsBI* gene (Matsutani et al., 2015). Furthermore, *Komagateeibacter europaeus* LMG18890 and LMG18494 strains are non-cellulose producers. They have some impairments in either *bcsBI* or *bcsCI* genes (Matsutani et al., 2015). A non-sense mutation in the *bcsCI* gene of *Komagataeibacter oboediens* MSKU3 has been reported to engender loss of cellulose productivity (Taweecheep et al., 2019). Therefore, we infer that the loss of cellulose productivity in the FNDCF1 strain is attributed to loss of a full functional set of the *bcs* operon.

Expression from the *bcs* operon in *Komagataeibacter* bacteria is known to be highly stimulated by c-di-GMP (Zhang et al., 2017). Therefore, the intracellular level of c-di-GMP is important for cellulose production. However, the FNDCF1 strain harbored five intact diguanylate cyclase/phosphodiesterase genes that are necessary to synthesize c-di-GMP, suggesting that the c-di-GMP level is normal in this strain. Similarly, the FNDCR1 and FNDCR2 strains, respectively, harbored five and two diguanylate cyclase/phosphodiesterase genes.

The genes for Tiorf138 protein, hypothetical protein of Cupin superfamily, hydrolase of NUDIX family, and chaperone protein HtpG are located immediately downstream of the *bcs* operon in many *Komagataeibacter* strains (**Figure 9** and **Supplementary Table 3**). Hydrolases of NUDIX (nucleoside diphosphates-linked moiety-X) family hydrolyzes the pyrophosphate linkage in various nucleoside diphosphates linked to another moiety X (NDP-X) such as pyridine nucleotides, coenzyme A, dinucleoside polyphosphates, and nucleotide sugars (Fisher et al., 2002). Uridine-5′-diphosphate-α-D-glucose (UDP-glucose), a substrate for cellulose polymerization in bacteria (Ryngajłło et al., 2019), is a kind of NDP-X (a nucleotide sugar). In the case of glycogen biosynthesis in which UDP-glucose is a polymerization substrate, NUDIX hydrolases are thought to influence the flux of glucose into glycogen (Heyen et al., 2009). Therefore, the NUDIX hydrolases near the *bcs* operons might be involved in UDP-glucose degradation and might regulate cellulose biosynthesis. In this respect, determining the substrate specificity of those NUDIX hydrolases is important. Actually, HtpG, a chaperon protein of hsp90 family, is known to interact with proteins containing tetratrico peptide repeat (TPR) domain (Scheufler et al., 2000). In fact, because the N-terminal portion of BcsC contains 17 TPR, HtpG might interact with BcsC (Du et al., 2016; Nojima et al., 2017). Additionally, in *Pseudomonas aeruginosa*, results demonstrated that low intracellular levels of c-di-GMP increase the intracellular HtpG level (Chua et al., 2013).

In the phylogenetic classification based on 16S rRNA genes, all three isolated strains were assigned to *Komagataeibacter*

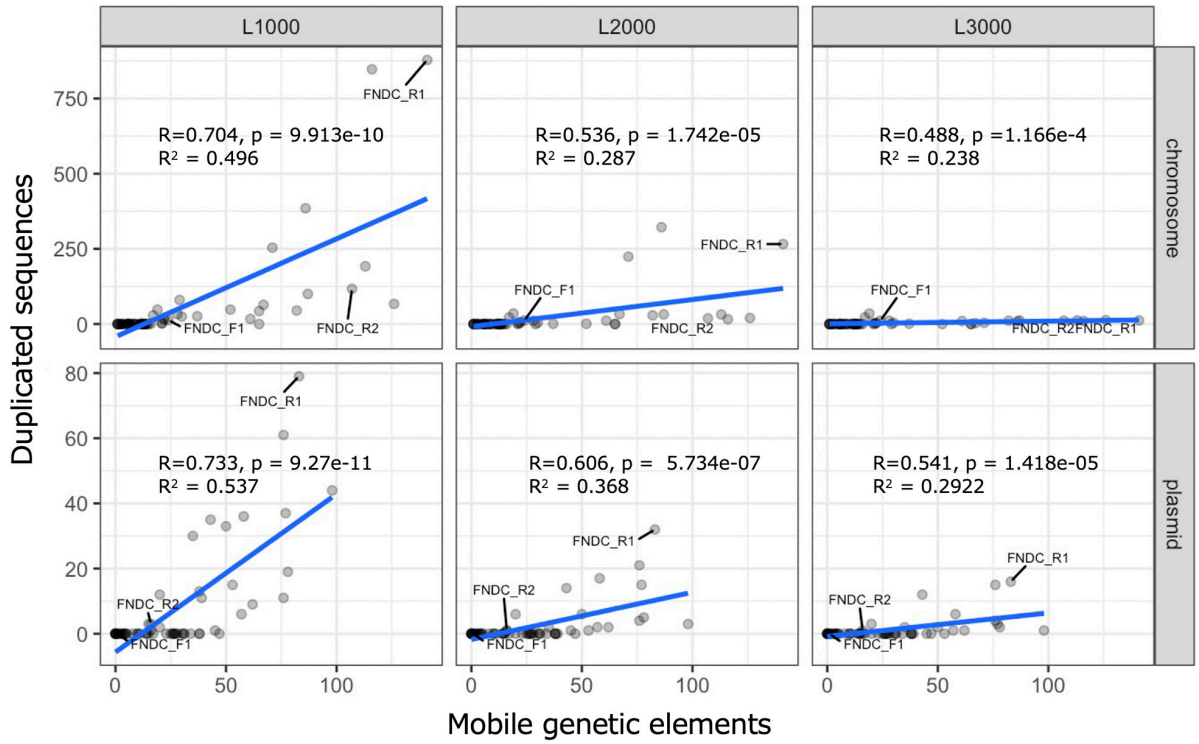


FIGURE 8 | Scatter plots for duplicated sequences and mobile genetic elements. These scatter plots show results of correlation and regression analysis of the number of duplicated sequences and mobile genetic elements in the genome. Correlation analysis and regression analysis were performed with functions implemented in R. The top and bottom rows, respectively, present results for chromosomes and plasmids. Columns L1000, L2000, and L3000, respectively, correspond to the conditions that set 1,000, 2,000, and 3,000 bp as the minimum length threshold during the duplicate sequence search. The line written in blue is the regression line. *R*, *p*, and R^2, respectively, denote the correlation coefficient, the *p*-value of the uncorrelation test, and the multiple correlation coefficient.

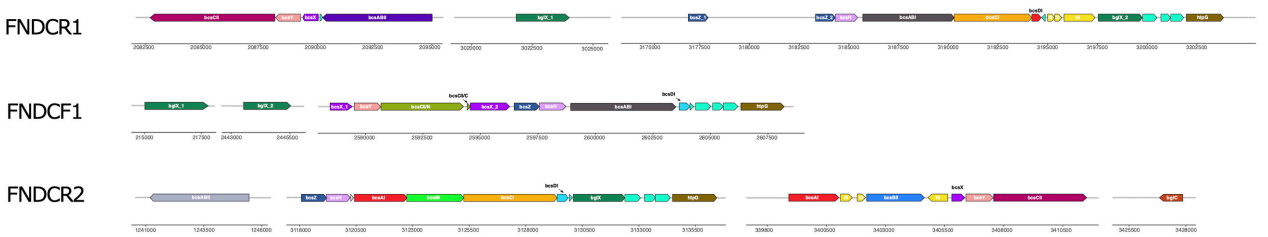


FIGURE 9 | Cellulose biosynthesis operons of the present strains. The figure portrays the operon structure for genes involved in cellulose biosynthesis. The structure was drawn based on results obtained for gene annotation (**Supplementary Table 1**). Rectangles with arrowheads represent regions that encode genes, open reading frames (ORFs), or transposons (ISs). The arrowhead direction shows the DNA strand orientation (forward/reverse). The numbers represent the genomic positions in the chromosomes. Colors correspond to the encoded genes or ISs. Those without names are the predicted ORFs.

genus. Phylogenetic tree analysis of the conserved core genes indicated that the three strains belong to mutually different clades. None of them matches any known strain. This finding suggests that the present strains are novel strains that have not been reported in the literature as *Komagataeibacter* strains. Duplicated genome regions as their genomic signature were also specifically examined. In fact, FNDCR1 and FNDCR2 have many duplicated sequences within their own genomes. By contrast, FNDCF1 has only a few duplicates in its own genome. It is particularly interesting that FNDCF1 has far fewer mobile genetic elements in its genome than the other two cellulose-producing lineages (FNDCR1 and FNDCR2), which might be attributable to the cross-strain variation in the sequence duplication among *Komagataeibacter* strains. After applying correlation analysis between the length of these homologous duplicated sequences and the number of mobile genetic elements, duplicates with minimum length of 1,000 bp were found to be positively correlated with the number of mobile genetic elements. Nevertheless, no correlation was found for duplicates with minimum length longer than 2,000 bp. This finding suggests that mobile genetic elements are related with transposition in the

genome, but they might be subject to negative selection in their duplicated sequence lengths.

Our initial motivation for this study was to understand the genetic instability of nata-de-coco producing bacteria and stabilization of nata-de-coco production yields. For industrial production of acetic acid, *Acetobacter pasteurianus* 386B strain is exploited frequently. Its low number of mobile genetic elements in the genome and absence of complete phage genomes makes this strain more genetically stable than other *A. pasteurianus* strains (Illeghems et al., 2013; Ryngajłło et al., 2019). From results of the comparative genomics study presented herein, although FNDCF1 lacks the ability to produce cellulose, it can be expected to have less influence of mobile factors on genome stability. Given these findings, after the *bcs* operon of FNDCR1 or FNDCR2 strains is genetically transferred to the chromosome of FNDCF1 strain, it can be anticipated for use in stable industrial nata-de-coco production. To this end, useful genetic engineering technologies including genome editing should be developed for *Komagataeibacter* bacteria.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA763499, PRJNA763499.

AUTHOR CONTRIBUTIONS

KI and NN designed the study, collected the data, conducted data analyses, and drafted the manuscript. HK, TI, and KK reviewed the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

We thank Masako Sato, Ayane Osanai, and Miyako Hata for their assistance. We are also grateful to members of our research group for their help and valuable discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.798010/full#supplementary-material

REFERENCES

Achaz, G., Boyer, F., Rocha, E. P. C., Viari, A., and Coissac, E. (2006). Repseek, a tool to retrieve approximate repeats from large DNA sequences. *Bioinformatics* 23:119121. doi: 10.1093/bioinformatics/btl519

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Azuma, Y., Hosoyama, A., Matsutani, M., Furuya, N., Horikawa, H., Harada, T., et al. (2009). Whole-genome analyses reveal genetic instability of *Acetobacter pasteurianus*. *Nucleic Acids Res.* 37:57685783. doi: 10.1093/nar/gkp612

Chawla, R. P., Bajaj, B. I., Survase, A. S., and Singhal, S. R. (2009). Microbial cellulose: fermentative production and applications. *Food Technol. Biotechnol.* 2:107124.

Chua, S. L., Tan, S. Y.-Y., Rybtke, M. T., Chen, Y., Rice, S. A., Kjelleberg, S., et al. (2013). Bis-(3′-5′)-Cyclic dimeric GMP regulates antimicrobial peptide resistance in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* 57:20662075. doi: 10.1128/aac.02499-12

Du, J., Vepachedu, V., Cho, S. H., Kumar, M., and Nixon, B. T. (2016). Structure of the cellulose synthase complex of *Gluconacetobacter hansenii* at 23.4 Å Resolution. *PLoS One* 11:e0155886. doi: 10.1371/journal.pone.0155886

Fisher, D. I., Safrany, S. T., Strike, P., McLennan, A. G., and Cartwright, J. L. (2002). Nudix hydrolases that degrade dinucleoside and diphosphoinositol polyphosphates also have 5-phosphoribosyl 1-pyrophosphate (prpp) pyrophosphatase activity that generates the glycolytic activator ribose 1,5-bisphosphate. *J. Biol. Chem.* 277:4731347317. doi: 10.1074/jbc.m209795200

Grant, J. R., and Stothard, P. (2008). The CGView server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.* 36:W181W184. doi: 10.1093/nar/gkn179

Grant, J. R., Arantes, A. S., and Stothard, P. (2012). Comparing thousands of circular genomes using the CGView Comparison Tool. *BMC Genomics* 13:202. doi: 10.1186/1471-2164-13-202

Heyen, C. A., Tagliabracci, V. S., Zhai, L., and Roach, P. J. (2009). Characterization of mouse UDP-glucose pyrophosphatase, a Nudix hydrolase encoded by the Nudt14 gene. *Biochem. Biophys. Res. Commun.* 390:14141418. doi: 10.1016/j.bbrc.2009.11.007

Iino, T., Suzuki, R., Tanaka, N., Kosako, Y., Ohkuma, M., Komagata, K., et al. (2012). *Gluconacetobacter kakiaceti* sp. nov., an acetic acid bacterium isolated from a traditional Japanese fruit vinegar. *Int. J. Syst. Evol. Microbiol.* 62:14651469. doi: 10.1099/ijs.0.031773-0

Illeghems, K., Vuyst, L. D., and Weckx, S. (2013). Complete genome sequence and comparative analysis of *Acetobacter pasteurianus* 386B, a strain well-adapted to the cocoa bean fermentation ecosystem. *BMC Genomics* 14:526. doi: 10.1186/1471-2164-14-526

Krawczyk, P. S., Lipinski, L., and Dziembowski, A. (2018). PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res.* 46:gkx1321. doi: 10.1093/nar/gkx1321

Lee, K., Buldum, G., Mantalaris, A., and Bismarck, A. (2014). More than meets the eye in bacterial cellulose: biosynthesis, bioprocessing, and applications in advanced fiber composites. *Macromol. Biosci.* 14:1032. doi: 10.1002/mabi.201300298

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293-W296. doi: 10.1093/nar/gkab301

Londrina, P. R., Gomes, R. J., Borges, M., de, F., Fortaleza, C. E., Rosa, M., et al. (2018). Acetic Acid Bacteria in the food industry: systematics, characteristics and applications. *Food Technol. Biotech.* 56:139151. doi: 10.17113/ftb.56.02.18.5593

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199

Matsutani, M., Ito, K., Azuma, Y., Ogino, H., Shirai, M., Yakushi, T., et al. (2015). Adaptive mutation related to cellulose producibility in *Komagataeibacter medellinensis* (*Gluconacetobacter xylinus*) NBRC 3288. *Appl. Microbiol. Biotechnol.* 99:72297240. doi: 10.1007/s00253-015-6598-x

Minkin, I., and Medvedev, P. (2020). Scalable multiple whole-genome alignment and locally collinear block construction with Sibelia Z. *Nat. Commun.* 11:6327. doi: 10.1038/s41467-020-19777-8

Nakashima, N., and Tamura, T. (2018). Whole-Genome Sequence of *Acetobacter orientalis* Strain FAN1, isolated from caucasian yogurt. *Genome Announc.* 6, e201–e218. doi: 10.1128/genomea.00201-18

Navarro, R. R., and Komagata, K. (1999). Differentiation of *Gluconacetobacter liquefaciens* and *Gluconacetobacter xylinus* based on DNA base composition, DNA relatedness, and oxidation products from glucose. *J. Gen. Appl. Microbiol.* 45:715. doi: 10.2323/jgam.45.7

Nguyen, L.-T., Schmidt, H. A., Haeseler, A., and von Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268274. doi: 10.1093/molbev/msu300

Nojima, S., Fujishima, A., Kato, K., Ohuchi, K., Shimizu, N., Yonezawa, K., et al. (2017). Crystal structure of the flexible tandem repeat domain of bacterial cellulose synthesis subunit C. *Sci. Rep.* 7:13018. doi: 10.1038/s41598-017-12530-0

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:36913693. doi: 10.1093/bioinformatics/btv421

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41:D590D596. doi: 10.1093/nar/gks1219

R Development Team (2021). *R: A Language and Environment for Statistical Computing*.

Römling, U., and Galperin, M. Y. (2015). Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends Microbiol.* 23:545557. doi: 10.1016/j.tim.2015.05.005

Ross, P., Mayer, R., and Benziman, M. (1991). Cellulose biosynthesis and function in bacteria. *Microbiol. Rev.* 55:3558. doi: 10.1128/mr.55.1.35-58.1991

Ryngajłło, M., Kubiak, K., Jêdrzejczak-Krzepkowska, M., Jacek, P., and Bielecki, S. (2019). Comparative genomics of the Komagataeibacter strains – Efficient bionanocellulose producers. *Microbiol. Open* 8:e00731. doi: 10.1002/mbo3.731

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., et al. (2000). Structure of TPR domain peptide complexes critical elements in the assembly of the Hsp70 Hsp90 multichaperone machine. *Cell* 101:199210. doi: 10.1016/s0092-8674(00)80830-2

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:20682069. doi: 10.1093/bioinformatics/btu153

Taweecheep, P., Naloka, K., Matsutani, M., Yakushi, T., Matsushita, K., and Theeragool, G. (2019). Superfine bacterial nanocellulose produced by reverse mutations in the bcsC gene during adaptive breeding of *Komagataeibacter oboediens*. *Carbohydr. Polym.* 226:115243. doi: 10.1016/j.carbpol.2019.115243

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Zhang, H., Xu, X., Chen, X., Yuan, F., Sun, B., Xu, Y., et al. (2017). Complete genome sequence of the cellulose-producing strain *Komagataeibacter nataicola* RZS01. *Sci. Rep.* 7:4431. doi: 10.1038/s41598-017-04589-6

Zhang, Q., Poehlein, A., Hollensteiner, J., and Daniel, R. (2018). Draft genome sequence of *Komagataeibacter maltaceti* LMG 1529T, a vinegar-producing acetic acid bacterium isolated from malt vinegar brewery acetifiers. *Genome Announc.* 6, e330–e418. doi: 10.1128/genomea.00330-18

Conflict of Interest: HK was employed by company Research and Development Department, Fujicco Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 March 2022
doi: 10.3389/fmicb.2022.856810

Unraveling the Genotypic and Phenotypic Diversity of the Psychrophilic *Clostridium estertheticum* Complex, a Meat Spoilage Agent

Joseph Wambui, Marc J. A. Stevens, Nicole Cernela and Roger Stephan*

Institute for Food Safety and Hygiene, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Caroline Barretto, Nestle, Switzerland

Reviewed by:
Monique Zagorec, INRA Centre Angers-Nantes Pays de la Loire, France
Ignacio Ferrés, Pasteur Institute of Montevideo, Uruguay

***Correspondence:**
Joseph Wambui
joseph.wambui@uzh.ch

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *17 January 2022*
Accepted: *04 March 2022*
Published: *28 March 2022*

Citation:
Wambui J, Stevens MJA, Cernela N and Stephan R (2022) Unraveling the Genotypic and Phenotypic Diversity of the Psychrophilic Clostridium estertheticum Complex, a Meat Spoilage Agent. Front. Microbiol. 13:856810. doi: 10.3389/fmicb.2022.856810

The spoilage of vacuum-packed meat by *Clostridium estertheticum* complex (CEC), which is accompanied by or without production of copious amounts of gas, has been linked to the acetone–butyrate–ethanol fermentation, but the mechanism behind the variable gas production has not been fully elucidated. The reconstruction and comparison of intra- and interspecies metabolic pathways linked to meat spoilage at the genomic level can unravel the genetic basis for the variable phenotype. However, this is hindered by unavailability of CEC genomes, which in addition, has hampered the determination of genetic diversity and its drivers within CEC. Therefore, the current study aimed at determining the diversity of CEC through comprehensive comparative genomics. Fifty CEC genomes from 11 CEC species were compared. Recombination and gene gain/loss events were identified as important sources of natural variation within CEC, with the latter being pronounced in genomospecies2 that has lost genes related to flagellar assembly and signaling. Pan-genome analysis revealed variations in carbohydrate metabolic and hydrogenases genes within the complex. Variable inter- and intraspecies gas production in meat by *C. estertheticum* and *Clostridium tagluense* were associated with the distribution of the [NiFe]-hydrogenase *hyp* gene cluster whose absence or presence was associated with occurrence or lack of pack distention, respectively. Through comparative genomics, we have shown CEC species exhibit high genetic diversity that can be partly attributed to recombination and gene gain/loss events. We have also shown genetic basis for variable gas production in meat can be attributed to the presence/absence of the *hyp* gene cluster.

Keywords: *Clostridium estertheticum*, *Clostridium tagluense*, comparative genomics, meat spoilage, hydrogenase, flagella, carbohydrates

INTRODUCTION

Clostridium estertheticum complex (CEC) comprises 10 closely related psychrophilic and anaerobic species (Palevich et al., 2021; Wambui et al., 2021a). Members of CEC have been isolated from different ecological niches including terrestrial and aquatic niches but, the majority of the species have been isolated from meat or meat processing facilities due to their involvement in meat spoilage (Collins, 1992; Spring, 2003; Suetin et al., 2009). In particular, *C. estertheticum* is the main causative agent of blown pack spoilage (BPS) of meat, which is characterized by pack distention arising from production of copious amounts of hydrogen and carbon dioxide (Wambui and Stephan, 2019). On the other hand, other species, such as *Clostridium tagluense,* have been linked with meat spoilage accompanied by little or no gas production (Dorn-In et al., 2022).

Microbial meat spoilage is usually attributed to an abundance of utilizable substrates including glycogen, lactate, amino acids, and low levels of glucose (Nychas et al., 2008). The model CEC species causing BPS, *C. estertheticum*, first utilizes glucose in the meat and once this substrate is exhausted, it switches to lactate utilization, which results in production of H_2 and CO_2 (Yang et al., 2009). The absence of ammonia and sulfur dioxide production by *C. estertheticum* suggests carbohydrates and lactate in meat are preferentially utilized over amino acids (Yang et al., 2009, 2010). Therefore, BPS has been directly linked to the utilization of fermentable substrates in meat. Compared to *C. estertheticum*, data on substrate utilization by other CEC species are limited, which further limits our understanding on the collective contribution of CEC to meat spoilage. The phenotypic data that are available for *C. tagluense*, suggest the species has a more narrow substrate spectrum than *C. estertheticum* (Suetin et al., 2009). The illustrated phenotypic heterogeneity among CEC species is indicative of genotypic diversity that may have an influence on metabolic pathways that are linked to meat spoilage.

The utilization of substrates coupled with the production of H_2 and CO_2 and other metabolites, such as acetate, butyrate, formate, butanol, and ethanol (Yang and Badoni, 2013) suggests that CEC species are capable of acetone–butyrate–ethanol fermentation, which can be linked to meat spoilage. Similarly, acetone–butyrate–ethanol fermentation has been linked with late blowing defect in hard and semi-hard cheese as a result of metabolism of lactate and other substrates by *Clostridium tyrobutyricum*, *Clostridium butyricum* and other clostridia resulting in the production of H_2, CO_2, and organic acids, mainly butyric acid, accompanied by abnormal aroma and flavor (Bassi et al., 2015; Gómez-Torres et al., 2015). The acetone–butyrate–ethanol fermentation is a biphasic process involving acidogenic and solventogenic phases. In the first phase, an organism converts diverse substrates in a typical acidogenic fermentation process consequently producing butyrate, acetate, H_2, and CO_2 (Patakova et al., 2013). The solventogenic phase occurs as the organism transitions to stationary phase, and it is characterized by growth cessation, a switch to butanol, acetone, and ethanol production and the initiation of endospore formation (Khamaiseh et al., 2014). The metabolic switch is necessary to avoid the deleterious effects of acid accumulation and to allow sufficient time for endospore formation (Zheng et al., 2009). Therefore, acetone–butyrate–ethanol fermentation is a critical and complex metabolic process needed for the development of acidogenic and solventogenic clostridia, such as CEC.

Comparative genomics allows multiple bacterial genomes to be more efficiently and conveniently compared in order to reveal inter- and intraspecies similarities or differences in particular groups of interest. Furthermore, it allows for better functional analysis at the genetic level (Yu et al., 2018). In this regard, a previous comparative genomic study involving six strains from four CEC species, *C. estertheticum*, *C. tagluense*, genomospecies1, and genomospecies2, identified genes involved in acetone–butyrate–ethanol fermentation and provided the first genomic link to the fermentation-mediated meat spoilage by CEC (Palevich et al., 2021). Although the fermentation has been linked with meat spoilage, the exact mechanism that results in variable gas production in chilled vacuum-packed meat has not been elucidated. Recently, more CEC genomes have been published (Wambui et al., 2020b, 2021a, 2022) thus providing an opportunity for comprehensive comparative genomics through which the genetic diversities within CEC can be assessed to reveal the underlaying causes of variable meat spoilage at the genomic level. At the same time, other unknown inter- and intraspecies genotypes of CEC can be unraveled. As a prerequisite to the analyses, it is imperative that efforts are made to increase the number of sequenced CEC genomes. Therefore, the current study aimed at (i) detecting, isolating, and sequencing new CEC strains from the meat processing environment and (ii) determining the genetic diversity of CEC through comprehensive comparative genomics.

MATERIALS AND METHODS

qPCR-Based Screening of *Clostridium estertheticum* Complex Strains

Four different sample types which included meat processing equipment ($n = 90$), bovine fecal samples ($n = 3$), and meat juice samples of unspoiled meat ($n = 153$) or BPS meat ($n = 3$) were screened for the presence of CEC by quantitative real-time PCR (qPCR) as previously described (Wambui et al., 2020c). Briefly, sponge samples obtained from the meat processing equipment were homogenized with Reinforced Clostridia Media (RCM; 25 ml) for 60 s using a stomacher. Hungate tubes containing 10 ml of the homogenized meat processing equipment sample were heated at 80°C for 10 min. The meat processing equipment and meat juice samples of unspoiled meat were incubated anaerobically for 10–14 days at 8°C prior to screening while meat juice samples from BPS samples were screened immediately after sampling. Current and subsequent anaerobic incubations were carried out in rectangular anerobic boxes (7.0 L; bioMérieux, Inc., Marcy l'Etoile, France), and the anaerobic conditions were generated by three 2.5 L AnaeroGen Sachets (Thermo Fisher Scientific, Waltham, MA, United States) per

box. DNA was extracted from 100 µl of each sample using the MagNa Pure LC DNA Isolation Kit III (Roche, Rotkreuz, Switzerland) by the MagNa Pure LC robotic workstation (Roche). Each sample was mixed with 10 µl lysozyme (20 mg/ml in phosphate-buffered saline) solution and incubated at 37°C for 30 min. Subsequent steps from lysis with proteinase K to DNA elution were carried out according to the instruction manual of the DNA extraction kit. The qPCR primers, probes and protocol used herein were as previously described (Brightwell and Clemens, 2012). qPCR-positive meat processing equipment and meat juice samples of unspoiled meat were enriched anaerobically 4 weeks at 8°C. The bovine fecal samples were obtained from a previous study (Wambui et al., 2021a) where attempts to isolate CEC strains were unsuccessful and thus had been stored anaerobically at 4°C.

Isolation and Presumptive Identification of CEC Strains

Isolation of CEC from meat juice samples of BPS samples was carried out by direct plating. Briefly, the samples were serially diluted 10-fold in saline solution and plated on Columbia agar supplemented with 5% defibrinated sheep blood (CBA). Isolation of CEC from the enriched meat processing equipment, bovine fecal samples, and meat juice samples of unspoiled meat was carried out after elimination of competitive microflora and activation of CEC spores as previously described without modification (Wambui et al., 2021b). Colonies on CBA displaying clostridial characteristics described previously (Húngaro et al., 2016) were selected and purified anaerobically on CBA for 3 weeks at 8°C. The isolates were confirmed as members of CEC by qPCR as described above.

DNA Extraction, Whole-Genome Sequencing, and Identification of CEC Strains

The extraction and whole-genome sequencing of CEC isolates was carried out as previously described (Wambui et al., 2020a). Briefly, the genomic DNA was isolated using the DNA blood and tissue kit (Qiagen, Hombrechtikon, Switzerland). The sequencing outputs, which were 150–300 bp pair-ended reads, were prepared using Nextera DNA Flex chemistry using the Nextera DNA Flex Sample Preparation Kit (Ill) as per manufacturer's guidelines (Illumina, San Diego, CA, United States). The resulting transposon-based libraries were sequenced on a MiniSeq sequencer (Illumina) with a minimal coverage of 50-fold. The MiniSeq MidOutput Reagent Cartridge (300 cycles) was used. Demultiplexing and adapter trimming was done using the Miniseq local run manager version 2.4.1 using standard settings. The reads were checked for quality using FastQC (Andrews, 2010) then assembled with SPAdes v. 3.12.0 (Bankevich et al., 2012) using Shovill 1.0.9.[1] Since *Clostridium algoriphilum* DSM 16153^T has not been sequenced yet, we purchased the strain from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, revived it as previously described, and sequenced it as described above. The quality of the genomes was checked using ContEST16S (Lee et al., 2017) and CheckM (Parks et al., 2015). The complete *rpoB* gene sequences were identified using blastn and extracted from the genomes with samtools using the faidx option and the blast results as input (Altschul et al., 1990; Li et al., 2009). These and the *rpoB* sequences of 34 publicly available CEC genomes (Wambui et al., 2021a), were aligned in CLC Workbench Genomics v. 8.1 (Qiagen, Aarhus, Denmark) using default settings while the phylogenetic tree was created with the same software using the Maximum likelihood Phylogeny method whereby the neighborhood joining method using the Jukes Cantor model were applied, respectively. Bootstraps were based on 1,000 replicates. The correct species assignment of the CEC isolates was carried out *in silico* from the pairwise nucleotide comparison of Average Nucleotide Identity (ANI), which was determined with pyANI (Pritchard et al., 2016) using the BLAST algorithm. The *rpoB*-based phylogeny consisting of the Type/Representative strains of CEC, *Clostridium algidicarnis*, *Clostridium frigidicarnis*, *Clostridium gasigenes*, *Clostridium putrefaciens*, and *Clostridium argentinense* was used to validate a divergent isolate as a member of CEC. Where necessary, resulting newick files were visualized in iTOL (Letunic and Bork, 2019). Core genome MLST was performed in Seqsphere+8.2.0 (Ridom, Münster, Germany). The *C._estertheticum_*cgMLST scheme was constructed using the type strain DSM 14864^T as seed and all other available *C._estertheticum_*genomes as testers. Standard settings were applied for target selection.

Characterization of the General Features of CEC Genomes and Pan-Genome Analysis

The 16 and 34 CEC genomes from the current and previous studies, respectively (**Supplementary Tables 1A,B**) were annotated using Prokka (Seemann, 2014) and RAST (Brettin et al., 2015). The GFF3 output files from Prokka were used for pan-genome analysis of the 50 CEC genomes in Roary v3.11.2 (Page et al., 2015) with the minimum percentage identity for BLASTP set at 60%, 70%, 80%, 90%, and 95%. For each setting, four different gene classes grouped into "core" (99% ≤ strains ≤100%), "soft-core" (95% ≤ strains ≤100%), "shell" (15% ≤ strains <95%), and "cloud" (0% ≤ strains <15%) were obtained with the -s setting for paralogues clustering. Pan-genome analysis were further carried out for the 25 *C. estertheticum* genomes in Roary with the minimum percentage identity for BLASTP set at 80%. Recombination events in CEC were estimated using Gubbins version 2.4.1. (settings: -f 30) using coreSNP alignment and tree as input (Croucher et al., 2015). For this purpose, the core genome alignment was created using Parsnp v1.5.3 (Treangen et al., 2014). The maximum likelihood phylogeny of *C. estertheticum* was determined using RAxML (Randomized Axelerated Maximum Likelihood; Stamatakis, 2014). The phylogenetic trees were visualized and annotated with metadata in iTOL (Letunic and Bork, 2019). Pan-genome characteristics were calculated using the package micropan in R using 500 the random permutations of genome ordering (Snipen and Liland, 2015).

[1]https://github.com/tseemann/shovill

Comparative Genomics and Phenotypic Characterization CEC

Function annotation and classification of proteins were performed by sequence comparison against the eggNOG database 5.0 (Huerta-Cepas et al., 2019), using the eggNOG-Mapper v2 (Cantalapiedra et al., 2021). Putative enzymes involved in the breakdown, biosynthesis, or modification of carbohydrates were identified using CAZy database using dbCAN using default settings (Drula et al., 2022). The genes involved in the acetone–butyrate–ethanol fermentation were identified from the previous work on CEC (Palevich et al., 2021) and other solventogenic clostridia (Li et al., 2020). In addition, the genes annotated as hydrogenases were manually identified in the RAST annotated CEC genomes. In total, a comprehensive list of 48 genes (**Supplementary Table 2**) was established and each gene was manually identified and counted in each of the RAST annotated CEC genome. Where applicable, the genes and pathways were validated using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa et al., 2016).

For phenotypic characterization, respective cultures were initially grown anaerobically at 8°C on CBA then standardized to McFarrand 3.0. Utilization of fermentable substrates by *C. estertheticum* and *C. tagluense* was determined using the standardized cultures as previously described (Wambui et al., 2020a) using the API20A kit (bioMeriéux, Marcy l'Etoile, France) and following manufacturer's specifications. The API20A strips were incubated at 8°C anaerobically for 2 weeks. For motility test and gas production in meat, standardized cultures were first inoculated 1:100 into RCM broth and incubated at 8°C for 10 days. The motility test was performed in liquid media which essentially determines swimming motility, requires shorter incubation time than growth on agar, and is easily reproducible. For this test, 200 µl of each culture was slowly injected into Hungate tubes containing fresh RCM broth using a syringe and the tubes incubated for 2 weeks at 8°C without shaking. Motility was determined as the movement of the cultures from the surface toward the bottom of the tubes. Gas production in meat was carried out as previously described with slight modifications (Dorn-In et al., 2022). Five hundred microliter of each culture was spread on the surface of approximately 500 g of beef sample. The samples were vacuum-packed and stored for 28 or 56 days at 8°C. Uninoculated vacuum-packed beef samples were included as negative controls. At the end of the storage period, meat spoilage was scored from 0 to 5 according to a previous description (Boerema et al., 2007). Loss of flagellar genes has been shown to increase exopolysaccharide synthesis (Watnick et al., 2001). Therefore, CEC colony morphology that could be related loss of flagellar genes, including sliminess, were determined on cultures grown on CBA for 8 weeks at 8°C. Each experiment was carried out in three biological replicates apart from the gas production in meat with two biological replicates.

Data Analysis

Data for substrate utilization, motility test, and gas production were described qualitatively. The distribution of carbohydrate-active enzymes was described as means and SD and the means compared using the Mann–Whitney U ($p = 0.05$) with Bonferroni correction. Data analysis and visualization were carried in R Studio Version 1.1.463 (RStudio, Inc., Boston, United States).

RESULTS

Identification of New Sequenced Strains, Including a Novel Species, Highlights the Diversity of CEC Species

In an effort to expand the existing collection of sequenced CEC strains, 15 new strains, which were confirmed by qPCR, were isolated from meat processing equipment ($n = 2$), bovine fecal samples ($n = 3$), and meat juice samples of unspoiled meat ($n = 7$) and BPS meat ($n = 3$). Based on the *rpoB* gene phylogeny (**Figure 1A**), 14 strains were identified as *C. tagluense* ($n = 8$) and *C. estertheticum* ($n = 6$). Their correct species identification was confirmed and validated by ANIb analysis (**Supplementary Figure 1**). The ANIb between the *C. tagluense* strains and *C. tagluense* DSM 17763^T ranged between 97.27% and 97.32% while the ANIb values between the *C. estertheticum* strains and *C. estertheticum* DSM 8809^T ranged from 95.80% to 98.18% confirming the correct species identification of the strains. Strain, CS001, formed an outgroup compared to other known 49 CEC strains (**Figure 1A**) suggesting a novel species within the CEC. Interspecies ANIb between CS001 and the other CEC strains ranged from 78.60% to 81.36% for *C. psychrophilum* DSM 14207^T and *C. tagluense* DSM 17763^T, respectively. Furthermore, strain CS001 had an ANIb of <81.7% with any of the 122 representative genomes of the genus *Clostridium* in the NBCI database, confirming it is a novel CEC species. The new species is herein referred to as genomospecies4. *rpoB*-based phylogenetic reconstruction with 16 type/representative strains from 11 CEC species, four non-CEC *Clostridium* species that are commonly isolated from meat (*C. algidicarnis*, *C. frigidicarnis*, *C. gasigenes*, and *C. putrefaciens*) and *C. argentinense*, which is the known closest *Clostridium* species to CEC, further provided evidence that genomospecies4 is a novel species within CEC despite its high divergence (**Figure 1B**). The novel species shows CEC is both expansive and diverse.

The strains, CM032, CM033, and CM034, isolated from BPS cases (**Figure 1C**; **Supplementary Figure 2**) were identified as *C. estertheticum*, and clustered together (**Figure 1A**). cgMLST of species *C. estertheticum* demonstrated the three strains had SNP differences of two to three, indicating that they are clonal (**Supplementary Figure 3**). Interestingly, strains CM033 and CM034 were isolated from BPS involving horse meat (**Figure 1C**), while their involvement in BPS was validated with CM034, which caused BPS in vacuum-packed beef (**Figure 1D**; **Table 1**). To the best of our knowledge, this is the first reported case of BPS in horse meat caused by *C. estertheticum*. Phenotypic analysis revealed CS001 from the novel species caused meat spoilage with a score of 3 which corresponds with blown puffy packs (**Table 1**). We further compared the new species with the other recently described species, genomospecies2 and

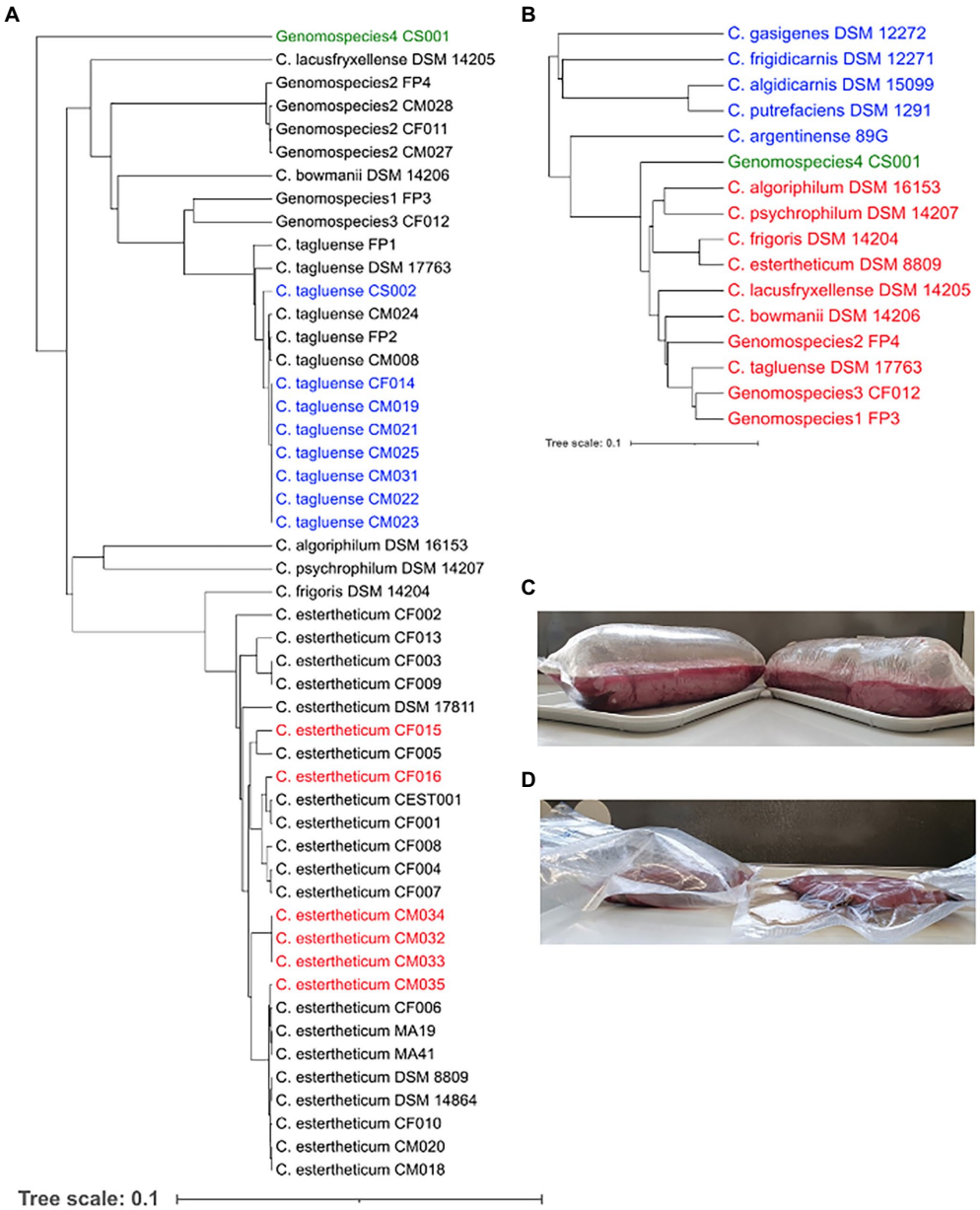


FIGURE 1 | Identification of new *Clostridium estertheticum* complex (CEC) isolates. **(A)** *rpoB* gene-based phylogeny of CEC strains. Newly isolated *C. estertheticum*, *C. tagluense,* and genomospecies4 strains are highlighted in red, blue, and green text, respectively. **(B)** Genomospecies4 CS001 (green) was confirmed as a novel member of CEC using the *rpoB* gene-based phylogenetic analysis of type/representative strains of CEC (red) and other related *Clostridia* spp. (blue). **(C)** Blown pack spoilage cases of horse meat that were likely caused by *C. estertheticum* CM033 and CM034. **(D)** Blowing ability of *C. estertheticum* CM034 (left) in beef against uninoculated beef (right).

genomspecies3 whose representative strains, CM028 and CF012, caused spoilage with a score of 3 and 2, respectively with the latter score corresponding to loss of vacuum without evidence of substantial gas production. This further demonstrates the significant involvement in meat spoilage of not only the known CEC species, but also other newly described species.

Clostridium algoriphilum DSM 16153^T was also sequenced within this study. On the *rpoB* phylogeny (**Figure 1A**), the strain was closely related to *C. psychrophilum* DSM 14207^T. This was further confirmed by the ANIb of 84.63% between the two genomes. *Clostridium algoriphilum* DSM 16153^T was distantly related to the newly identified genomospecies4 CS001 (ANIb = 79.53%).

Genome and Pan-Genome Analyses Reveal High Genetic Diversity Within CEC

Although CEC forms a single clade within the *Clostridium* genus, their genomes demonstrated high variability with respect

TABLE 1 | Blown pack spoilage (BPS) potential of selected *Clostridium estertheticum* complex (CEC) isolates.

Strain	Species	BPS score†	Characteristic
CM034	*C. estertheticum*	4	Fully distended pack, but not tightly stretched
CM028	Genomospecies2	3	Blown puffy pack
CF012	Genomospecies3	2	Loss of vacuum within the pack
CS001	Genomospecies4	3	Blown puffy pack
DSM 14864*	*C. estertheticum*	5	Tightly stretched overblown pack
Negative control (no inoculum)		0	No gas bubbles in the drip

**Positive control.*
†Mean of two biological replicates.

to size, number of coding sequences (CDS), and G + C content (**Supplementary Table 1**). The size varied from 3.662 Mbp in genomospecies2 CM028 to 5.555 Mbp in genomospecies1 FP3 while the CDS ranged from 3.337 in genomosspecies2 CM028 to 5.271 in *C. tagluense* FP2. The G + C content ranged from 29.9% in *C. psychrophilum* DSM 14207 to 31.7% in genomospecies2 CM027 and genomospecies3 CF012. The size, CDS and G + C of CS001 from the newly identified species, genomospecies4, were 4.139 Mbp, 3,831 and 31.6%, respectively, and 4.520 Mbp, 4,524, and 30.7%, respectively, in the newly sequenced *C. algoriphilum* DSM 16153^T. Overall, genomospecies2 has the smallest genome size within CEC (**Supplementary Figure 4**).

The variability of CEC genomes may have contributed to interspecies or intraspecies diversity, population genetics, and evolution. We therefore explored the entire genomic repertoire of the CEC by systematically determining its pan-genome with the BLAST identity cutoff at 60%, 70%, 80%, 90%, and 95%. The cutoff had a significant effect on the CEC pan-genome whereby its increase corresponded with a decrease in core- and soft-core genomes and an increase in shell- and cloud-genomes (**Supplementary Table 3**). Because of these effects, further pan-genome analyses were performed on the data set from 80% BLAST identity cutoff whereby the core- and soft-core genomes comprised of 748 and 942 genes, respectively, while the shell- and cloud-genomes had 7,296 and 28,741 genes, respectively (**Figure 2A**). The core genome represented 14.19%–22.40% of the gene content of each strain, while the cloud genes represented 77.72% of the pan-genome. The large accessory genome highly suggests a significant degree of genomic diversity within CEC. The pan-genome frequency showed a proportional relationship between the number of genes and genomes and the curve analysis indicates the CEC pan-genome is open (**Figure 2B**). The pan-genome of CEC had an alpha of 0.456. Similarly, the pan-genome of *C. estertheticum* had an alpha of 0.392. Given that the pan-genome is closed if the estimated alpha is above 1.0, both pan-genomes can be considered to be open.

Evaluation of Cluster of Orthologous Groups Reveal Variability in Carbohydrate Metabolism and Transport in CEC

In a next step, we evaluated the cluster of orthologous groups (COGs) functional classification to define possible differences in the functions encoded by the core, soft-core, cloud-, and shell-genomes of CEC (**Figures 2C–E**). In all four cases, the highest proportion of genes were COG category S (Function unknown). Among the known functions, both core and soft-core gnomes consisted mostly of COG categories J (Translation), L (Replication and repair), K (Transcription), E (Amino Acid metabolism and transport), and F (Nucleotide metabolism and transport), which suggest that the processes controlling the flow of genetic information in CEC are conserved. On the other hand, the shell-genome was biased toward K (Transcription), T (Signal Transduction), E (Amino Acid metabolism and transport), C (Energy production and conversion), G (Carbohydrate metabolism and transport) while the cloud genome consisted mostly of L (Replication and repair), K (Transcription), M (Cell wall/membrane/envelop biogenesis), G (Carbohydrate metabolism and transport) and T (Signal Transduction) categories. The identification of carbohydrate metabolism and transport category in both cloud- and shell-genomes suggests that these processes are among the most genetically variable within CEC.

Recombination Influences the Population Structure of CEC

To determine whether recombination events have influenced the population structure of CEC, we first determined the occurrence of recombination within the core genome of CEC. The core-genome-based phylogenetic tree (**Figure 3A**) confirmed all species identified by the *rpoB* gene (**Figure 1A**), but the population structure varied slightly whereby it was divided into Clade 1 consisting of *C. estertheticum*, *C. frigoris*, and *C. psychrophilum*, Clade 2 consisting of *C. bowmanii C. algoriphilum C. lacusfryxellense* genomospecies2 and genomospecies4 and Clade 3 consisting of *C. tagluense*, genomospecies1, and genomospecies3. Within the core genome, we identified 252 recombination events, out of which 70 were ancestral, that is, they occurred between or among strains, while 182 were strain-specific (**Supplementary Figure 5**). Among, the ancestral recombination events, the majority occurred within the species and more specifically between strains that were more closely related than distantly related. The reconstruction of the core phylogenetic tree without the recombined regions (**Figure 3B**) revealed the influence of recombination on the population structure of CEC. While three clades were identified, the species within these clades varied slightly with those before the removal of recombination events. Clade 1 consisted of *C. estertheticum* and *C. frigoris*, Clade 2 consisted of *C. algoriphilum*, *C. lacusfryxellense,* and *C. psychrophilum* and genomspecies2 while Clade 3 consisted of *C. tagluense*, *C. bowmanii*, genomospecies1, genomospecies3, and genomospecies4. The data show that recombination events within CEC occur regularly and have also contributed to the current population structure of the complex.

The Reduced Genome and Speciation of Genomospecies2 Correspond With Absence of Flagella Related Genes

The variable interspecies genome sizes within CEC shows genes have been acquired or lost within the complex, which might contribute to speciation. This prompted us to determine signature

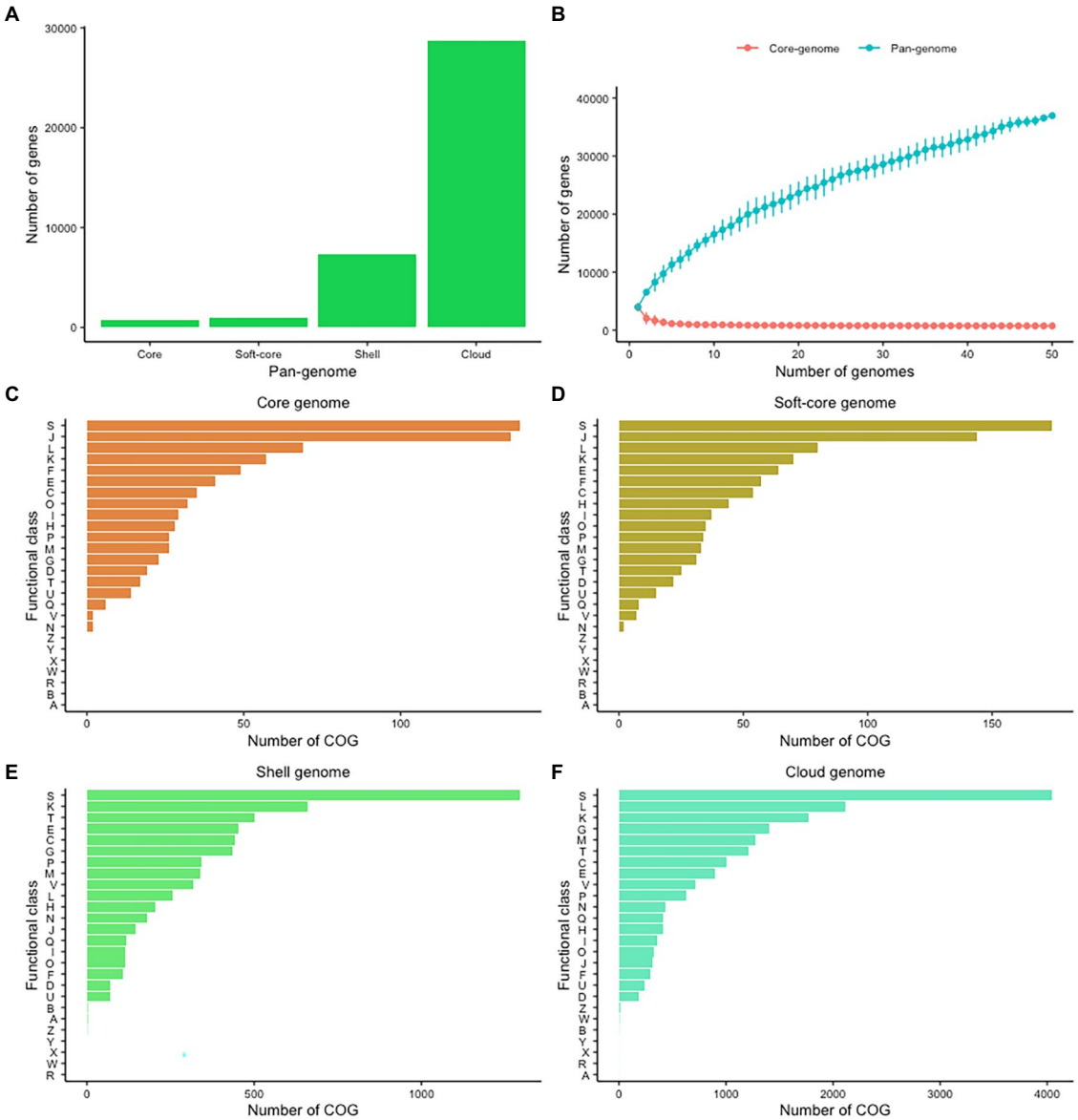


FIGURE 2 | The pan-genome of CEC. **(A)** The sizes of core-, soft-core, shell-, and cloud-genomes at 80% BLAST cutoff identity. **(B)** The sizes of core- and pan-genomes in relation to numbers of genomes added into the gene pool as a function of the number of genomes included. **(C–F)** Distribution of cluster of orthologous genes (COGs) in the core, soft-core, shell-, and cloud-genomes. COG categories: A, RNA processing and modification; C, energy production and conversion; D, cell cycle control and mitosis; E, amino acid metabolism and transport; F, nucleotide metabolism and transport; G, carbohydrate metabolism and transport; H, coenzyme metabolism; I, lipid metabolism; J, translation; K, transcription; L, replication and repair; M, cell wall/membrane/envelop biogenesis; N, cell motility; O, post-translational modification, protein turnover, chaperone functions; P, inorganic ion transport and metabolism; Q, secondary structure; R, general functional prediction only; S, function unknown; T, signal transduction; U, intracellular, trafficking and secretion; V, defense mechanisms; W, extracellular structures; X, mobilome: prophages, transposons; and Z, cytoskeleton.

genes that have been acquired or lost within CEC and determine whether the loss or gain may have contributed to speciation events within the complex. In this case, we investigated genomospecies2, which has the least genome size of less than 4.0 Mbp (**Supplementary Figure 4**). An analysis of the CEC accessory genome led to the identification of nine genes that were exclusively absent in genomospecies2. These included chemotaxis-related genes (*cheA*, *cheD*, and *cheW*), RNA polymerase sigma-D factor, and two putative type III secretion system genes that were annotated as putative ATP synthase YscN and Yop proteins translocation protein U. Further analysis of the genes showed each CEC strain had 23–28 genes related to chemotaxis and flagella assembly while genomospecies2 strains had one gene each (data not shown). A whole-genome sequence-based comparison of genomospecies2 and the two closely related species, *C. bowmanii* and *C. lacusfryxellense* using RAST

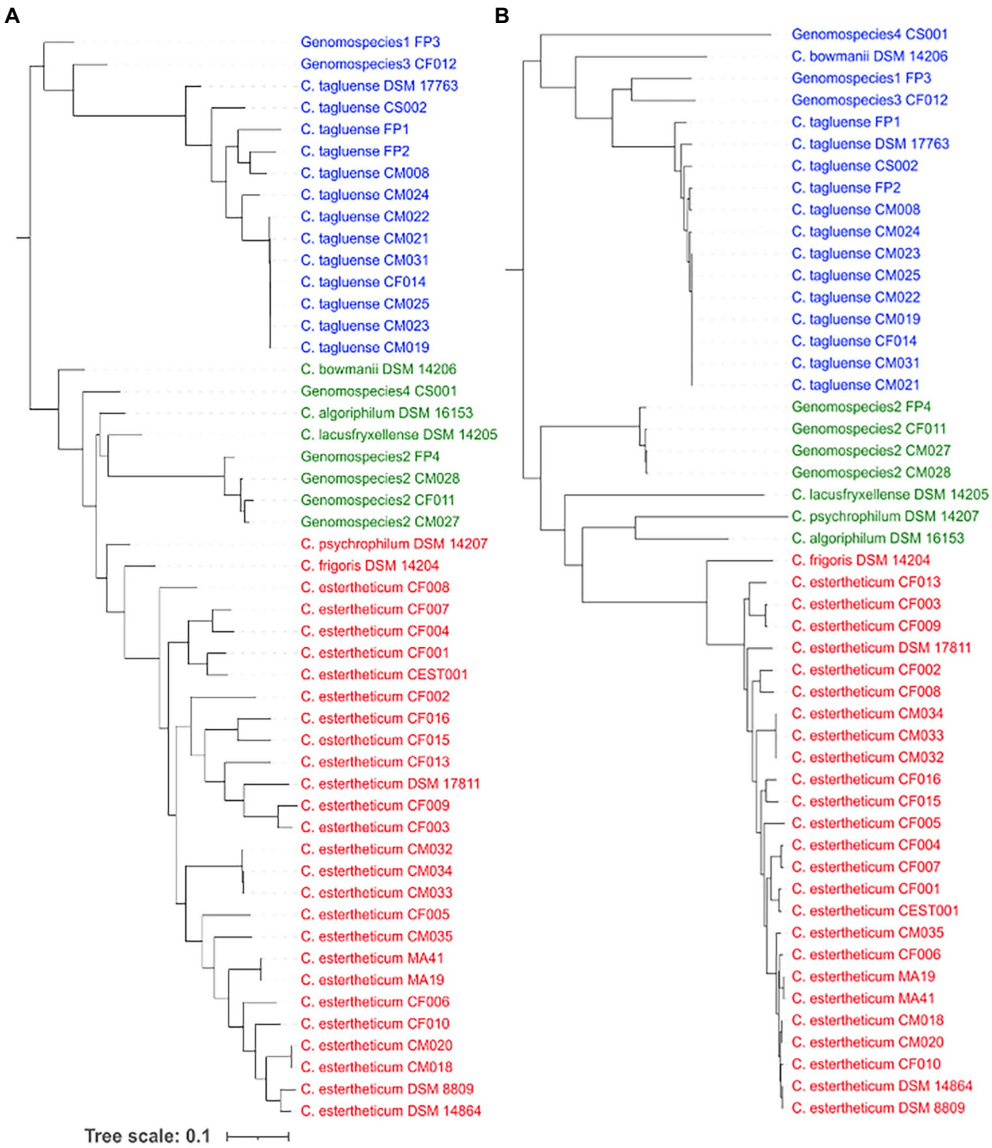


FIGURE 3 | Influence of recombination events on the population structure of CEC. The core genome-based phylogeny of CEC **(A)** before and **(B)** after the removal of recombination events and three corresponding clades from each tree are highlighted in blue, green, and red fonts.

annotation server showed up to 55 related genes were absent in genomospecies2 (**Figure 4A**) further confirming the absence of these genes. The lack of these genes was consistent with reduced motility of the genomospecies2 CF011 when compared with *C. bowmanii* DSM 14206 in RCM media (**Figure 4B**). An investigation of genomospecies2 strains CF011, CM027, and CM028 colony morphology showed they form significantly larger and slimy colonies after extended incubation periods (>8 weeks) compared with all other CEC species (**Figure 4C**). Further analysis of gene presence/absence in the pan-genome matrix did not reveal any exopolysaccharide genes that were specific to genomospecies2. Therefore, genomospecies2 is currently the only known CEC species lacking genes related to flagella signaling and assembly and this loss influences the motility and may also be linked with the unique colony morphology.

Identification of Conserved Metabolic Genes That Are Present/Absent in Two CEC Phylogroups

Although CEC causes meat spoilage, the spoilage is distinguished as either corresponding with formation of copious or low amounts of gas (**Table 1**). This prompted us to investigate the presence/absence of conserved metabolic genes that can explain these differences using *C. estertheticum* and *C. tagluense* as representative species. The analysis of the CEC accessory genome revealed 166 and 412 genes that were only present in *C. estertheticum*

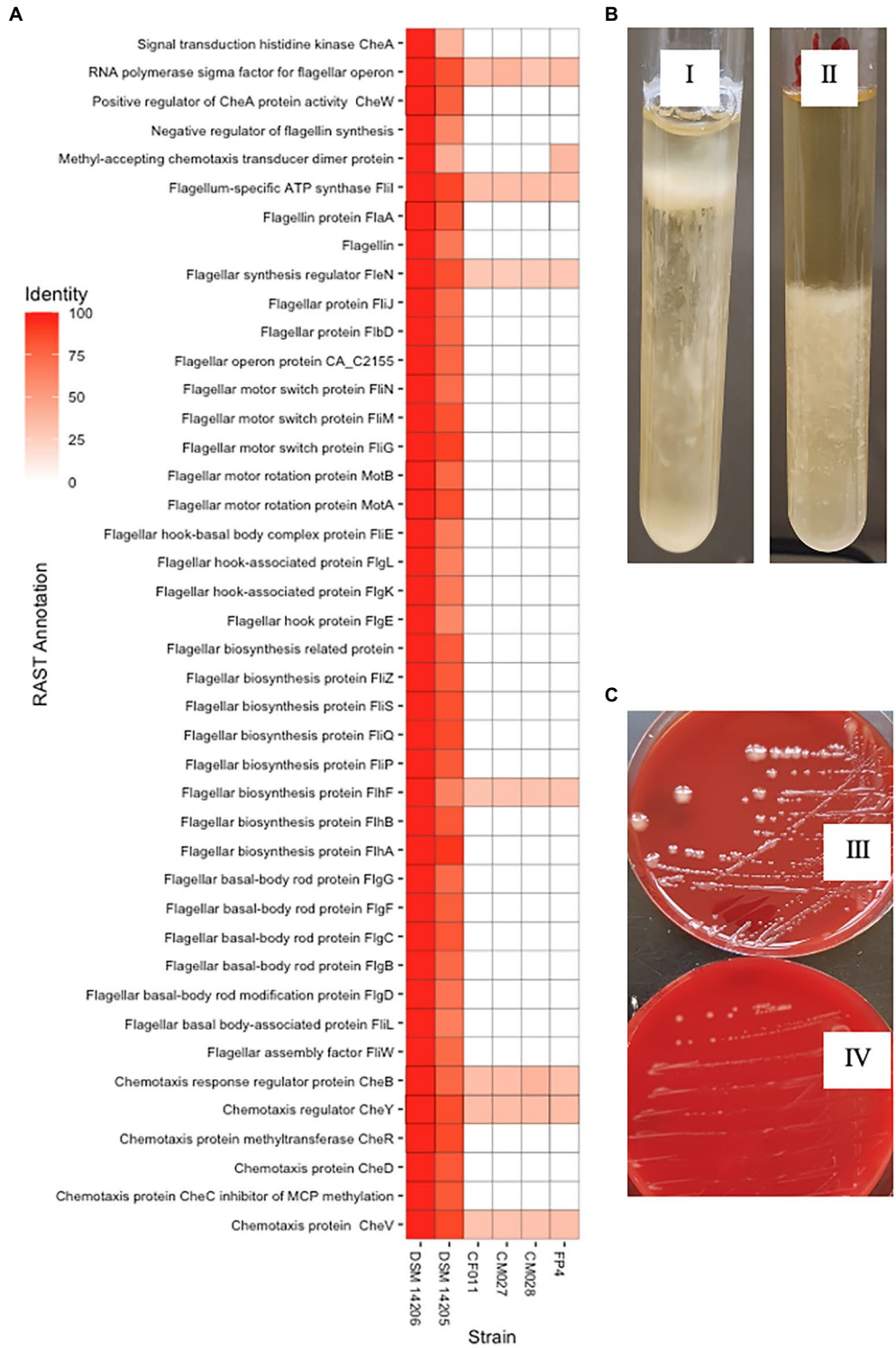


FIGURE 4 | Unique genotypic and phenotypic characteristics of genomospecies2. **(A)** Whole-genome sequence-based comparison in RAST server of genomospecies2 strains FP4, CF011, CM027, and CM028 against *Clostridium bowmanii* DSM 14206 and *Clostridium lacusfryxellense* DSM 14205 reveals the lack of flagellar assembly and signaling genes in genomospecies2. **(B)** Motility test in Reinforced Clostridium Media broth shows genomospecies2 CF011 has reduced motility compared to *C. bowmanii* DSM 14206. Both strains were incubated for 14 days without shaking. **(C)** On Columbia Blood Agar plates, genomospecies2 CF011 forms larger and slimy colonies unlike *C. bowmanii* DSM 14206. Both strains were incubated for 8 weeks.

and *C. tagluense*, but no obvious genes could be linked to meat spoilage that is characteristic of the two species or CEC in general (data not shown). Therefore, we hypothesized that these metabolic genes are also conserved with other species given the high relatedness of closely related CEC species. Therefore, the CEC species were divided into two major phylogroups, phylogroup 1 (PG1) and Phylogroup 2 (PG2), which is consistent with the *rpoB* gene phylogenetic tree. PG1 consisted of *C. estertheticum*, *C. frigoris*, *C. psychrophilum*, and *C. algoriphilum* while PG2 consisted of *C. tagluense*, *C. lacusfryxellense*, *C. bowmanii*, genomospecies1, genomospecies2, genomospecies3, and genomospecies4. The analysis identified 63 and 27 genes that were conserved in PG1 and PG2, respectively. Elimination of seven genes with similar annotations between the two groups and 12 hypothetical proteins allowed 71 (49 and 14 genes in PG1 and PG2, respectively) genes to be compared. Interestingly, genes linked to hydrogen and carbon dioxide production could be identified among the 71 genes. Among these included formate hydrogenlyase subunits 4 and 7 and Hydrogenase-4 components B and G that were specific to PG1 while hydrogenase isoenzymes formation protein HypE was specific to PG2. This suggested that PG1 and PG2 have evolved different fermentation pathways that might influence gas production. Further comparison of the 71 genes revealed pentose metabolic genes (arabinose metabolism transcriptional repressor, arabinose operon regulatory protein, arabinose-proton symporter L-arabinose isomerase, L-ribulose-5-phosphate 4-epimerase, ribulokinase, and transaldolase) and lactose and galactose metabolic genes (lactose permease and tagatose-6-phosphate kinase) were identified in PG1 only. This supports the COG analysis and further suggests PG1 and PG2 have evolved different carbohydrate metabolic pathways which might not only influence their habitat adaptation, but also meat substrate utilization and their respective meat spoilage mechanisms.

Variable Utilization of Carbohydrates Is Linked With Variable Distribution of Carbohydrate-Active Enzymes in *Clostridium estertheticum* and *Clostridium tagluense*

To determine the differences in carbohydrate metabolic pathways between PG1 and PG2, we specifically investigated the distribution of six groups of carbohydrate-active enzymes (CAZy), namely, carbohydrate esterases (CE), carbohydrate-binding modules (CBM), glycoside hydrolases (GH) glycosyltransferases (GT), polysaccharide lyases (PL), and auxiliary activities (AA), in *C. estertheticum* and *C. tagluense* genomes (**Supplementary Figure 6**), which represented PG1 and PG2, respectively. In *C. estertheticum*, GHs were the most frequent CAZy genes and ranged from 40 to 69 per genome. In contrast, GTs, which ranged from 30 to 42 per genome, were the most frequent in *C. tagluense*. Overall, the frequencies of AAs, PLs, CBMs, GTs, and GHs were higher ($p < 0.05$) in *C. estertheticum* than in *C. tagluense*. However, GTs were more frequent ($p < 0.05$) in *C. tagluense* than in *C. estertheticum*. An in-depth analysis of the GHs revealed the two species encoded 53 different types of enzymes (**Figure 5**), with *C. estertheticum* having the highest diversity of the enzymes. Specifically, 36 GHs were only identified in *C. estertheticum* and eight of these GHs (GH3, GH20, GH32, GH36, GH42, GH43, GH51, and GH115) were present in all 25 *C. estertheticum* genomes while two GHs (GH28 and GH105) were present in 24 *C. estertheticum* genomes. In contrast, only three GHs were exclusively identified in *C. tagluense*. Among these, GH171 was present in all *C. tagluense* genomes while GH129 was absent in only one genome. Despite the differences, nine GHs (GH13, GH18, GH23, GH39, GH65, GH73, GH94, GH109, and GH156) were present in all *C. estertheticum* and *C. tagluense* genomes. Among these, GH13, GH94, and GH109 were significantly higher ($p < 0.05$) in *C. estertheticum* than in *C. tagluense* while GH73, and GH156 were significantly higher ($p < 0.05$) in *C. tagluense* than in *C. estertheticum* (**Supplementary Figure 7**). Therefore, this confirmand the variable distribution of carbohydrate utilization genes between *C. estertheticum* and *C. tagluense*.

The variable distribution of carbohydrate utilization genes between the two species suggested the range of substrates they can utilize also differs. To validate this, we determined the utilization of 16 substrates (**Table 2**). *Clostridium estertheticum* strains utilized D-glucose, D-lactose, D-sucrose, D-maltose, D-salicin, D-xylose, L-arabinose, D-cellobiose, and D-raffinose. The strains showed variable utilization of six other substrates and did not utilize D-trehalose. Comparatively, *C. tagluense* strains utilized only three substrates, D-glucose, D-maltose, and D-trehalose while three strains utilized glycerol. Therefore, the high number of GHs in *C. estertheticum* corresponded to utilization of diverse fermentable substrates compared to *C. tagluense*.

Variable Gas Production in Vacuum-Packed Meat by *Clostridium estertheticum* and *Clostridium tagluense* Corresponds With the Distribution of [NiFe] Hydrogenase *hyp* Gene Cluster

Carbohydrate utilization is liked with gas production in *Clostridium* genus through the acetone–butanol–ethanol fermentation pathway. To determine whether the differences in gas evolution during meat spoilage between the *C. estertheticum* and *C. tagluense* is solely linked with differences in the carbohydrate utilization genes or there exists other underlaying genetic influences, we first reconstructed the acetone–butyrate–ethanol fermentation pathway to include steps for the production of lactate, formate, ethanol, acetate, butyrate, acetone, acetoin, and butanol (**Figure 6**). The genes involved in the pathway were present in both species with minor variations including copy numbers (**Figure 7**). This suggested the central acetone–butyrate–ethanol fermentation pathway was conserved between the two species at the genetic level and cannot fully explain the differences in gas production. We therefore carried in-depth analysis whereby we identified and compared two different processes that produce H_2 and CO_2 through the oxidation of pyruvate in the acetone–butyrate–ethanol fermentation pathway

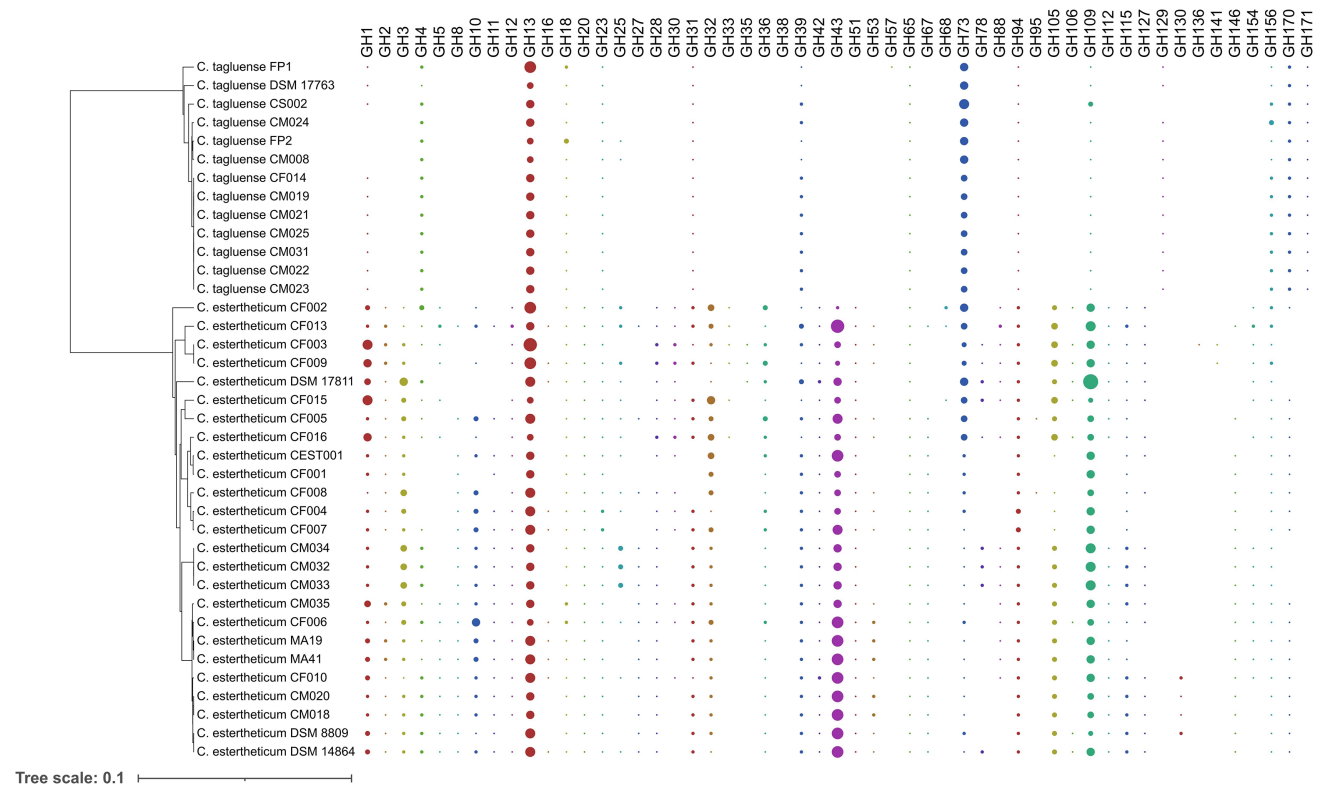


FIGURE 5 | Distribution of glycoside hydrolases in *C. estertheticum* and *C. tagluense*. The size of each circle corresponds with the frequency of the gene. A blank shows the gene is absent. The phylogeny is based on the *rpoB* gene.

(**Figure 6**). The first process involves the oxidative decarboxylation of pyruvate to acetyl-coA and CO_2 by pyruvate ferrodoxin oxidoreductase (PFOR). This process requires the reduction of ferredoxin, which is further oxidized simultaneously by [FeFe] hydrogenase, HydA, and ferredoxin NADH oxidoreductase (FNOR) resulting in the formation of H_2 (**Figure 6**). All genes involved in this process including *pfor*, *fnor*, and *hydA* were present in both species although the copy number of *hydA* in *C. tagluense* was higher than in *C. estertheticum* (**Figure 7**). The genes involved in the maturation of HydA, that is *hydE*, *hydF*, and *hydG*, were also identified (**Figure 7**). The second process involves the decarboxylation of pyruvate to acetyl-coA and formate by pyruvate formate lyase (PFL; **Figure 6**). The formate is then oxidized to CO_2 by a formate dehydrogenase (Fdh) or to CO_2 and H_2 by formate-hydrogenase lyase complex (FHL; comprising of Fdh and a hydrogenase; **Figure 6**). The presence of the *pfl* gene indicated the process of formate production is conserved in both species (**Figure 7**). However, the *fdh* gene was present in all *C. tagluense* genomes and six *C. estertheticum* genomes (**Figure 7**).

Interestingly, we identified other gene clusters, which likely encode putative hydrogenases that can play an important role in production of molecular hydrogen. A gene cluster for bifurcating [FeFe] hydrogenase encoding genes for *alpha*, *beta,* and *gamma* subunits was identified in both species. The bifurcating [FeFe] hydrogenase is herein referred to as Hya and is encoded by the *hya* cluster consisting of *hyaABC* genes, encoding the alpha, beta, and gamma subunits, respectively (**Figure 7**). The presence of *hyaABC* genes suggests the metabolic process of its product is conserved in both species and together with HydA hydrogenase may also not fully explain the differences observed in gas production at the genetic level. Additionally, we found three different hydrogenase gene clusters in both species (**Figure 7**). One cluster herein referred to as *hyc* cluster consists of six genes *hycBCDEFG* was exclusive to *C. estertheticum*. The second cluster, herein referred to as *hyp* cluster consisting of nine [NiFe] hydrogenase genes in the order *hypABCDEFGHB* was present in all *C. tagluense* strains and nine *C. estertheticum* strains. The strains with the *hyp* cluster formed two distinct clusters within *C. estertheticum* suggesting the intraspecies presence or absence of the cluster was evolutionary. The third [FeFe] hydrogenase gene cluster, here in referred to as *hyb* cluster consisted of four genes *hybABCD*, and was exclusive to *C. tagluense*. We speculated the variable distribution of the *hyb*, *hyc,* and *hyp* hydrogenase gene clusters could explain the phenotypic differences in gas production in meat between *C. estertheticum* and *C. tagluense*.

To link the genome and phenotype, we tested gas production by both species in meat. Three out of the four *C. estertheticum* strains DSM 14864, CM020, and CM034 caused pack distention while one *C. estertheticum* strain, CEST001, and all four *C. tagluense* strains CM008, CM023, CM024, and CS002, did

TABLE 2 | Substrate utilization by in-house *C. estertheticum* and *Clostridium tagluense* (+, positive reaction and –, negative reaction).

Strain ID	GLU	MAN	LAC	SAC	MAL	SAL	XYL	ARA	GLY	CEL	MNE	MLZ	RAF	SOR	RHA	TRE
C. tagluense CM008	+	–	–	–	+	–	–	–	+	–	–	–	–	–	–	+
C. tagluense CM019	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+
C. tagluense CM021	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+
C. tagluense CM022	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+
C. tagluense CM023	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+
C. tagluense CM024	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+
C. tagluense CM025	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+
C. tagluense CM031	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+
C. tagluense CF014	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	+
C. tagluense CS002	+	–	–	–	+	–	–	–	+	–	–	–	–	–	–	+
C. estertheticum CEST001	+	+	+	+	+	+	+	+	–	+	+	–	+	–	–	–
C. estertheticum CF001	+	+	+	+	+	+	+	+	–	+	+	–	+	–	–	–
C. estertheticum CF002	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	–
C. estertheticum CF003	+	–	+	+	+	+	+	+	–	+	–	–	+	–	+	–
C. estertheticum CF004	+	+	+	+	+	+	+	+	–	+	+	–	+	+	–	–
C. estertheticum CF005	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CF006	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CF007	+	+	+	+	+	+	+	+	–	+	+	–	+	–	–	–
C. estertheticum CF008	+	–	+	+	+	+	+	+	+	+	+	–	+	–	–	–
C. estertheticum CF009	+	–	+	+	+	+	+	+	–	+	–	–	+	–	+	–
C. estertheticum CF010	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CF013	+	–	+	+	+	+	+	+	–	+	–	–	+	–	+	–
C. estertheticum CF015	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CF016	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CM018	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CM020	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CM032	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CM033	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CM034	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum CM035	+	+	+	+	+	+	+	+	–	+	+	–	+	+	+	–
C. estertheticum DSM 14864	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	–

GLU, glucose; MAN, mannose; LAC, lactose; SAC, sucrose; MAL, maltose; SAL, salicin; XYL, xylose; ARA, arabinose; GLY, glycerol; CEL, cellobiose; MNE, mannose; MLZ, melezitose; RAF, raffinose; SOR, sorbitol; RHA, rhamnose; and TRE, trehalose.

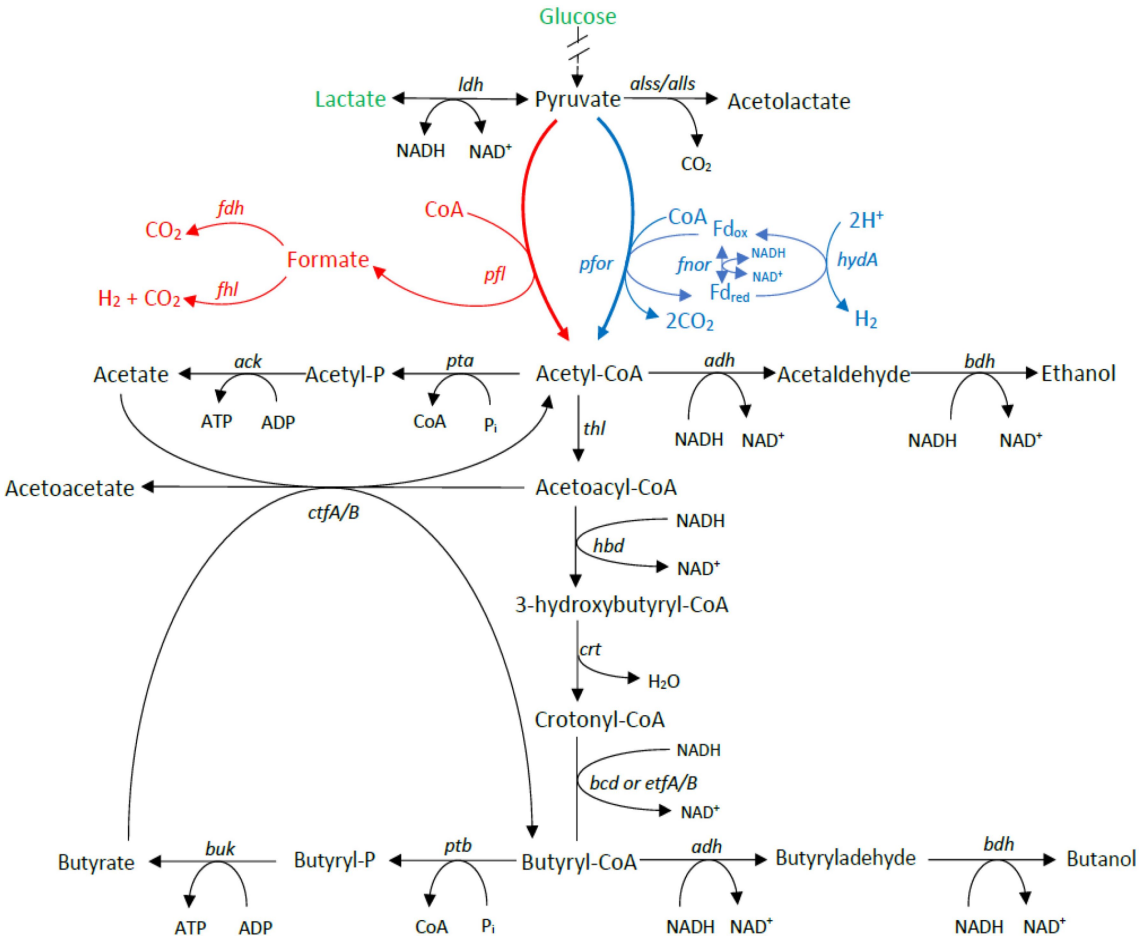


FIGURE 6 | Genome-based reconstruction of *C. estertheticum* and *C. tagluense* central fermentation pathway. Two key pathways involved in gas production are highlighted in red or blue colors. The main substrates utilized by both species in meat, glucose, and lactate are highlighted in green. *ldh*, L-lactate dehydrogenase; *alls* and *alss*, acetolactate synthase large and small subunits; *pfor*, NADP+ pyruvate ferrodoxin oxidoreductase (*fnor-nfn*); *pta*, phosphate acetyltransferase; *ack*, acetate kinase; *adh*, alcohol dehydrogenase; *thl*, thiolase; *hbd*, hydroxybutyryl-coA dehydrogenase; *crt*, crotonase; *bcd/etfAB*, butyryl-CoA dehydrogenase/electron transfer flavoprotein NAD+ ferredoxin, subunits alpha and beta; *ptb*, phosphotransbutyrylase; *buk*, butyrate kinase; *ctfA/B*, coA-transferase subunits A/B; *bdh*, butanol dehydrogenase; *pfl*, pyruvate formate lyase; and *hydA*, hydrogenase A.

not cause pack distention (**Figures 8A,B**). This was surprising since variable differences in pack distention by *C. estertheticum* have not been reported before. At the genomic level, the four *C. estertheticum* strains had *hyd*, *hya,* and *hyc* gene clusters while all four *C. tagluense* strains had *hyd*, *hya*, *hyb*, and *hyp* genes. *Clostridium estertheticum* CEST001, which did not cause pack distention also had the *hyp* genes, which led to the suggestion that the presence of the *hyp* gene cluster was linked with absence of pack distention. Further analysis of selected *C. estertheticum* strains with or without the *hyp* genes consistently showed *C. estertheticum* CM035 (*hyp* −) caused pack distention while *C. estertheticum* CF003 and CF007 (*hyp* +) did not cause pack distention despite extended incubation periods of 56 days (**Figure 8C**). Therefore, the differences in the intra- and interspecies distribution of hydrogenases gene clusters, and specifically the *hyp* cluster, may be associated with meat spoilage with or without gas production by *C. estertheticum* and *C. tagluense* species.

BPS Causing Clades of *Clostridium estertheticum* Possess Unique Spore Germination Genes

We finally took advantage of the large repertoire of *C. estertheticum* genomes ($n = 25$) to perform an intraspecies comparative genomics. Using a 80% BLAST identity cutoff, we showed *C. estertheticum* has a pan-genome of 20,111 genes. The core- and soft-core genomes comprised of 1,771 and 2,036 genes, respectively, while the shell- and cloud-genomes had 4,354 and 13,721 genes, respectively. The large accessory genome highly suggests a significant degree of genomic diversity within *C. estertheticum*.

We further created a maximum likelihood phylogenetic tree which revealed six *C. estertheticum* putative clades (**Supplementary Figure 8**). Interestingly, the *C. estertheticum* reference strains known to cause BPS, DSM 8809, DSM 14864, MA19, and MA41 and the BPS causing strains in the present study, CM020 and CM035 clustered together in Clade 5

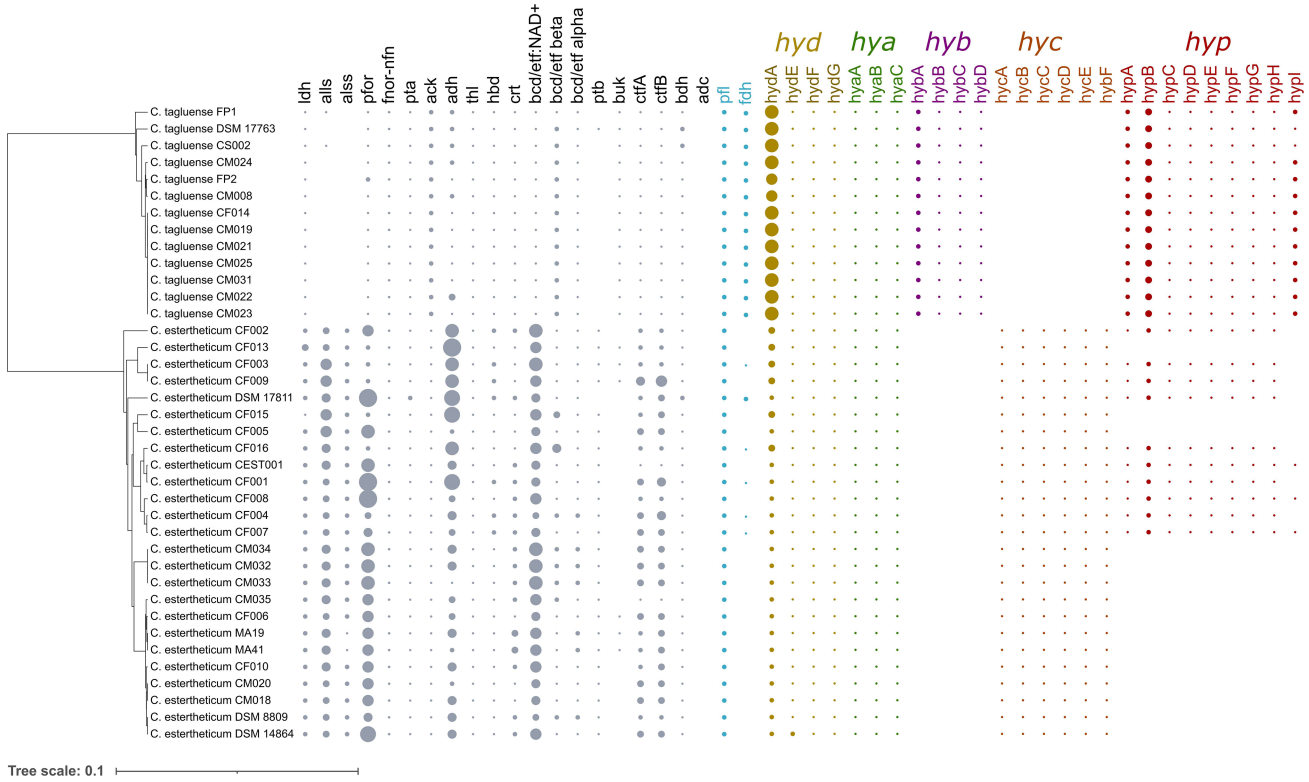


FIGURE 7 | Distribution of acetone–butyrate–ethanol fermentation pathway and hydrogenases genes within *C. estertheticum* and *C. tagluense*. The five hydrogenase gene clusters, *hyd*, *hya*, *hyb*, *hyc*, and *hyp* are highlighted in different colors. The size of each circle corresponds with the frequency of the gene. A blank shows the gene is absent. The phylogeny is based on the *rpoB* gene. *ldh*, L-lactate dehydrogenase; *alls* and *alss*, acetolactate synthase large and small subunits; *pfor*, $NADP^+$ pyruvate ferrodoxin oxidoreductase (*fnor-nfn*); *pta*, phosphate acetyltransferase; *ack*, acetate kinase; *adh*, alcohol dehydrogenase; *thl*, thiolase; *hbd*, hydroxybutyryl-coA dehydrogenase; *crt*, crotonase; *bcd/etfAB*, butyryl-CoA dehydrogenase/electron transfer flavoprotein NAD^+ ferredoxin, subunits alpha and beta; *ptb*, phosphotransbutyrylase; *buk*, butyrate kinase; *ctfA/B*, coA-transferase subunits A/B; *bdh*, butanol dehydrogenase; *adc*, acetone dehydrogenase; *pfl*, pyruvate formate lyase; and *hyd*, *hya*, *hyb*, *hyc*, and *hyp*, hydrogenase gene clusters.

(**Supplementary Figure 8**), which is *hyp* − (**Figure 7**). The three strains causing BPS cases in Switzerland CM032, CM033, and CM034 clustered in Clade 3, which is also *hyp* − (**Figure 7**). Because of the putative clades' relevance to BPS, we determined the presence of other important genes that are uniquely specific to these clades. Four genes annotated as "putative response regulatory protein," "Spore germination protein B2," "Spore germination protein B1," and "DegV domain-containing protein" were identified. The identification of the two spore germination proteins genes that are specific to these two putative clades suggests they have evolved different spore germination pathways.

DISCUSSION

CEC includes diverse psychrophilic and anaerobic species that are commonly isolated from the meat processing environment, and are collectively recognized as a major cause of meat spoilage in chilled and vacuum-packed meat with *C. estertheticum* being regarded as the most important among the named CEC species (Wambui and Stephan, 2019). Previously, *C. estertheticum* has been reported to cause BPS in a diverse range of meat including lamb, beef and pork (Collins, 1992; Byrne et al., 2009; Zhang et al., 2020) but for the first time, we have shown here that it can also cause spoilage in horse meat (**Figure 1C**). Interestingly, we noted that *C. estertheticum* CM032, which caused BPS in beef, and CM033 and CM034, which caused BPS in horse meat samples, were clonal (**Supplementary Figure 3**). This is also the first reported case of a *C. estertheticum* clone causing meat spoilage in two different types of meat and suggests a similar source of contamination. Therefore, this study not only demonstrates the suitability of whole-genome sequencing in the identification of CEC strains, but also determination of the relatedness of strains causing meat spoilage thus facilitating traceability and inference of the source of contamination.

So far, 10 CEC species had been identified (Palevich et al., 2021; Wambui et al., 2021a). The identification of strain CS001 which forms the novel species, genomospcies4 (**Figures 1A,B**), has expanded the number of species to 11 further revealing the diversity of CEC. Interestingly, strain CS001 forms a divergent branch within CEC (**Figures 1A,B**) suggesting the probable existence of yet unsampled intermediate CEC species between

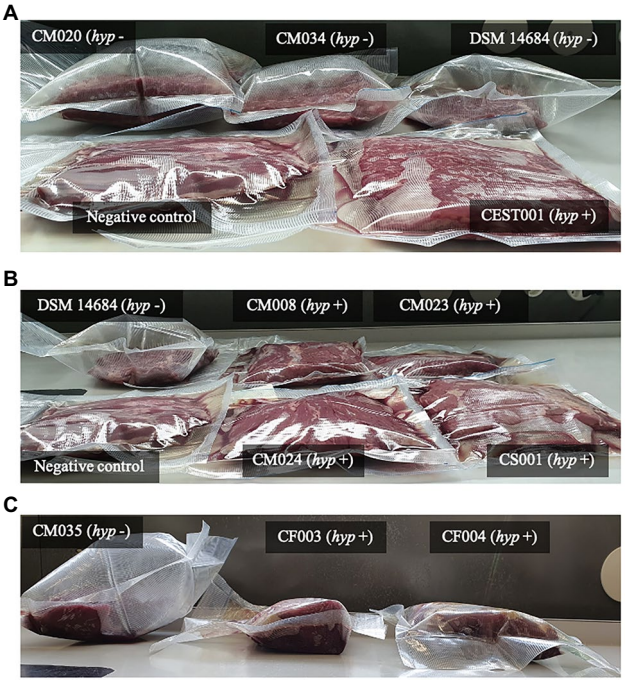


FIGURE 8 | Gas production in vacuum-packed beef at 8°C by *C. estertheticum* and *C. tagluense* with (*hyp* +) or without (*hyp* –) *hyp* hydrogenase gene cluster. **(A)** The *hyp* – *C. estertheticum* CM020, CM034, and DSM 14864 caused pack distention while the *hyp* + *C. estertheticum* CEST001 only caused loss of vacuum. **(B)** The *hyp* + *C. tagluense* CM008, CM023, CM024, and CS002 caused loss of vacuum. *Clostridium estertheticum* DSM 14864 (*hyp* –) was the positive control. **(C)** The *hyp* + *C. estertheticum* CF003 and CF004 did not cause pack distention even after 56 days of incubation compared to the *hyp* – *C. estertheticum* CM035. In **(A)** and **(B)**, the samples were incubated for 28 days.

genomspecies4 and other CEC species. The fact that the strain CS001 from the novel genomospecies4 species and other recently described species (**Table 1**) could also cause meat spoilage, further reinforces CEC's role in the spoilage of chilled vacuum-packed meat.

Difficulties in isolation and culturing of CEC strains have continuously limited the number of CEC genomes available for the inter- and intraspecies genomic characterization. This is evident in a previous study where only six strains from four CEC species were genomically characterized (Palevich et al., 2021). Accordingly, our work has not only increased the availability of CEC genomes, but has gone ahead to perform a comprehensive comparative genomic analysis that has revealed the main forces underpinning the genetic diversity and evolution of 11 CEC species. The analysis of recombination events has shown that although recombination is limited to closely related strains, the events are high and have an influence on the population structure of CEC, which was demonstrated by the reconstruction of the core-genome-based phylogeny of CEC before and after removal of recombined regions (**Figure 3**). This is consistent with a previous report that recombination can influence the bacterial phylogeny (van Rossum et al., 2020; Lin et al., 2021). Although we have identified recombination as one of the drivers of genetic diversity within CEC, gene gain and loss events are also recognized as major drivers for genetic diversity and result in the evolution of bacteria with unique genotypes and phenotypes (Li et al., 2017; van Rossum et al., 2020). In this regard, we have identified several candidate species-specific genes that have been lost by genomospecies2 and likely contributed to its reduced genome size and speciation. Among these genes were the flagellar assembly and signaling genes (**Figure 4A**). Flagella are particularly important for bacteria motility (Moens and Vanderleyden, 1996). Consistently, the absence of the genes corresponded with reduced motility of genomospecies2 strains (**Figure 4B**). The current analysis was carried out in broth culture, which might indicate swimming motility of genomospecies2 strains in liquid media is impaired. Further motility studies including swarming motility, which determines the movement of bacteria across a surface with the aid of flagella are proposed. We also noted that the absence of these genes corresponded with formation of larger and slimy colonies by genomospecies2 compared to other CEC species (**Figure 4C**) suggesting altered colony morphology and production of extracellular matrix. Similar results have been reported in *Vibrio cholerae* and it has been suggested that the absence of flagella constitutes a signal to increase exopolysaccharide synthesis (Watnick et al., 2001). Further studies will be required to identify genes responsible for the synthesis of the matrix by genomospecies2 and determination of its composition.

By exploiting the pan-genome of CEC, its genetic diversity was further unraveled and attributed to the large accessory genome. Accessory genes are important for fundamental bacterial functions, but can also determine the diversity of bacteria (Adeniji et al., 2019). Through the analysis of cluster of orthologous groups, carbohydrate metabolism and transport emerged as one of the key functions of the accessory genome (**Figures 2E,F**) while the analysis of gene presence/absence showed that the two CEC phylogroups differed with carbohydrate metabolism genes indicating they have evolved different saccharolytic abilities. This is particularly interesting because several studies have reported differences in fermentable substrate utilization among CEC members (Spring, 2003; Suetin et al., 2009; Yang and Badoni, 2013). In the current study, we have used a diverse range of strains from *C. estertheticum* and *C. tagluense* and confirmed previous reports that utilization of fermentable substrates by *C. estertheticum* is higher than in *C. tagluense* (Suetin et al., 2009). Using a higher repertoire of CEC strains and genomes than any previous study, we showed the phenotypes (**Table 2**) were consistent with the distribution of carbohydrate utilization genes (**Figure 5**) in both species. These data suggest *C. estertheticum* is highly adaptable to environment containing diverse carbohydrate substrates.

With respect to meat spoilage, the utilization of intramuscular carbohydrates by members of CEC has been reported to be an important link to the acetone–butyrate–ethanol fermentation pathway, which ultimately results in meat spoilage (Palevich et al., 2021). In the current study, we have gone ahead and determined the genetic basis for differences in gas production during meat spoilage by CEC

using *C. estertheticum* and *C. tagluense* as representative species. While meat spoilage accompanied by gas production by *C. estertheticum* could have been as a consequence of higher number of carbohydrates metabolizing genes (**Figure 5**), which corresponded with high range of substrate utilization compared to *C. tagluense* (**Table 2**), the revelation that some *C. estertheticum* strains also caused meat spoilage with less gas production (**Figures 8A,C**) did not fully support this hypothesis. Further analyses have however linked the gas production to the inter-genomic distribution of hydrogenases (**Figure 7**). Hydrogenases (H_2ases) are redox metalloenzymes that catalyze the reversible oxidation of molecular hydrogen (H_2; Schuchmann et al., 2018). Anaerobic bacteria, such as *Clostridium* spp., produce H_2 after the reduction of protons by H_2ases as means of disposing excess reducing equivalents while H_2 produced from other metabolic pathways can be recycled by uptake H_2ases (Vignais et al., 2001). Based on their metal content, there are three types of hydrogenase enzymes: mononuclear [Fe] and dinuclear [FeFe] and [NiFe] (Kaur-Ghumaan and Stein, 2014). Accordingly, we have shown that the [FeFe] and [NiFe] hydrogenases are encoded in the genomes of *C. estertheticum* and *C. tagluense* (**Figure 6**; **Supplementary Table 2**). The *hya* and *hyb* gene clusters encode [FeFe]-H_2ases while the *hyc* and *hyp* gene clusters encode [NiFe]-H_2ases gene clusters with each cluster consisting of a repertoire of genes necessary for the assembly and maturation a functional hydrogenases. These include genes necessary for catalytic activity of the hydrogenase, proteases involved in formation or maturation of subunits and membrane proteins involved in membrane binding (Trchounian et al., 2012).

The intragenomic presence of different [FeFe]-H_2ase and [FeFe]-H_2ase gene clusters in both *C. estertheticum* and *C. tagluense* suggest functional redundance of each type of hydrogenase. In *Clostridium* spp., this allows strains to adopt a wide range of environmental conditions (Calusinska et al., 2010). The [FeFe]-H_2ases are typically associated with H_2 production while [NiFe]-H_2ases are mainly associated with H_2 uptake (Peters et al., 2015). Interestingly, we have shown the presence *hyp* gene cluster, which encodes a putative [NiFe]-H_2ase corresponds with absence of pack distention in both *C. estertheticum* and *C. tagluense* (**Figure 8**). Similar to other characterized homologous *hyp* gene clusters (Buhrke et al., 2001), we hypothesize that the gene products of *C. estertheticum* and *C. tagluense hyp* gene clusters are responsible for the assembly and maturation of hydrogen oxidizing [NiFe]-H_2ases. Its absence in strains causing pack distention, such as *C. estertheticum* CM034 and DSM 14864 (**Figure 8A**), suggests these strains are incapable of reincorporating the hydrogen that is produced as a metabolite hence resulting in the accumulation and hence the phenotype that is characteristic of BPS. On the other hand, the strains with the *hyp* gene cluster, such as *C. estertheticum* CEST001 (**Figure 8A**), CF003, and CF004 (**Figure 8C**) and *C. tagluense* CM008, CM023, and CM024 (**Figure 8B**), are capable of recycling the hydrogen which prevents its accumulation in the packs hence lack of pack distention. The current lack of genetic tools amenable to CEC and the slow growth rates of CEC strains are a major limitation to the validation of these hypotheses through mutagenesis. Despite of these limitations, our study has provided a basis for future characterization of the *hyp* gene cluster.

The availability of 25 *C. estertheticum* genomes in the current study could provide some insights into the population structure and genetic diversity of the species. The pan-genome analysis demonstrated high genetic diversity within the species. Similarly, we identified six putative clades of the species (**Supplementary Figure 8**), two of which consist of strains known to cause BPS. Through intraspecies genomic comparison of *C. estertheticum* we identified that these two putative clades, which consist of *hyp* − strains, possess unique spore germination genes. Spore germination is critical process for the outgrowth and development of spoilage potential of sporulating bacteria (Delbrück et al., 2021). Therefore, further determination of spore germination process of the *hyp* − strains is necessary. Due to the limited number of genomes of the species, these clades are postulated to be putative and the current intraspecies comparative genomics are preliminary. More genomes are required for robust analysis to determine the most probable population structure and intraspecies genetic diversity of the species.

CONCLUSION

This study provides a comprehensive genomic analysis of the *C. estertheticum* complex (CEC), a diverse group of psychrophilic bacteria associated with the spoilage of chilled vacuum-packed meat. Following the isolation, identification, and sequencing of 15 new CEC strains, we have expanded the number of CEC species to 11 and the number of publicly available genomes to 50. Presently, we have shown that recombination and gene gain/loss events are important sources of natural variation within CEC. Accordingly, genomospecies2 has lost genes related to flagellar assembly and signaling, which influences its motility and colony morphology. Through pan-genome analysis and phenotypic validation using the *C. estertheticum* and *C. tagluense* as representative species, we have shown the variable carbohydrate utilization within CEC is linked with variable distribution of carbohydrate-active enzymes, and more specifically, the glycoside hydrolases. Our findings that some *C. estertheticum* strains caused meat spoilage without pack distention suggest that these differences do not fully account for differences in gas production in vacuum-packed meat by *C. estertheticum* and *C. tagluense*. However, an important finding of our study is the revelation that the inter- and intraspecies differences exhibited by the two species during meat spoilage are associated with the distribution of the [NiFe]-hydrogenase *hyp* gene cluster. The absence of the cluster is associated with pack distention and its presence is associated with lack of pack distention suggesting the hydrogenase catalyzes the oxidation of molecular hydrogen preventing its accumulation in vacuum packs. We propose mutagenesis of the [NiFe]-hydrogenase *hyp* gene cluster be carried out in future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

JW and RS designed the study. MS and NC carried out whole-genome sequencing and assembly. JW and MS carried out the bioinformatic analysis. JW carried out the phenotypic tests and wrote the initial draft manuscript. JW, MS, NC, and RS revised the final manuscript. RS supervised the study. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.856810/full#supplementary-material

REFERENCES

Adeniji, A. A., Loots, D. T., and Babalola, O. O. (2019). *Bacillus velezensis*: phylogeny, useful applications, and avenues for exploitation. *Appl. Microbiol. Biotechnol.* 103, 3669–3682. doi: 10.1007/S00253-019-09710-5/FIGURES/7

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed November 2, 2021).

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bassi, D., Puglisi, E., and Cocconcelli, P. S. (2015). Understanding the bacterial communities of hard cheese with blowing defect. *Food Microbiol.* 52, 106–118. doi: 10.1016/J.FM.2015.07.004

Boerema, J. A., Broda, D. M., Penney, N., and Brightwell, G. (2007). Influence of peroxyacetic acid–based carcass rinse on the onset of "blown pack" spoilage in artificially inoculated vacuum-packed chilled beef. *J. Food Prot.* 70, 1434–1439. doi: 10.4315/0362-028X-70.6.1434

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5, 1–6. doi: 10.1038/srep08365

Brightwell, G., and Clemens, R. (2012). Development and validation of a real-time PCR assay specific for *Clostridium estertheticum* and *C. estertheticum*-like psychrotolerant bacteria. *Meat Sci.* 92, 697–703. doi: 10.1016/j.meatsci.2012.06.025

Buhrke, T., Bleijlevens, B., Albracht, S. P. J., and Friedrich, B. (2001). Involvement of *hyp* gene products in maturation of the H_2-sensing [NiFe] hydrogenase of *Ralstonia eutropha*. *J. Bacteriol.* 183, 7087–7093. doi: 10.1128/JB.183.24.7087-7093.2001

Byrne, B., Monaghan, A. M., Lyng, J. G., Sheridan, J. J., and Bolton, D. J. (2009). A case of "blown pack" meat linked to *Clostridium estertheticum* in Ireland. *J. Food Saf.* 29, 629–635. doi: 10.1111/j.1745-4565.2009.00182.x

Calusinska, M., Happe, T., Joris, B., and Wilmotte, A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* 156, 1575–1588. doi: 10.1099/MIC.0.032771-0/CITE/REFWORKS

Cantalapiedra, C. P., Hern Andez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/MOLBEV/MSAB293

Collins, M. D. (1992). Taxonomic studies on a psychrophilic *Clostridium* from vacuum-packed beef: description of *Clostridium estertheticum* sp. nov. *FEMS Microbiol. Lett.* 96, 235–239. doi: 10.1016/0378-1097(92)90410-P

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using gubbins. *Nucleic Acids Res.* 43:e15. doi: 10.1093/NAR/GKU1196

Delbrück, A. I., Zhang, Y., Heydenreich, R., and Mathys, A. (2021). *Bacillus* spore germination at moderate high pressure: a review on underlying mechanisms, influencing factors, and its comparison with nutrient germination. *Compr. Rev. Food Sci. Food Saf.* 20, 4159–4181. doi: 10.1111/1541-4337.12789

Dorn-In, S., Mang, S., and Schwaiger, K. (2022). Unknown cold-tolerant *Clostridium* spp.: characteristics and potential to cause meat spoilage. *Food Microbiol.* 102:103916. doi: 10.1016/J.FM.2021.103916

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* 50, D571–D577. doi: 10.1093/NAR/GKAB1045

Gómez-Torres, N., Garde, S., Peirotén, Á., and Ávila, M. (2015). Impact of *Clostridium* spp. on cheese characteristics: microbiology, color, formation of volatile compounds and off-flavors. *Food Control* 56, 186–194. doi: 10.1016/J.FOODCONT.2015.03.025

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/NAR/GKY1085

Húngaro, H. M., Caturla, M. Y. R., Horita, C. N., Furtado, M. M., and Sant'Ana, A. S. (2016). Blown pack spoilage in vacuum-packaged meat: a review on clostridia as causative agents, sources, detection methods, contributing factors and mitigation strategies. *Trends Food Sci. Technol.* 52, 123–138. doi: 10.1016/J.TIFS.2016.04.010

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/NAR/GKV1070

Kaur-Ghumaan, S., and Stein, M. (2014). [NiFe] hydrogenases: how close do structural and functional mimics approach the active site? *Dalton Trans.* 43, 9392–9405. doi: 10.1039/C4DT00539B

Khamaiseh, E. I., Abdul Hamid, A., Abdeshahian, P., Wan Yusoff, W. M., and Kalil, M. S. (2014). Enhanced butanol production by *Clostridium acetobutylicum* NCIMB 13357 grown on date fruit as carbon source in P2 medium. *Sci. World J.* 2014:395754. doi: 10.1155/2014/395754

Lee, I., Chalita, M., Ha, S. M., Na, S. I., Yoon, S. H., and Chun, J. (2017). ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 67, 2053–2057. doi: 10.1099/ijsem.0.001872

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, S., Huang, L., Ke, C., Pang, Z., and Liu, L. (2020). Pathway dissection, regulation, engineering and application: lessons learned from biobutanol production by solventogenic clostridia. *Biotechnol. Biofuels* 13:39. doi: 10.1186/S13068-020-01674-3

Li, N., Wang, K., Williams, H. N., Sun, J., Ding, C., Leng, X., et al. (2017). Analysis of gene gain and loss in the evolution of predatory bacteria. *Gene* 598, 63–70. doi: 10.1016/J.GENE.2016.10.039

Lin, E., Zou, S., Wang, Y., Lee, C. C., Chiu, C. H., and Feng, Y. (2021). Phylogeny, recombination, and invasiveness of group B *Streptococcus* revealed by genomic comparisons of its global strains. *Eur. J. Clin. Microbiol. Infect. Dis.* 40, 581–590. doi: 10.1007/S10096-020-04067-4/FIGURES/2

Moens, S., and Vanderleyden, J. (1996). Functions of bacterial flagella. *Crit. Rev. Microbiol.* 22, 67–100. doi: 10.3109/10408419609106456

Nychas, G.-J. E., Skandamis, P. N., Tassou, C. C., and Koutsoumanis, K. P. (2008). Meat spoilage during distribution. *Meat Sci.* 78, 77–89. doi: 10.1016/J.MEATSCI.2007.06.020

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/BIOINFORMATICS/BTV421

Palevich, N., Palevich, F. P., Maclean, P. H., Altermann, E., Gardner, A., Burgess, S., et al. (2021). Comparative genomics of *Clostridium* species associated with vacuum-packed meat spoilage. *Food Microbiol.* 95:103687. doi: 10.1016/j.fm.2020.103687

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Patakova, P., Linhova, M., Rychtera, M., Paulova, L., and Melzoch, K. (2013). Novel and neglected issues of acetone–butanol–ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. *Biotechnol. Adv.* 31, 58–67. doi: 10.1016/J.BIOTECHADV.2012.01.010

Peters, J. W., Schut, G. J., Boyd, E. S., Mulder, D. W., Shepard, E. M., Broderick, J. B., et al. (2015). [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim. Biophys. Acta* 1853, 1350–1369. doi: 10.1016/j.bbamcr.2014.11.021

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/c5ay02550h

Schuchmann, K., Chowdhury, N. P., and Müller, V. (2018). Complex multimeric [FeFe] hydrogenases: biochemistry, physiology and new opportunities for the hydrogen economy. *Front. Microbiol.* 9:2911. doi: 10.3389/FMICB.2018.02911/BIBTEX

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/BIOINFORMATICS/BTU153

Snipen, L., and Liland, K. H. (2015). Micropan: An R-package for microbial pan-genomics. *BMC Bioinformatics* 19:79. doi: 10.1186/s12859-015-0517-0

Spring, S. (2003). Characterization of novel psychrophilic clostridia from an Antarctic microbial mat: description of *Clostridium frigoris* sp. nov., *Clostridium lacusfryxellense* sp. nov., *Clostridium bowmanii* sp. nov. and *Clostridium psychrophilum* sp. nov. and reclassification of *Clostridium laramiense* as *Clostridium estertheticum* subsp. *laramiense* subsp. nov. *Int. J. Syst. Evol. Microbiol.* 53, 1019–1029. doi: 10.1099/ijs.0.02554-0

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/BIOINFORMATICS/BTU033

Suetin, S. V., Shcherbakova, V. A., Chuvilskaya, N. A., Rivkina, E. M., Suzina, N. E., Lysenko, A. M., et al. (2009). *Clostridium tagluense* sp. nov., a psychrotolerant, anaerobic, spore-forming bacterium from permafrost. *Int. J. Syst. Evol. Microbiol.* 59, 1421–1426. doi: 10.1099/ijs.0.002295-0

Trchounian, K., Poladyan, A., Vassilian, A., and Trchounian, A. (2012). Multiple and reversible hydrogenases for hydrogen production by *Escherichia coli*: dependence on fermentation substrate, pH and the F_0 F_1 -ATPase. *Crit. Rev. Biochem. Mol. Biol.* 47, 236–249. doi: 10.3109/10409238.2012.655375

Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. (2014). The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15:524. doi: 10.1186/S13059-014-0524-X/TABLES/4

van Rossum, T., Ferretti, P., Maistrenko, O. M., and Bork, P. (2020). Diversity within species: interpreting strains in microbiomes. *Nat. Rev. Microbiol.* 18, 491–506. doi: 10.1038/s41579-020-0368-1

Vignais, P. M., Billoud, B., and Meyer, J. (2001). Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* 25, 455–501. doi: 10.1111/j.1574-6976.2001.tb00587.x

Wambui, J., Cernela, N., Corti, S., and Stephan, R. (2020a). Comparative genome analysis and phenotypic characterization of *Clostridium gasigenes* CGAS001 isolated from chilled vacuum-packed lamb meat. *Front. Microbiol.* 11:2048. doi: 10.3389/FMICB.2020.02048

Wambui, J., Cernela, N., Stevens, M. J. A., and Stephan, R. (2020b). Draft genome sequence of *Clostridium estertheticum* CEST001, belonging to a novel subspecies of *C. estertheticum*, isolated from chilled vacuum-packed lamb meat imported to Switzerland. *Microbiol. Resour. Announc.* 9, 13–14. doi: 10.1128/MRA.00806-20

Wambui, J., Cernela, N., Stevens, M. J. A., and Stephan, R. (2021a). Whole genome sequence-based identification of *Clostridium estertheticum* complex strains supports the need for taxonomic reclassification within the species *Clostridium estertheticum*. *Front. Microbiol.* 12:727022. doi: 10.3389/fmicb.2021.727022

Wambui, J., Ghielmetti, G., Morach, M., Hochreutener, M., and Stephan, R. (2021b). Detection of psychrophilic *Clostridium* spp. in fecal samples from cattle of different ages sampled at the slaughterhouse level. *J. Food Prot.* 84, 58–62. doi: 10.4315/JFP-20-259

Wambui, J., Püntener, S., Corti, S., Cernela, N., and Stephan, R. (2020c). Detection of psychrophilic *Clostridium* spp. causing "blown pack" spoilage in meat juice samples from chilled vacuum-packed beef and lamb meat imported from different countries to Switzerland. *J. Food Prot.* 83, 56–59. doi: 10.4315/0362-028X.JFP-19-321

Wambui, J., and Stephan, R. (2019). Relevant aspects of *Clostridium estertheticum* as a specific spoilage organism of vacuum-packed meat. *Microorganisms* 7:142. doi: 10.3390/microorganisms7050142

Wambui, J., Stevens, M. J. A., Sieber, S., Cernela, N., Perreten, V., and Stephan, R. (2022). Targeted genome mining reveals the psychrophilic *Clostridium estertheticum* complex as a potential source for novel bacteriocins, including cesin A and estercticin A. *Front. Microbiol.* 12:801467. doi: 10.3389/fmicb.2021.801467

Watnick, P. I., Lauriano, C. M., Klose, K. E., Croal, L., and Kolter, R. (2001). The absence of a flagellum leads to altered colony morphology, biofilm development and virulence in *Vibrio cholerae* O139. *Mol. Microbiol.* 39, 223–235. doi: 10.1046/J.1365-2958.2001.02195.X

Yang, X., and Badoni, M. (2013). Substrate utilization during incubation in meat juice medium of psychrotolerant clostridia associated with blown pack spoilage. *Food Microbiol.* 34, 400–405. doi: 10.1016/j.fm.2013.02.003

Yang, X., Balamurugan, S., and Gill, C. O. (2009). Substrate utilization by *Clostridium estertheticum* cultivated in meat juice medium. *Int. J. Food Microbiol.* 128, 501–505. doi: 10.1016/j.ijfoodmicro.2008.10.024

Yang, X., Gill, C. O., and Balamurugan, S. (2010). Products of glucose and lactate fermentation, and utilization of amino acids by *Clostridium estertheticum* subspp. *laramiense* and *estertheticum* growing in meat juice medium. *J. Food Prot.* 73, 1348–1352. doi: 10.4315/0362-028X-73.7.1348

Yu, J., Zhao, J., Song, Y., Zhang, J., Yu, Z., Zhang, H., et al. (2018). Comparative genomics of the herbivore gut symbiont *Lactobacillus reuteri* reveals genetic diversity and lifestyle adaptation. *Front. Microbiol.* 9:1151. doi: 10.3389/FMICB.2018.01151/BIBTEX

Zhang, P., Ward, P., McMullen, L. M., and Yang, X. (2020). A case of 'blown pack' spoilage of vacuum-packaged pork likely associated with *Clostridium estertheticum* in Canada. *Lett. Appl. Microbiol.* 70, 13–20. doi: 10.1111/lam.13236

Zheng, Y. N., Li, L. Z., Xian, M., Ma, Y. J., Yang, J. M., Xu, X., et al. (2009). Problems with the microbial production of butanol. *J. Ind. Microbiol. Biotechnol.* 36, 1127–1138. doi: 10.1007/S10295-009-0609-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fmicb.2022.850720

The Microbial Quality of Commercial Chopped Romaine Lettuce Before and After the "Use By" Date

*Chao Liao and Luxin Wang**

Department of Food Science and Technology, University of California, Davis, Davis, CA, United States

In the United States, due to the limited information about the microbial quality and safety of fresh produce after the labeled open dates, unnecessary discarding of fresh produce in good conditions and food loss have been caused. The aim of this study was to address this knowledge gap and evaluate the microbial quality of commercial chopped Romaine lettuce (RL) on the "Use By" dates (UBD) and 5 days after the "Use By" dates (UBD5). The microbial quality was evaluated using culture-dependent and culture-independent methods. Three brands of RL samples, from early and late harvest seasons, were purchased from local grocery stores and stored at 4°C until 5 days after their UBD. On the UBD and UBD5, bagged lettuce was opened, homogenized, diluted, and plated onto plate count agar and anaerobic agar to obtain total aerobic plate counts (APC) and total anaerobic plate counts (AnPC). For the culture-independent method, DNA was extracted from each sample homogenate and used for 16S rRNA gene sequencing. The culture-dependent results showed that there was no significant change in APC or AnPC between UBD and UBD5 samples. The APC and AnPC ranged from 5.71 ± 0.74 to 7.89 ± 0.10 Log CFU/g and 1.75 ± 0.08 to 7.32 ± 0.61 Log CFU/g, respectively. No significant difference in alpha diversity, based on observed features and Shannon index values, was detected between UBD and UBD5 samples using 16S rRNA sequencing. Similarly, no difference was observed in beta diversity based on the Jaccard distance matrixes and the weighted Unifrac distance matrixes. Taxonomic analysis revealed 128 genera in all RL samples. The top five genera were *Pseudomonas* (with relative abundance ranging from 16.47 to 92.72%), *Serratia* (0–52.35%), *Weissella* (0–42.42%), *Pantoea* (0.17–21.33%), and *Lactococcus* (0–24.30%). The differential abundance analysis based on the ANCOM test showed that no bacteria were detected to have significantly differential abundance in RL between UBD and UBD5. In summary, both the culture-dependent and culture-independent results showed that there was no significant difference in the microbial quality of RL before and shortly after the UBD.

Keywords: Romaine lettuce, "Use By" date, microbial food quality, plate counting, 16S rRNA gene sequencing, bacterial communities

OPEN ACCESS

Edited by:
Nicholas Bokulich,
ETH Zürich, Switzerland

Reviewed by:
Karl Matthews,
Rutgers, The State University of New Jersey, United States
John Phillip Bowman,
University of Tasmania, Australia

***Correspondence:**
Luxin Wang
lxwang@ucdavis.edu

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *08 January 2022*
Accepted: *07 March 2022*
Published: *11 April 2022*

Citation:
Liao C and Wang L (2022) The Microbial Quality of Commercial Chopped Romaine Lettuce Before and After the "Use By" Date. Front. Microbiol. 13:850720. doi: 10.3389/fmicb.2022.850720

INTRODUCTION

Although fruits and leafy greens are essential components of a healthy diet, they are also one of the most wasted foods, with up to 63% of fresh fruits and 70% of leafy greens wasted (Mijares et al., 2021). The United States Department of Agriculture (USDA) estimated that approximately 31% of the 430 billion pounds of food produced in 2010 (valued at ca. $161.6 billion) was lost and not available for human consumption at the retail and consumer levels (Buzby et al., 2014). Vegetables (19%) ranked second on the list of food groups with high loss rates (Buzby et al., 2014). In the USDA and Environmental Protection Agency (2015) announced that the US 2030 Food Loss and Waste Reduction goal was to reduce 50% of food loss and waste by the year 2030 (Environmental Protection Agency, 2015; United States Department of Agriculture, 2015). There are many causes and drivers for food loss and waste at the retail level; consumers' confusion over "Use By" and "Best Before" dates and other date labeling is one of them (Newsome et al., 2014; Buzby et al., 2015). Date labeling has been practiced for decades, and its main aims include informing stock rotation at retail and facilitating potential recalls and traceability. However, the terminologies used vary widely around the world (Newsome et al., 2014).

Such variations in date labeling terms have contributed to substantial misunderstanding by the industry and consumers and led to significant unnecessary food loss and waste, especially in the United States (Newsome et al., 2014). Recent studies showed that approximately 37% of American consumers usually discarded fresh produce when the open date is past even when the product had no quality and safety problems (Leib et al., 2016; Wilson et al., 2017). In the United States, three date labels, namely, "Best if Use By" or "Best Before," "Use By," and "Sell By," are the most applied for food products (Wilson et al., 2017). Tsiros and Heilman (2005) summarized the three commonly used label phrases: (1) "Best before," which indicates the date after which a product loses the best quality for consumption. (2) "Use By," indicating the date after which a product is no longer remaining sufficient quality and should not be consumed, but not necessarily associated with food safety. (3) "Sell by," suggesting the last day on which a product should be sold (Tsiros and Heilman, 2005; Wilson et al., 2017). Among these three phrases, "Use By" generates the greatest value of predicted waste (Wilson et al., 2017).

In contrast to the United States, Regulation No. 1169/2011 of the European Parliament and the Council of the European Union (EU) mandates defined the use of date of minimum durability or best before date, and "Use By" date (European Commission, 2011). Foods that are highly perishable and likely to constitute an immediate danger to human health should carry a "Use By" date (UBD), after which date the "food shall be deemed unsafe" (European Food Safety Authority Panel on Biological Hazards [BIOHAZ] et al., 2020). In Australia and New Zealand, date labeling with a best before date or "Use By" date is required for most packaged foods with a shelf life of less than 2 years (Australian Government, 2012). "Best before" date indicates "the last date on which you can expect a food to retain all of its quality attributes, provided it has been stored according to any stated storage conditions and the package is unopened" (Food Standards Australia New Zealand, 2013). UBD is "the last date on which the food may be eaten safely, provided it has been stored according to any stated storage conditions and the package is unopened" and the food should not be eaten due to the health and safety reasons after the date (Food Standards Australia New Zealand, 2013).

Microorganisms play significant roles in food spoilage and loss (Lorenzo et al., 2018). The microbiological quality and safety of fresh produce are determined by factors from pre-harvesting to post-harvest. To better characterize the microorganisms present in fresh produce during storage, a number of research projects have applied culture-dependent and culture-independent methods to investigate the microbiological quality of fresh produce during storage (Jeddi et al., 2014; Berthold-Pluta et al., 2017; Cruz et al., 2019; Arienzo et al., 2020). For instance, Jeddi et al. (2014) enumerated the total aerobic bacteria present in 116 samples of fresh-cut vegetables, ready-to-eat salads, and wheat and mung bean sprouts before their open dates. The total aerobic mesophilic bacterial counts were 5.3–7.5, 5.5–7.4, 5.5–8.4, and 6.4–8.5 Log CFU/g, respectively for fresh-cut vegetable, ready-to-eat salads, and wheat and mung bean sprouts, respectively. Similarly, Berthold-Pluta et al. (2017) reported that the total aerobic mesophilic bacterial counts in lettuces, sprouts, and non-pasteurized fruits and vegetables during cold storage (below 4°C) before the "Best Before" dates were in ranges of 5.6–7.6, 6.8–8.4, and 2.9–7.7 Log CFU/g, respectively.

By using 16S rRNA sequencing, Leff and Fierer (2013) profiled the bacterial communities present on 11 produce varieties purchased from grocery stores, such as apples, grapes, lettuce, mushrooms, peaches, peppers, spinach, strawberries, tomatoes, alfalfa sprouts, and mung bean sprouts. Their results showed that the bacterial communities associated with each product type differed remarkably from each other, but certain produce types appeared to share more similar bacterial communities, such as sprouts, spinach, lettuce, tomatoes, peppers, and strawberries. Keshri et al. (2019) determined the dynamics of the bacterial communities of fresh alfalfa and mung bean sprouts right after purchase and every week until the end of their shelf-life. They found that both sprout types contained spoilage-related bacteria, *Pseudomonas* and *Pantoea*, at the time of purchase. The abundance of *Pseudomonas* increased in alfalfa after 3 weeks of cold storage at 4°C, while *Pantoea* dominated in mung bean sprouts after 2 weeks of storage at the same temperature. While previous studies provide important insights into the microbial population composition and potential changes during the shelf life of fresh produce, most of their sampling ended on the UBD. No information is available about bacterial populations present in fresh produce after the UBD. However, such information is very much needed for the industry as well as consumers to be better informed about what happens to the microbial quality after the UBD.

To fill the above knowledge gap, the objectives of this study were to investigate the bacterial populations of three brands of bagged chopped Romaine lettuce (RL) purchased during different seasons (early season *vs.* late season) and then evaluate the

microbial quality of RL on the UBD and 5 days after the "Use By" dates (UBD5), using both culture-dependent and culture-independent methods.

MATERIALS AND METHODS

Romaine Lettuce Samples

Three brands (Brands A, B, and C) of commercial chopped RL products were purchased from local grocery stores during the early (September–October) and late (March–April) harvest seasons (Williams et al., 2013). Three batches of RL for each season were purchased for the triplicate experiment. Upon arrival at the laboratory, all samples were stored at 4°C and sampled on their labeled UBD and UBD5. Photos were also taken for these samples on each sampling day (**Supplementary Figure 1**). As shown in **Supplementary Figure 1**, no significant visual differences were observed between UBD and UBD5.

Culture-Based Analysis of the Microbial Quality of Romaine Lettuce

At each sampling point presenting different shelf lives (UBD or UBD5), 80 g of RL was taken from each bag and added in a 55 oz Whirl-Pak filter bag (Nasco, Fort Atkinson, WI, United States) with 320 ml of phosphate-buffered saline (PBS, pH 7.4). The filter bag with RL samples was homogenized by using the Smasher^{TM} Lab Blender (AES-Chemunix, Bruz, France) for 120 s at the fast speed (620 strokes/min). One milliliter of the homogenate was serially diluted in 9 ml of PBS in 15 ml Falcon tubes (VWR, Atlanta, GA, United States), and four 100 μl of each dilute was plated on two plate count agar (PCA, BD Biosciences, Sparks, MD, United States) and two anaerobic agar (AA, BD Biosciences). PCA plates were incubated at 37°C for 48 h before colony enumeration. Anaerobic agar count medium supplemented with 2 mg/L of Methylene blue as the indicator of oxygen levels (light blue color for the presence of oxygen and yellowish color for the absence of oxygen) was used for culturing and enumerating anaerobes. Plated AA agar was incubated in Mitsubishi Gas Chemical (MGC) AnaeroPack-Jars with two MGC AnaeroPack-Anaero packs (Mitsubishi Gas Chemical Co., Tokyo, Japan) in each jar at 37°C for 48 h before enumeration (Liao and Wang, 2021). The generated total aerobic plate counts (APC) and total anaerobic plate counts (AnPC) were calculated and expressed as mean ± standard deviation (*SD*) Log CFU/g. The limit of detection of this plating method was 1.7 Log CFU/g of lettuce.

DNA Extraction and 16S rRNA Gene Sequencing

From each homogenate prepared above at each sampling point, 10 ml of the homogenate was transferred into a 15 ml Falcon tube (VWR, Atlanta, GA, United States) and centrifuged at 3,000 × *g* for 10 min at 4°C by using the Eppendorf Centrifuge 5810 R (Eppendorf, Hamburg, Germany) to collect the cell pellets. Each cell pellet was then washed twice by using 10 ml of PBS *via* centrifugation. Each washed pellet was then re-suspended with 1 ml of PBS and transferred into a 1.5 ml microcentrifuge tube (VWR, Atlanta, GA, United States). DNA was extracted from these pellets by using the DNeasy Powersoil kit (Qiagen, Gaithersburg, MD, United States) following the manufacturer's instructions. A total of 36 DNA samples were extracted and stored at −80°C for the subsequent 16S rRNA sequencing process. The library construction was carried out based on the amplification of V3–V4 region of the 16S rRNA gene (341F: 5′-CCTACGGGNGGCWGCAG-3′ and 785R: 5′-GACTACHVGGGTATCTAATCC-3′). Every PCR amplicon was tagged with forward and reverse barcodes (7 bp) (Sinclair et al., 2015; Liao and Wang, 2021). The sequencing was performed using the Illumina® MiSeq instrument with the MiSeq Reagent Kit v3 (Illumina, San Diego, CA, United States) to produce 2 × 300 bp paired-end reads. Library preparation and sequencing were conducted at the HudsonAlpha Genomic Service Laboratory (Huntsville, AL, United States).

Microbiome Data Analysis

The 16S rRNA gene sequencing data were processed using the Quantitative Insights Into Microbial Ecology (QIIME 2 version 2021.4) pipeline (Bolyen et al., 2019). De-multiplexed sequences were obtained from the Illumina BaseSpace platform by assigning reads to each sample based on sample-specific barcodes. For quality control, barcodes and primers were trimmed from the raw sequences using a q2-cutadapt plugin (Martin, 2011), followed by the denoising process to filter out the noisy reads, remove chimeric and singleton sequences, join denoised pair-end sequences, and cluster sequences using q2-dada2 plugin (Callahan et al., 2016). The principle of the DADA2 plugin in the QIIME 2 pipeline (q2-dada2 plugin) was based on the interactive quality plots of the Phred score as a function of each base in paired-end sequences (**Supplementary Figure 2**). The base position after which the median Phred quality scores of bases dropping below 30 was applied as the cut-off point for truncation of sequences in the same length (Estaki et al., 2020). With parameters "–p-trunc-len-f" (truncating length for forward reads), "–p-trunc-len-r" (truncating length for reverse reads), "–p-max-ee-f" (the number of maximum expected errors for forward reads), and "–p-max-ee-r" (the number of maximum expected errors for reverse reads) set as 260, 185, 2, and 4, respectively, approximately 62% of the sequences with median Pred quality score greater than 30 were kept for the downstream sequence analysis (**Supplementary Table 1**).

In this step, the amplicon sequence variants (ASVs) were clustered based on 100% of sequence similarity (Callahan et al., 2017), and the feature table of ASVs with frequency and feature data of representative ASVs was generated. A phylogenetic tree was constructed based on the feature data by using the align-to-tree-mafft-fasttree plugin (Katoh and Standley, 2013). Diversity analyses were conducted by using the core-metrics-phylogenetic plugin (Caporaso et al., 2010) with the parameter of sampling depth set as the minimum library size across all samples (80,938 reads). Alpha diversity was evaluated based on the observed features (ASVs) and Shannon index values (Kõiv et al., 2019). The beta diversity was evaluated by using the Jaccard distance matrix and the weighted Unifrac distance matrix (Lozupone et al., 2011; Edgar, 2017). Beta diversity was then visualized *via*

the two-dimensional principal coordinate analysis (PCoA) plots. Taxonomic analysis was conducted at the genus level using the classify-sklearn plugin (Bokulich et al., 2018a,b), which employed the pre-trained Naive Bayes classifier based on SILVA 138 small subunit rRNA[1] database (Quast et al., 2013; Bokulich et al., 2018a).

Statistical Analyses

All experiment trials were carried out in three independent replicates. Analysis of variance (ANOVA) and Tukey's Honestly Significant Difference (HSD) test (Abdi and Williams, 2010) were applied to analyze the mean differences of bacterial amounts (APC or AnPC) obtained from RL of different brands (Brands A, B, and C) purchased during different seasons (early *vs*. late season), as well as at different points in their shelf lives (UBD *vs*. UBD5). The non-parametric Wilcoxon rank sum test and the Kruskal–Wallis test were applied to test the difference between alpha diversities among groups (Xia and Sun, 2017). The permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2014) was used for analyzing the difference of beta diversities and the impact of shelf life, brand, and harvest season on beta diversity. Analysis of composition of microbiomes (ANCOM) (Mandal et al., 2015) was applied to identify bacteria that had significantly differential abundance between groups when samples were grouped and compared by shelf life, brands, or harvest seasons. The identification of these bacteria was based on the calculation of the centered log ratio (clr) F statistic and the W statistic. The clr F statistic measured the differences in effect size between groups based on the clr transformed bacterial abundance data, while the W statistic represented the number of times the log-ratio of a taxon with every other taxon being tested was identified to be significantly different across groups (Stevens et al., 2019). All the above analyses were performed using R scripts (version 4.1.0). Differences were considered statistically significant when the probability (p) value was less than 0.05.

RESULTS

Changes in Aerobic Plate Counts and Anaerobic Plate Counts in Romaine Lettuce on the UBD and UBD5

As shown in **Figure 1**, the APC of RL samples ranged from 5.71 ± 0.74 to 7.27 ± 0.20 Log CFU/g when purchased and sampled in the early season (**Figure 1A**) and ranged from 6.61 ± 0.79 to 7.89 ± 0.10 Log CFU/g in the late season (**Figure 1B**). At each harvest season, no difference in APC was detected between samples plated on the UBD and UBD5. When comparing the APC among three brands, brand C (6.98 ± 0.46 Log CFU/g in the early season and 7.57 ± 0.56 Log CFU/g in the late season) had statistically higher APC counts than brand B (5.75 ± 0.61 Log CFU/g in the early season and 6.65 ± 0.74 Log CFU/g in the late season). The AnPC of RL ranged from 1.75 ± 0.08 to 6.05 ± 0.44 Log CFU/g in samples from the early season (**Figure 1C**) and 2.26 ± 0.97 to 7.32 ± 0.61 Log CFU/g in samples from the late-season (**Figure 1D**). No significant difference in RL AnPC was detected between UBD and UBD5 except for brand B when samples were tested in the late season. The AnPC of brand B was 2.26 ± 0.97 Log CFU/g on UBD and was 4.34 ± 0.84 Log CFU/g on UBD5 when sampled in the late season. For samples purchased and analyzed in the early season, regardless of shelf-life, brand C had the highest AnPC among all three brands (5.98 ± 0.57 Log CFU/g), while brand B showed the lowest AnPC (1.97 ± 0.36 Log CFU/g). For samples from the late harvest season, brand B had the lowest AnPC (3.30 ± 1.40 Log CFU/g) among all three brands.

Amplicon Sequence Variant Identification and the Diversity Analyses of Bacterial Communities

A total of 36 DNA samples were sequenced using 16S rRNA sequencing. With the denoising process in QIIME 2 pipeline, 1,179 ASVs were identified from the total 6,051,758 sequences (**Supplementary Table 1**). The minimum sequence frequency, 80,938, was chosen for rarefying the reads across all samples to avoid false positive detection (Saary et al., 2017; **Supplementary Figure 3**). The impacts of shelf life (UBD and UBD5), brand (brands A, B, and C), and harvest season (early and late season) on the alpha and beta diversity of bacterial communities were analyzed by using the QIIME 2 pipeline.

The alpha diversity of bacterial communities was evaluated by measuring observed features (ASVs) and the Shannon index values (**Figure 2**). As shown in **Figure 2A**, the numbers of the observed features (left) and the Shannon index values (right) showed no difference between UBD and UBD5 ($p > 0.05$). When comparing alpha diversity among the three brands, brand C had 131 ± 18 observed features (**Figure 2B** left), which was remarkably higher than the observed features from brands A and B ($P = 0.043$ when comparing brand A and brand C; $p = 0.0072$ when comparing brands B and C based on the Wilcoxon rank sum test). No difference of observed features was detected between brands A and C. Brand C samples also had the highest Shannon index value of 3.95 ± 0.80, compared with brand A (3.09 ± 0.82, $p = 0.01$) and brand B (2.19 ± 0.71, $p = 5 \times 10^{-5}$) (**Figure 2B**, right). Brand A had a higher Shannon index value than brand B ($p = 0.012$). The Kruskal–Wallis test showed that the factor of "brand" pronouncedly impacted the observed features ($p = 0.017$) and Shannon index values ($p = 1.7 \times 10^{-4}$) of RL bacterial communities. When focusing on the impact of harvest seasons, early harvest-season RL had higher observed features (129 ± 17) than late-season RL (104 ± 21) ($p = 6.3 \times 10^{-4}$, **Figure 2C** left). However, such impact was not detected when focusing on the Shannon index values (**Figure 2C** right).

Changes in the alpha diversity of bacterial communities in each brand of RL between UBD and UBD5 were also studied (**Supplementary Figure 4**). It can be seen that the brand A RL had a significant increase of observed feature numbers from UBD to UBD5, while no difference of Shannon index values was

[1]https://www.arb-silva.de/

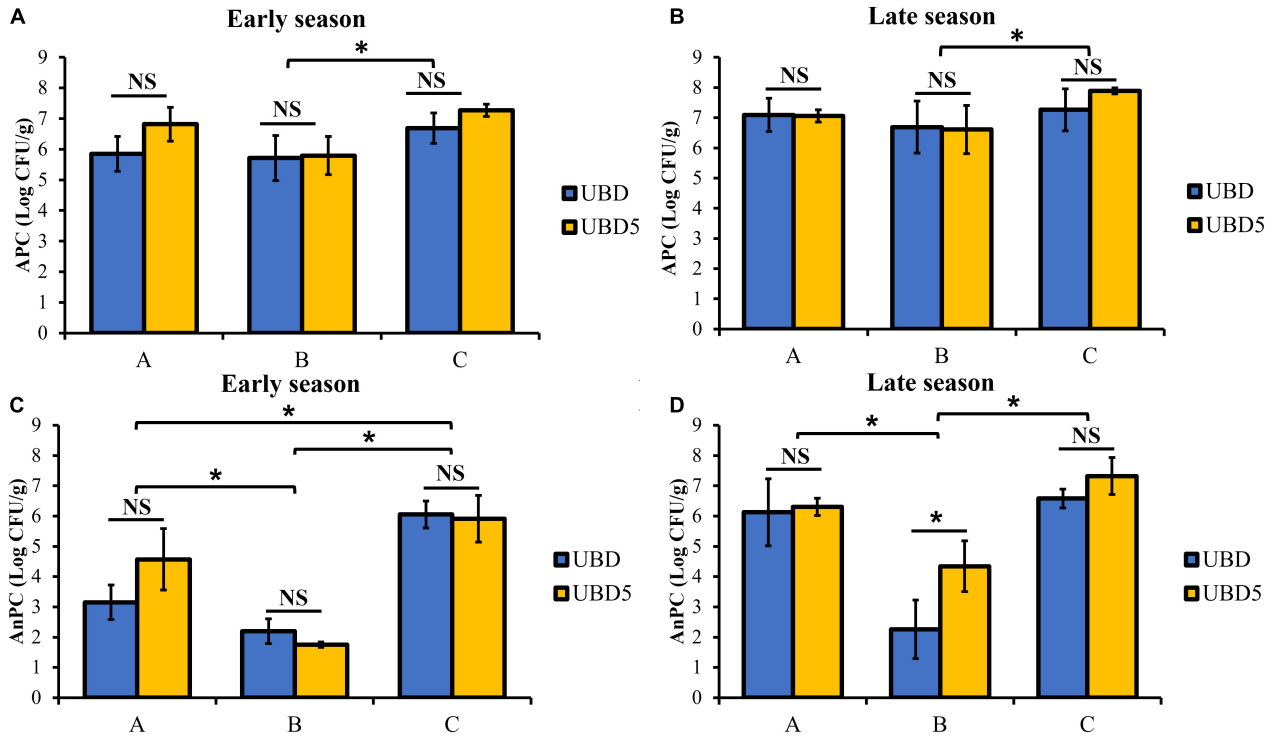


FIGURE 1 | Total aerobic plate count (APC, panels **A,B**) and anaerobic bacterial counts (AnPC, panels **C,D**) of three brands of RL purchased in the early and late seasons and analyzed on the their "Use By" dates (UBD) and 5 days after the "Use By" dates (UBD5). *Represents significant differences of APC or AnPC between RL samples of three brands or between UBD and UBD5 ($p < 0.05$).

observed between RL on the UBD and UBD5. Brand B RL had no changes in neither observed feature number nor Shannon index value from UBD to UBD5. For the brand C RL, Shannon index value had a remarkable increase from UBD to UBD5, while no difference of observed feature numbers was detected between UBD and UBD5.

The beta diversity of the RL bacterial community was calculated based on the Jaccard distance matrix and the weighted Unifrac distance matrix. PCoA was used to visualize the beta diversity of bacterial communities in RL samples grouped by shelf life, brands, or seasons. The PERMANOVA test was employed to statistically evaluate the impacts of shelf life, brands, or seasons on the beta diversity. In general, PCoA plots based on the Jaccard distance matrix (**Figure 3** left) showed that the principal coordinates 1 and 2 explained 14.88 and 11.25% of variances, respectively, and PCoA plots based on the weighted Unifrac distance (**Figure 3** right) showed that the PCo1 and PCo2 explained 84.76 and 6.76% of variances, respectively, suggesting that the weighted Unifrac distance matrix captured more commensal bacteria variances than the Jaccard distance matrix, as illustrated on the two-dimensional PCoA plots. No obvious visual separation was observed between UBD and UBD5 samples. This result was further confirmed by using the PERMANOVA test with *p*-values of 0.767 and 0.077 for the Jaccard distance-based comparison and the weighted Unifrac-based comparison, respectively. The factor of "brand," on the other hand, more significantly impacted the beta diversity of the RL bacterial community, as the PCoA plots based on the Jaccard distance matrix and the weighted Unifrac distance both showed clear visual separation among the three brands. The PERMANOVA test confirmed the significant impact of the factor "brand" on the beta diversity of the RL bacterial community by generating *p*-values of 0.001 for both the Jaccard distance-based comparison and the weighted Unifrac distance-based comparison. When analyzing the impact of harvest season on beta diversity, analyses based on the Jaccard distance was different from the results obtained based on the weighted Unifrac. The PCoA plot presented a clear visual separation between early and late season samples when using the Jaccard distance-based comparison (*p*-value = 0.01 in the PEMANOVA test), while no separation was observed from the PCoA plot based on the weighted Unifrac distance matrix (*p*-value = 0.55 in the PEMANOVA test).

Taxonomic Analysis of Commensal Bacteria in Romaine Lettuce

The taxonomic analysis of commensal bacteria identified 128 genera across all RL samples based on the SILVA database. The top five genera were *Pseudomonas* (with relative abundances ranging from 9.95 to 94.73%), *Weissella* (0–42.42%), *Serratia* (0–52.35%), *Leuconostoc* (0–31.56%), and *Lactococcus* (0–24.30%) (**Figure 4**). When focusing on shelf life, the top five

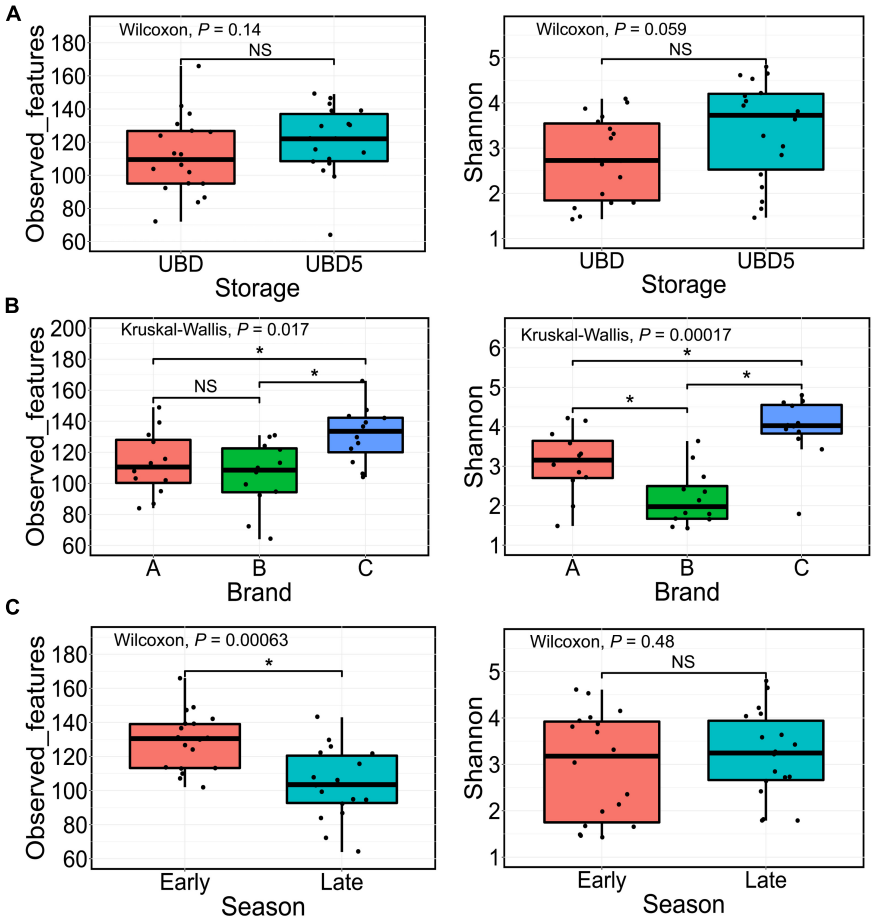


FIGURE 2 | Comparison of the alpha diversity based on the observed features (left) or the Shannon index values (right) of microbial populations present in RL between UBD and UBD5 **(A)**, among three brands **(B)**, and between two seasons **(C)**. The Wilcoxon rank test was used for pairwise comparison while the Kruskal–Wallis test was applied for comparisons among three groups. *Represents significant differences between groups ($p < 0.05$). NS means no significant difference.

genera identified from the UBD samples were *Pseudomonas* (16.47–92.72%), *Serratia* (0–52.35%), *Weissella* (0–42.42%), *Pantoea* (0.17–21.33%), and *Lactococcus* (0–24.30%), while the top five genera identified on UBD5 were *Pseudomonas* (9.95–94.73%), *Pantoea* (0–31.56%), *Leuconostoc* (0.23–28.50%), *Serratia* (0–25.66%), and *Weissella* (0.10–39.62%).

When focusing on different brands, the top five genera identified in brand A samples were *Pseudomonas* (15.32–91.62%), *Leuconostoc* (0.01–31.56%), *Serratia* (0–52.35%), *Pantoea* (0.60–19.48%), and *Lactococcus* (0.2–0.42%); the top five genera identified in brand B were *Pseudomonas* (25.46–94.73%), *Pantoea* (0.17–28.50%), *Janthinobacterium* (0.47–24.01%), *Xanthomonas* (0–30.26%), and *Serratia* (0.05–16.77%); while the top five genera identified in brand C were *Pseudomonas* (9.95–56.41%), *Weissella* (9.21–42.42%), *Serratia* (0.28–46.06%), *Lactococcus* (0.94–24.30%), and *Leuconostoc* (1.14–22.52%). When grouping the samples by harvest seasons, the top five genera identified in the early season were *Pseudomonas* (9.95–91.62%), *Pantoea* (1.64–28.50%), *Serratia* (0–46.06%), *Weissella* (0–42.42%), and *Leuconostoc* (0–31.56%), while the top five genera identified in the late season were *Pseudomonas* (15.33–94.73%), *Weissella* (0–34.69%), *Leuconostoc* (0–26.41%), *Serratia* (0.05–52.35%), and *Lactococcus* (0–24.30%).

Differential Abundance Analysis of Commensal Bacteria in Romaine Lettuce Based on Shelf Life, Brand, and Harvest Season

To identify the critical features (biomarkers) associated with samples when grouped by shelf life, brand, and harvest season, the ANCOM analysis was employed for the differential abundance analysis of commensal bacteria in RL. Results of ANCOM analysis are shown in **Figure 5A**. No bacteria were identified to be significantly different in abundance between UBD and UBD5, while six genera were identified to have significantly differential abundance among three brands; these were *Weissella* (W statistics = 144, clr F statistics = 67.7), *Leuconostoc* (W statistics = 144, clr F statistics = 35.3), *Lactococcus* (W statistics = 144, clr F statistics = 33.9), *Pseudarcobacter* (W

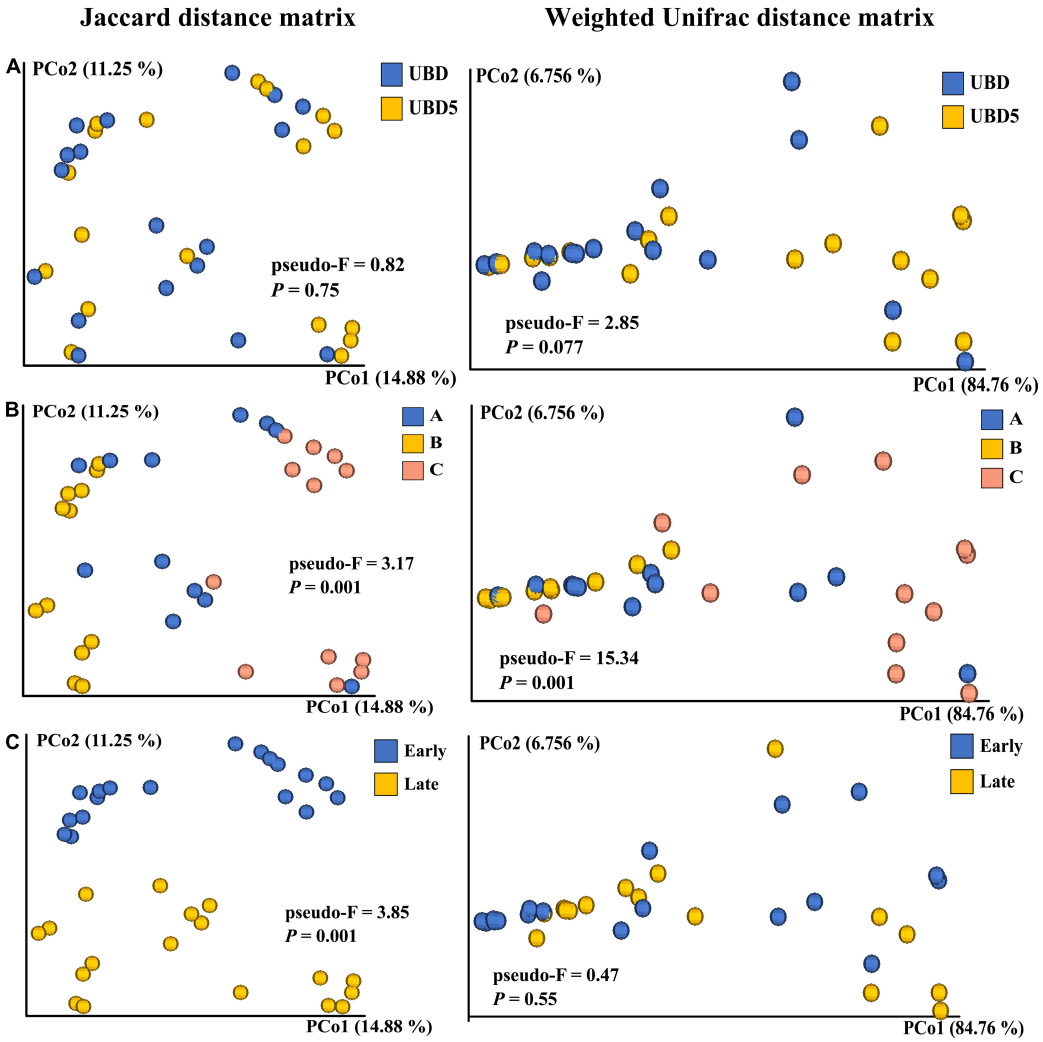


FIGURE 3 | Comparison of the beta diversity based on the Jaccard distance matrix (left) and the weight Unifrac distance matrix (right) of microbial communities present in RL between UBD and UBD5 **(A)**, among three brands **(B)**, or between different seasons **(C)**. The PERMANOVA test was applied to analyze the differences in each group when samples were grouped by shelf life, brand, or seasons. "pseudo-F" values are measures of the effect size, indicating the differences within each group of RL. The *p*-values of less than 0.05 indicate that the factor of shelf life, brand, or season significantly impacts the beta diversity.

statistics = 137, clr F statistics = 19.1), *Yersinia* (W statistics = 134, clr F statistics = 26.3), and *Massilia* (W statistics = 132, clr F statistics = 20.3). *Leuconostoc* (RA 0.011–31.56%) had the highest RA in brand A, while *Massilia* (0–2.12%) had the highest abundance in brand B. *Weissella* (9.21–42.42%), *Lactococcus* (RA 0.94–24.30%), *Pseudarcobacter* (0–4.51%), and *Yersinia* (0–0.50%) had the highest RA in brand C. *Lactococcus*, *Leuconostoc*, *Weissella*, *Pseudarcobacter*, and *Yersinia* had the lowest RA in brand B (**Figure 5B**). When grouping RL samples by harvest seasons, only one unannotated genus (UAG) of Oxalobacteraceae (W statistics = 137, clr F statistics = 3.1) had significantly higher RA in the samples from the late harvest season (0–5.24%) than in the samples from the early harvest season (0–0.083%) (**Figure 5B**). Based on the ANCOM test, no bacteria in the RL of brands A, B, and C individually could be identified as biomarkers for between UBD and UBD5 samples (**Supplementary Figure 4**).

DISCUSSION

This study applied culture-dependent and culture-independent methods to investigate the commensal bacteria present in bagged RL on their UBD and UBD5. The combination of both methods allows us to better characterize the bacterial populations and discover potential changes associated with abundance, diversity, and composition of different bacterial groups. The application of 16S rRNA gene sequencing is advantageous for microbiome profiling and analyzing bacterial community dynamics, as it is culture-independent and relatively unbiased compared to traditional culture methods that rely highly on selective or differentiated media (Grim et al., 2017).

During the quality control step of analyzing the sequence data in the QIIME 2 pipeline, parameters "–p-trunc-len-f," "–p-trunc-len-r," "–p-max-ee-f," and "–p-max-ee-r" were set at

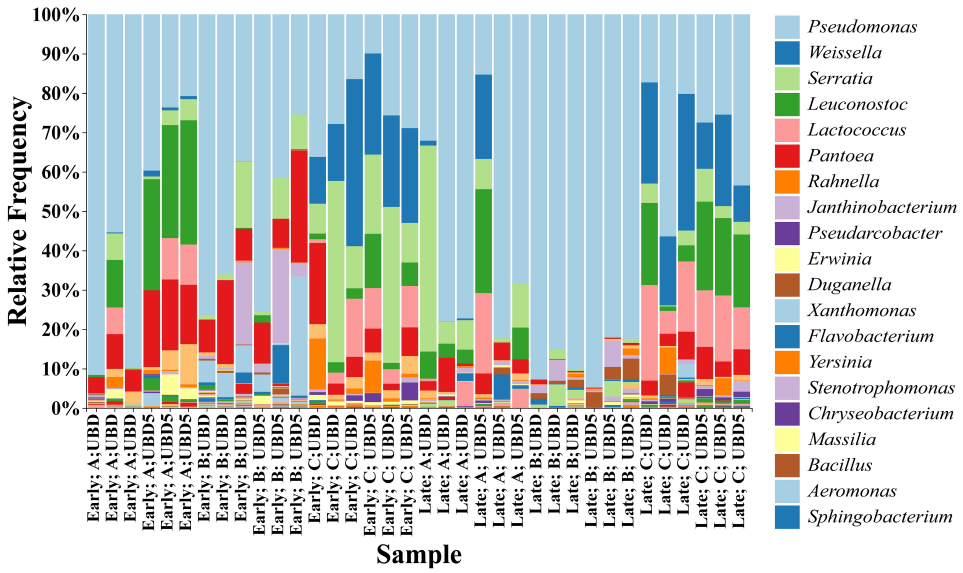


FIGURE 4 | Relative abundance of the top 20 bacteria genera identified from all three brands of RL purchased from two harvest seasons at UBD and UBD5. "UBD" and "UBD5" mean "Use By" dates and 5 days after the UBD. "A," "B," and "C" stand for three brands and "Early" and "Late" represent the early and late harvest season.

260, 185, 2, and 4, respectively, in the q2-dada2 plugin command. Estaki et al. (2020) recommended the setting of the read length at which the median Phred quality score began to drop below 30 or 20 if the entire read quality was too low. With these settings, the forward reads and reverse reads were truncated at 260 and 185 bp as their median Phred quality score of the next base started to drop below 30 as shown in the interactive quality plots (**Supplementary Figure 2**). In addition, the forward reads with more than two erroneous bases and reverse reads with more than four erroneous bases were discarded, as the reverse reads had lower Phred quality scores than the forward reads (**Supplementary Figure 2**). Similar observations have been reported by D'Amore et al. (2016) and Gerasimidis et al. (2016).

As shown in **Figure 1**, culture-dependent results showed no difference in APC or AnPC of RL between UBD and UBD5, except the AnPC of late-season brand B. The population density of culturable commensal bacteria in the RL on the UBD and UBD5 remained at the same level for aerobic bacteria and anaerobic bacteria, indicating that the short 5-day storage after the UBD had no significant impact on the total culturable bacterial abundance. It can be explained by that bacterial populations on UBD and UBD5 had reached stationary levels under this particular storage condition. The APC of bagged RL ranged from 5.71 ± 0.74 to 7.89 ± 0.10 Log CFU/g and the AnPC ranged from 1.75 ± 0.08 to 7.32 ± 0.61 Log CFU/g. RL samples obtained from the late season had higher APC ($p = 3.08 \times 10.5$) and AnPC ($p = 1.06 \times 10.4$) than RL samples obtained from the early season. Culturable bacteria counts obtained from this study were consistent with results from previous studies. Aycicek et al. (2006) reported that the outer leaves of RL and iceberg-lettuce samples harbored 3.3–7.4 Log CFU/g of aerobic bacteria. Korir et al. (2016) reported that the mean abundance of total aerobic bacteria was 7.76 Log CFU/g for lettuce. Similarly, Liao and Wang (2021) reported that the total aerobic bacteria in Spring Mix salad samples were at the level of approximately 6.6 Log CFU/g. The differences observed between different seasons were also reported by Williams et al. (2013) in which the abundance of aerobic bacteria in the phyllosphere was found to be lower on RL planted in the early season (June) (approximately 3.8–5.5 Log CFU/g) than on RL planted in the late season (August and October) (approximately 5.0–6.2 Log CFU/g).

The diversity analysis of the 16S rRNA gene sequencing detected no significant difference of alpha diversity in RL between UBD and UBD5, indicating that neither the richness nor the evenness of the RL bacterial communities had changed. Beta diversity based on the Jaccard distance matrix and the weighted Unifrac distance matrix also showed no difference between UBD and UBD5, as visualized on the PCoA plots. The results were further confirmed by the PERMANOVA test. These results indicate that the additional 5-day storage after the UBD has no impact on the diversity of RL bacterial communities. Similar results have been reported by our previous Spring Mix study in which no significant difference in the beta diversity was observed among Spring Mix samples collected at different storage times during 15 days of cold storage at 4°C (Liao and Wang, 2021).

The top six identified genera, *Pseudomonas*, *Serratia*, *Weissella*, *Pantoea*, *Lactococcus*, and *Leuconostoc*, were the same across RL samples, but in different orders. The exception was that the top six genera in brand B samples included *Janthinobacterium*, *Xanthomonas*, and *Flavobacterium* instead of *Weissella, Lactococcus*, and *Leuconostoc*. Among the top six genera, *Pseudomonas* consistently dominated the RL bacterial communities with RA ranging from 9.95 to 94.73%. The previous study carried out by (Gu et al., 2018) reported that the *Pseudomonas* species consistently dominated microbial communities across all spinach samples during week-long

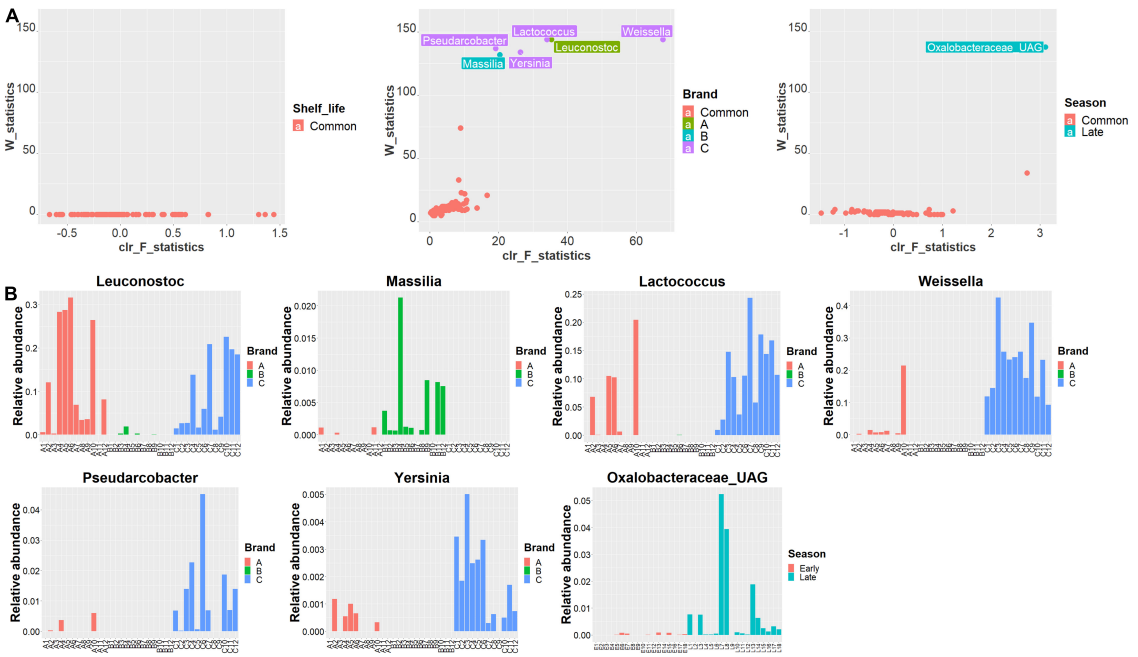


FIGURE 5 | (A) Volcano plots of bacteria, at the genus level, identified to have different abundances when RL are grouped and compared based on their shelf-life (left), brand (middle), and harvest season (right) with the ANCOM test. The abundance differences of bacteria were evaluated by the clr F-statistic and W-statistic. Bacteria with no different abundance when RL samples were grouped and compared based on shelf life, brand, and season factors are marked in red. Bacteria with significantly higher abundances in the brands A, B, and C RL are marked in green, blue, and purple colors, respectively, when comparing the RL of three brands. Bacteria with significantly higher abundances in the late season RL are marked in blue when comparing RL of different seasons. **(B)** Barplots of the relative abundance changes of *Leuconostoc*, *Massilia*, *Weissella*, *Lactococcus*, *Pseudarcobacter*, *Yersinia* identified as the biomarkers for the three brands of RL and Oxalobacteraceae identified as the biomarker for the late season RL.

storage at 4°C (up to 34.20%). Keshri et al. (2019) reported that *Pseudomonas* dominated the bacterial communities of alfalfa sprouts throughout 3 weeks of cold storage at 4°C, with an RA of greater than 60%. The *Pseudomonas* genus contains a group of species associated with fresh produce spoilage, such as *Pseudomonas fluorescents*, *P. marginalis*, *P. viridiflava*, and *P. chloroaphis,* and is able to outcompete other bacteria in food matrixes (Hibbing et al., 2010; Langsrud et al., 2016). Another potential spoilage genus is *Pantoea* (Bae et al., 2014; Keshri et al., 2019). The study from Bae et al. (2014) and Keshri et al. (2019) reported that the *Pantoea* RA was up to 50.40% in mung bean sprouts after 2 weeks of storage at 4°C. A previous study (Leff and Fierer, 2013) also reported that *Pantoea* was highly abundant in pepper (11.5%), spinach (32.5%), and sprouts (57.5%) analyzed right after purchase.

Weissella, *Lactococcus*, and *Leuconostoc* have been reported as psychrotrophic lactic acid bacteria (LAB) potentially linked to the spoilage of fresh produce (Säde et al., 2016; Keshri et al., 2019). Pothakos et al. (2014) reported that *Leuconostoc* spp. was the most dominant population in ready-to-eat salads at the end of their shelf-life (7 days) at 4°C, which caused the early spoilage of vegetables before the end of shelf-life. Keshri et al. (2019) found that the abundance of *Leuconostoc* increased after 2 weeks of cold storage at 4°C in mung bean sprouts, the bacterium is considered to play a role in the spoilage of sprouts. In addition, greater numbers of *Lactococcus* and *Weissella* were reported in Spring Mix salad when being stored at 4°C for 15 days (Liao and Wang, 2021). *Serratia* is a plant-associated genus that is a non-pathogenic symbiont that has also been reported to be present at high levels in organic green leafy lettuce with up to 66% RA (Jackson et al., 2013). *Yersinia* (0–0.50%) and *Bacillus* (0–3.0%), two genera containing human pathogens, e.g., *Yersinia enterocolitica* (Cristiano et al., 2021) and *Bacillus cereus* (Chon and Seo, 2021), were also identified in RL. No significant change in their relative abundance was detected from UBD to UBD5 based on the ANCOM test either.

In this study, ANCOM analysis was applied to analyze and identify biomarkers (bacterial genera) with differential RA in RL samples when grouped by different shelf lives (UBD *vs.* UBD5), brands (brands A, B, and C), and harvest seasons (early *vs.* late). For RL on the UBD and UBD5, the ANCOM test showed no biomarker with significantly differential RA between them, suggesting that the composition of bacteria in the RL had no change 5 days past the UBD. When comparing the microbial populations among three brands, the ANCOM test identified that five genera had significantly different RAs among the three brands. Among them, *Leuconostoc*, *Weissella*, *and Lactococcus* contain species associated with food spoilage, such as *Leuconostoc citreum*, *Leuconostoc mesenteroides*, and *Weissella confuse* (Säde et al., 2016). For the RL from early and late seasons, ANCOM identified only one genus, Oxalobacteraceae_UAG (up to 0.80%), was identified as higher RA in the RL from late

season compared to that from early season. Other than this genus in a low RA, RL from the two seasons had no difference in composition of bacterial communities, which was also reflected in the above Shannon diversity of the two-seasonal RL. Williams et al. (2013) illustrated that the season-of-planting factor more strongly impacted RL bacterial communities after 4–8 weeks of planting in the field. Specifically, *Pantoea*, *Pseudomonas*, *Erwinia*, and members of Enterobacteriaceae were identified as the most abundant genera in late season plants. *Leuconostoc*, *Lactococcus*, *Bacillus*, and *Exiguobacterium* were predominant in early season plants. By comparison, the weaker impact of the seasonal factor on the diversity and composition of RL bacterial communities in our study could be explained by postharvest processing (e.g., washing, packaging, and cold storage), which might alter the RL bacterial communities and mitigate the seasonal impact.

In the United States, except for infant formula, the U.S. Food and Drug Administration (FDA) does not require manufacturers to display specific open date labeling on their food products (Wilson et al., 2017). Although a few poultry, meat, and egg products under the jurisdiction of USDA need data documentation, no strict guidelines for using the food date terminology on food products have been set by the USDA (Leib et al., 2013; Newsome et al., 2014). The current open-date labels used in food markets are not well regulated, as the "UBD," "Best By," and "Sell By" are interchangeably applied by manufacturers (Wilson et al., 2017). This ambiguous food date labeling has fostered confusion about food product safety and quality among consumers, which subsequently causes unnecessary levels of food waste (Wansink and Wright, 2006; Newsome et al., 2014). Evidence showed that consumers waste food products when they are near the open dates perceived for food quality reasons or food safety reasons (Tsiros and Heilman, 2005; Theotokis et al., 2012; Newsome et al., 2014; Wilson et al., 2017).

In January 2017, the Grocery Manufacturers Association (GMA) and Food Marketing Institute (FMI) recommended the use of two introductory phrases for food date labeling, such as "Best If Used By" and "Use By." They recommended that "Best If Used By" is applied to "indicate to the consumer that, after a specified date, the product may not taste or perform as expected but is safe to be used or consumed" and "Use By" is used to "applies to perishable products that should be consumed by the date on the package and discarded after that date" (Grocers Manufacturers Association, 2019). The US FDA strongly supported the industry's voluntary use of the "Best If Use By" introductory phrase for the quality-based food date labels but was not addressing the use of a "Use by" label for safety reasons (Food and Drug Administration, 2019). Other than the United States, most developed countries required food date labeling for most food products. The EU mandates the use of food date labeling for food products. Australia, New Zealand, and the EU clearly regulate the UBD related to food safety and "Best Before" related to food quality (Newsome et al., 2014).

To mitigate food waste, stakeholders need to well clarify and regulate the food date labeling, and customers need more information about fresh produce quality and about safety after the open date. The US Food and Drug Administration (2010) reported that the expected shelf-life for bagged fresh-cut leafy green is approximately 12–16 days after production at 4°C and the sensory quality of fresh leafy green can last at least a week after the open date labels (US Food and Drug Administration, 2010). However, limited study on microbial food quality after "open dates" has been reported. As fresh fruits and vegetables comprise the largest part of food waste in the food system (Augustin et al., 2020), this study presented bacterial RL quality on and after the UBD based on culture-dependent and culture-independent methods. The results showed that RL bacterial communities had no change in abundance, diversity, or composition from the UBD to UBD5. RL on the UBD5 had no increase of microbial quality or safety risk compared to RL on the UBD.

CONCLUSION

In summary, this study determined there was no difference in bacterial abundance, diversity, and composition of bacterial communities in three brands of RL on the UBD5 compared with RL on the UBD, suggesting that the microbial quality of RL remained the same at two storage time points. Factors of "brand" and "harvesting season" played more significant roles in shaping the bacterial communities present in RL samples. The study provides first-hand information about the microbial quality of fresh produce on the UBD and a few days after the UBD.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

CL contributed to experimental design, experiment, manuscript writing, and data analysis. LW contributed to funding acquisition, experimental design, manuscript revising, and supervision. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the Alabama Agricultural Experiment Station, the USDA HATCH Multistate Research Project S1077 "Enhancing Microbial Food Safety by Risk Analysis," and the startup funding from the University of California, Davis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.850720/full#supplementary-material

REFERENCES

Abdi, H., and Williams, L. J. (2010). "Tukey's honestly significant difference (HSD) test," in *Encyclopedia of Research Design,* eds N. Salkind (Thousand Oaks, CA: SAGE publication, Inc.), 1–5. doi: 10.4135/9781412961288

Anderson, M. J. (2014). "Permutational multivariate analysis of variance (PERMANOVA)," in *Wiley Statsref: Statistics Reference Online*, eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, and J. L. Teugels (Chichester, UK: John Wiley & Sons, Ltd), 1–15. doi: 10.1002/9781118445112.stat07841

Arienzo, A., Murgia, L., Fraudentali, I., Gallo, V., Angelini, R., and Antonini, G. (2020). Microbiological quality of ready-to-eat leafy green salads during shelf-life and home-refrigeration. *Foods* 9:421. doi: 10.3390/foods9101421

Augustin, M. A., Sanguansri, L., Fox, E. M., Cobiac, L., and Cole, M. B. (2020). Recovery of wasted fruit and vegetables for improving sustainable diets. *Trends Food Sci. Technol.* 95, 75–85. doi: 10.1016/j.tifs.2019.11.010

Australian Government (2012). *Australia New Zealand Food Standards Code - Standard 1.2.5 - Date Marking of Packaged Food.* Available online at: http://www.comlaw.gov.au/Details/F2012C00762 (accessed February 20, 2022)

Aycicek, H., Oguz, U., and Karci, K. (2006). Determination of total aerobic and indicator bacteria on some raw eaten vegetables from wholesalers in ankara, turkey. *Int. J. Hyg. Environ. Health* 209, 197–201. doi: 10.1016/j.ijheh.2005.07.006

Bae, Y.-M., Zheng, L., Hyun, J.-E., Jung, K.-S., Heu, S., and Lee, S.-Y. (2014). Growth characteristics and biofilm formation of various spoilage bacteria isolated from fresh produce. *J. Food Sci.* 79, M2072–M2080. doi: 10.1111/1750-3841.12644

Berthold-Pluta, A., Garbowska, M., Stefańska, I., and Pluta, A. (2017). Microbiological quality of selected ready-to-eat leaf vegetables, sprouts and non-pasteurized fresh fruit-vegetable juices including the presence of cronobacter spp. *Food Microbiol.* 65, 221–230. doi: 10.1016/j.fm.2017.03.005

Bokulich, N. A., Dillon, M. R., Bolyen, E., Kaehler, B. D., Huttley, G. A., and Caporaso, J. G. (2018a). q2-sample-classifier: machine-learning tools for microbiome classification and regression. *J. Open Res. Softw* 3:934. doi: 10.21105/joss.00934

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018b). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6:90. doi: 10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Buzby, J., Bentley, J., Padera, B., Ammon, C., and Campuzano, J. (2015). Estimated fresh produce shrink and food loss in U.S. supermarkets. *Agriculture* 5, 626–648. doi: 10.3390/agriculture5030626

Buzby, J. C., Farah-Wells, H., and Hyman, J. (2014). *The Estimated Amount, Value, and Calories of Postharvest Food Losses at the Retail and Consumer Levels in the United States*. EIB-121. Washington, DC: U.S. Department of Agriculture, Economic Research Service. doi: 10.2139/ssrn.2501659

Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639–2643. doi: 10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from *Illumina amplicon* data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Chon, J. W., and Seo, K. H. (2021). Evaluation of ceftazidime as an antibiotic supplement in mannitol-yolk-polymyxin b agar used for enumeration of bacillus cereus in ready-to-eat vegetables. *J. Food Prot.* 84, 1698–1703. doi: 10.4315/JFP-20-405

Cristiano, D., Peruzy, M. F., Aponte, M., Mancusi, A., Proroga, Y. T. R., Capuano, F., et al. (2021). Comparison of droplet digital PCR vs real-time PCR for yersinia enterocolitica detection in vegetables. *Int. J. Food Microbiol.* 354, 109321. doi: 10.1016/j.ijfoodmicro.2021.109321

Cruz, M. R. G., Leite, Y. J. B., de, S., Marques, J., de, L., Pavelquesi, S. L. S., et al. (2019). Microbiological quality of minimally processed vegetables commercialized in brasilia, DF, brazil. *FST* 39, 498–503. doi: 10.1590/fst.16018

D'Amore, R., Ijaz, U. Z., Schirmer, M., Kenny, J. G., Gregory, R., Darby, A. C., et al. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genom.* 17:55. doi: 10.1186/s12864-015-2194-9

European Commission (2011). *Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the Provision of Food Information to Consumers.* Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011R1169&from=EN (accessed March 26, 2022).

Edgar, R. C. (2017). Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ* 5:e3889. doi: 10.7717/peerj.3889

European Food Safety Authority Panel on Biological Hazards [BIOHAZ], Koutsoumanis, K., Allende, A., Alvarez-Ordóñez, A., Bolton, D., Bover-Cid, S., et al. (2020). Guidanceon date marking and related food information: part 1 (date marking). *EFSA J.* 18:e06306. doi: 10.2903/j.efsa.2020.6306

Environmental Protection Agency (2015). *United States 2030 Food Loss and Waste Reduction Goal*. Available online at: https://www.epa.gov/sustainable-management-food/united-states-2030-food-loss-and-waste-reduction-goal (accessed February 20, 2022)

Estaki, M., Jiang, L., Bokulich, N. A., McDonald, D., González, A., Kosciolek, T., et al. (2020). QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Curr. Prot. Bioinform.* 70:e100. doi: 10.1002/cpbi.100

Food and Drug Administration (2019). *Confused by Date Labels on Packaged Foods?*. Available online at: https://www.fda.gov/consumers/consumer-updates/confused-date-labels-packaged-foods (accessed February 20, 2022)

US Food and Drug Administration (2010). *Program Information Manual Retail Food Protection: Recommendations for the Temperature Control of Cut Leafy Greens During Storage and Display in Retail Food Establishments*. Available online at: https://www.fda.gov/media/78982/download (accessed February 20, 2022)

Food Standards Australia New Zealand (2013). *Date Marking User Guide to Standard 1.2.5: Date Marking of Packaged Food*. Available online at: http://www.foodstandards.gov.au/code/userguide/Documents/Guide%20to%20Standard%201.2.5%20-%20Date%20Marking%20of%20Food.pdf (accessed February 20, 2022)

Gerasimidis, K., Bertz, M., Quince, C., Brunner, K., Bruce, A., Combet, E., et al. (2016). The effect of DNA extraction methodology on gut microbiota research applications. *BMC Res. Notes* 9:365. doi: 10.1186/s13104-016-2171-7

Grim, C. J., Daquigan, N., Lusk Pfefer, T. S., Ottesen, A. R., White, J. R., and Jarvis, K. G. (2017). High-resolution microbiome profiling for detection and tracking of *Salmonella enterica*. *Front. Microbiol.* 8:1587. doi: 10.3389/fmicb.2017.01587

Grocers Manufacturers Association (2019). *Best if Clearly Labeled: How the Consumer Packaged Goods Industry is Reducing Confusion and FoodWaste*. Available online at: http://healthdocbox.com/Nutrition/112254594-Best-if-clearly-labeled-how-the-consumer-packaged-goods-industry-is-reducing-confusion-and-food-waste.html (accessed February 20, 2022).

Gu, G., Ottesen, A., Bolten, S., Ramachandran, P., Reed, E., Rideout, S., et al. (2018). Shifts in spinach microbial communities after chlorine washing and storage at compliant and abusive temperatures. *Food Microbiol.* 73, 73–84. doi: 10.1016/j.fm.2018.01.002

Hibbing, M. E., Fuqua, C., Parsek, M. R., and Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* 8, 15–25. doi: 10.1038/nrmicro2259

Jackson, C. R., Randolph, K. C., Osborn, S. L., and Tyler, H. L. (2013). Culture dependent and independent analysis of bacterial communities associated with commercial salad leaf vegetables. *BMC Microbiol.* 13:274. doi: 10.1186/1471-2180-13-274

Jeddi, M. Z., Yunesian, M., Gorji, M. E., Noori, N., Pourmand, M. R., and Khaniki, G. R. J. (2014). Microbial evaluation of fresh, minimally-processed vegetables and bagged sprouts from chain supermarkets. *J. Health Popul. Nutr.* 32, 391–399.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Keshri, J., Krouptiski, Y., Abu-Fani, L., Achmon, Y., Bauer, T. S., Zarka, O., et al. (2019). Dynamics of bacterial communities in alfalfa and mung bean sprouts during refrigerated conditions. *Food Microbiol.* 84:103261. doi: 10.1016/j.fm.2019.103261

Kõiv, V., Arbo, K., Maiväli, Ü, Kisand, V., Roosaare, M., Remm, M., et al. (2019). Endophytic bacterial communities in peels and pulp of five root vegetables. *PLoS One* 14:e0210542. doi: 10.1371/journal.pone.0210542

Korir, R. C., Parveen, S., Hashem, F., and Bowers, J. (2016). Microbiological quality of fresh produce obtained from retail stores on the eastern shore of maryland, united states of america. *Food Microbiol.* 56, 29–34. doi: 10.1016/j.fm.2015.12.003

Langsrud, S., Moen, B., Møretrø, T., Løype, M., and Heir, E. (2016). Microbial dynamics in mixed culture biofilms of bacteria surviving sanitation of conveyor belts in salmon-processing plants. *J. Appl. Microbiol.* 120, 366–378. doi: 10.1111/jam.13013

Leff, J. W., and Fierer, N. (2013). Bacterial communities associated with the surfaces of fresh fruits and vegetables. *PLoS One* 8:e59310. doi: 10.1371/journal.pone.0059310

Leib, E. B., Gunders, D., Ferro, J., and Nielsen, A., Nosek, G., and Qu, J. (2013). *The Dating Game: How Confusing Food Date Labels Lead to Food Waste in America.* NRDC Rep. R:13-09-A. Available online at: http://www.nrdc.org/food/files/dating-game-report.pdf (accessed March 26, 2022).

Leib, E. B., Rice, C., Neff, R., Spiker, M., and Schklair, A. (2016). *Consumer Perceptions of Date Labels: National Survey*. Harvard Food Law and Policy Clinic, National Consumers League, and Johns Hopkins University Center for a Livable Future.

Liao, C., and Wang, L. (2021). Evaluation of the bacterial populations present in spring mix salad and their impact on the behavior of *Escherichia coli* O157:H7. *Food Control.* 2021:107865. doi: 10.1016/j.foodcont.2021.107865

Lorenzo, J. M., Munekata, P. E., Dominguez, R., Pateiro, M., Saraiva, J. A., and Franco, D. (2018). "Main groups of microorganisms of relevance for food safety and stability," in *Innovative Technologies for Food Preservation,* eds F. J. Barba, A. S. Sant'Ana, V. Orlie, and M. Koubaa (Amsterdam, NL: Elsevier, Inc.), 53–107. doi: 10.1016/B978-0-12-811031-7.00003-0

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169–172. doi: 10.1038/ismej.2010.133

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., and Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* 26:27663. doi: 10.3402/mehd.v26.27663

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10. doi: 10.14806/ej.17.1.200

Mijares, V., Alcivar, J., and Palacios, C. (2021). Food waste and its association with diet quality of foods purchased in south florida. *Nutrients* 13:535. doi: 10.3390/nu13082535

Newsome, R., Balestrini, C. G., Baum, M. D., Corby, J., Fisher, W., Goodburn, K., et al. (2014). Applications and perceptions of date labeling of food. *Comput. Rev. Food Sci. Food Safety* 13, 745–769. doi: 10.1111/1541-4337.12086

Pothakos, V., Snauwaert, C., De Vos, P., Huys, G., and Devlieghere, F. (2014). Monitoring psychrotrophic lactic acid bacteria contamination in a ready-to-eat vegetable salad production environment. *Int. J. Food Microbiol.* 185, 7–16. doi: 10.1016/j.ijfoodmicro.2014.05.009

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Saary, P., Forslund, K., Bork, P., and Hildebrand, F. (2017). RTK: efficient rarefaction analysis of large datasets. *Bioinformatics* 33, 2594–2595. doi: 10.1093/bioinformatics/btx206

Säde, E., Lassila, E., and Björkroth, J. (2016). Lactic acid bacteria in dried vegetables and spices. *Food Microbiol.* 53, 110–114. doi: 10.1016/j.fm.2015.09.005

Sinclair, L., Osman, O. A., Bertilsson, S., and Eiler, A. (2015). Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the illumina platform. *PLoS One* 10:e0116955. doi: 10.1371/journal.pone.0116955

Stevens, A. J., Purcell, R. V., Darling, K. A., Eggleston, M. J. F., Kennedy, M. A., and Rucklidge, J. J. (2019). Human gut microbiome changes during a 10 week randomised control Trial for micronutrient supplementation in children with attention deficit hyperactivity disorder. *Sci. Rep.* 9:10128. doi: 10.1038/s41598-019-46146-3

Theotokis, A., Pramatari, K., and Tsiros, M. (2012). Effects of expiration date-based pricing on brand image perceptions. *J. Retail.* 88, 72–87. doi: 10.1016/j.jretai.2011.06.003

Tsiros, M., and Heilman, C. M. (2005). The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. *J. Mark.* 69, 114–129. doi: 10.1509/jmkg.69.2.114.60762

United States Department of Agriculture (2015). *Food Loss and Waste 2030 Champions*. Available online at: https://www.usda.gov/foodlossandwaste/champions (accessed February 20, 2022)

Wansink, B., and Wright, A. O. (2006). best if used by …?" how freshness dating influences food acceptance. *J. Food Sci.* 71, S354–S357. doi: 10.1111/j.1750-3841.2006.00011.x

Williams, T. R., Moyne, A.-L., Harris, L. J., and Marco, M. L. (2013). Season, irrigation, leaf age, and *Escherichia coli* inoculation influence the bacterial diversity in the lettuce phyllosphere. *PLoS One* 8:e68642. doi: 10.1371/journal.pone.0068642

Wilson, N. L. W., Rickard, B. J., Saputo, R., and Ho, S.-T. (2017). Food waste: The role of date labels, package size, and product category. *Food Qual Prefer* 55, 35–44. doi: 10.1016/j.foodqual.2016.08.004

Xia, Y., and Sun, J. (2017). Hypothesis testing and statistical analysis of microbiome. *Genes Dis.* 4, 138–148. doi: 10.1016/j.gendis.2017.06.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fmicb.2022.849883

Similar Carcass Surface Microbiota Observed Following Primary Processing of Different Pig Batches

Charlotte Braley[1,2], Philippe Fravalo[3,4], Marie-Lou Gaucher[1,2,3,5], Guillaume Larivière-Gauthier[4], Fanie Shedleur-Bourguignon[1,2], Jessie Longpré[5,6] and Alexandre Thibodeau[1,2,3,5]*

[1] Chaire de Recherche en Salubrité des Viandes (CRSV), Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, QC, Canada, [2] Département de Pathologie et Microbiologie, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, QC, Canada, [3] Groupe de Recherche et d'Enseignement en Salubrité Alimentaire (GRESA), Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, QC, Canada, [4] Le Conservatoire National des Arts et Métiers (CNAM), Paris, France, [5] Center de Recherche en Infectiologie Porcine et Avicole (CRIPA), Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, QC, Canada, [6] F. Ménard, Division d'Olymel s.e.c., Ange-Gardien, QC, Canada

OPEN ACCESS

Edited by:
Jerome Combrisson, Mars, United States

Reviewed by:
Beatrix Stessl, University of Veterinary Medicine Vienna, Austria
Héctor Argüello, University of Córdoba, Spain

***Correspondence:**
Charlotte Braley
braley.braley@umontreal.ca

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *06 January 2022*
Accepted: *09 May 2022*
Published: *27 May 2022*

Citation:
Braley C, Fravalo P, Gaucher M-L, Larivière-Gauthier G, Shedleur-Bourguignon F, Longpré J and Thibodeau A (2022) Similar Carcass Surface Microbiota Observed Following Primary Processing of Different Pig Batches. Front. Microbiol. 13:849883. doi: 10.3389/fmicb.2022.849883

Bacterial contamination during meat processing is a concern for both food safety and for the shelf life of pork meat products. The gut microbiota of meat-producing animals is one of the most important sources of surface contamination of processed carcasses. This microbiota is recognized to vary between pigs from different farms and could thus be reflected on the bacterial contamination of carcasses at time of processing. In this study, the microbiota of 26 carcasses of pigs originating from different farms (i.e., batches) were compared to determine if an association could be observed between carcass surface microbiota (top and bottom) and the origin of slaughtered animals. The microbiota of the top and bottom carcass surface areas was analyzed by culturing classical indicator microorganisms (mesophilic aerobic bacteria, Enterobacteria, *Escherichia coli*, *Pseudomonas*, and lactic bacteria), by the detection of *Salmonella*, and by 16S rRNA gene sequencing. Culture results showed higher Enterobacteria, *E. coli,* and lactic bacteria counts for the bottom areas of the carcasses (neck/chest/shoulder) when compared to the top areas. *Salmonella* was not detected in any samples. Globally, 16S rRNA gene sequencing showed a similar composition and diversity between the top and bottom carcass areas. Despite the presence of some genera associated with fecal contamination such as *Terrisporobacter*, *Escherichia-Shigella*, *Turicibacter*, *Clostridium sensustricto1*, and *Streptococcus* on the carcass surface, sequencing analysis suggested that there was no difference between the different batches of samples from the top and bottom areas of the carcasses. The primary processing therefore appears to cause a uniformization of the carcass global surface microbiota, with some specific bacteria being different depending on the carcass area sampled.

Keywords: pig slaughterhouse, pork carcass, farm influence, microbiota, primary processing

INTRODUCTION

The production of proteins derived from animal sources is an important part of the food chain for humans. Pork products are a growing part of the world economy. Since 1961, the world pork production has increased by four to five times, reaching 112 million tons in 2014 (Ritchie and Roser, 2017). Moreover, consumers are demanding premium food commodities, forcing producers to further reduce bacterial contamination on pig carcasses to improve the quality of the final pork meat products. Bacteria from different sources such as the skin and the digestive tract of the animals, the slaughterhouse environment and equipment contribute to the constitution of the carcass microbiota (Biasino et al., 2018; Zwirzitz et al., 2020) during processing. In pig meat processing, evisceration is universally recognized as a critical step that often results in carcass contamination by bacterial populations, particularly when the application of good practices is either lacking or sub-optimal (Rivas et al., 2000). Therefore, the gut microbial community is thought to be the most important source of carcass contamination by both non-pathogenic and pathogenic bacteria during the pigs meat processing in commercial conditions (Wheatley et al., 2014). The presence of these microorganisms can speed-up meat spoilage, a process that ultimately leads to products that are not suitable for human consumption, in turn leading to economic losses and food waste (Gram et al., 2002). This also impacts the microbial safety of pork meat products, raising concerns over associated foodborne diseases (Borch et al., 1996).

Salmonella is one of the most common foodborne pathogens, and contaminated pork meat products are a major source of human foodborne infections, especially in Europe and in the United States (US) (Martínez-Avilés et al., 2019). For example, in the United States, the Centers for Disease Control and Prevention (CDC) estimated that *Salmonella* caused 35% of the illnesses linked to pork meat (Self et al., 2017). In European countries, the prevalence of *Salmonella* on pig carcass surfaces ranged from 0.35% to 17.41% (Pala et al., 2019). The most effective interventions that are recognized to contribute to lowering the presence of *Salmonella* on pig carcasses at the slaughterhouse plant level are the rigorous application of good hygiene practices as well as the slaughter of *Salmonella*-free animals [Bonardi, 2017; Regulation (EC) No 2073/2005, 2005]. Indeed, *Salmonella* is often found within the pig intestinal microbiota without causing any clinical signs in the colonized animals.

The presence of foodborne pathogens on carcasses is routinely accompanied by spoilage microorganisms. Spoilage is defined as unfavorable changes of the organoleptic properties of the meat—such as off-flavors, texture, and poor taste (Casaburi et al., 2015)—rendering the meat unsuitable for human consumption. The initial number and type of microorganisms directly influence the time needed to reach a sufficient level to cause these changes (Gram et al., 2002; Wheatley et al., 2014). Psychrotrophic aerobes and facultative anaerobes such as members of the Enterobacteria, *Pseudomonas spp.*, and lactic acid bacteria (Pothakos et al., 2015) are examples of spoilage microorganisms that originate from the food processing environment and the intestinal microbiota of the animals (Wheatley et al., 2014).

The swine gut microbiota is dominated by Firmicutes, Bacteroidetes, and Proteobacteria at the phylum level (Wang et al., 2019). This gut microbiota plays important roles in the animal host metabolism, immune system development, and resistance to pathogens (Fouhse et al., 2016). Several factors can influence the composition of the gut microbiota: genetics, diet, age of the host, antibiotic treatments, and the presence of foodborne pathogens such as *Salmonella* (Hasan and Yang, 2019). The gut microbiota has also been shown to vary between pigs originating from different farms (Yang et al., 2018).

Until now, most studies conducted on carcass microbiological quality at slaughterhouses focused on the transmission of gut bacteria to the pig carcass surface using culture-dependent methods such as the counting of fecal indicator bacteria that indicate the microbial quality of the final meat product. Studies showed that these fecal indicators were not uniformly distributed on the pig carcass surface, that the level of contamination was higher on the bottom (neck) area, and that this contamination varied between carcasses (Wheatley et al., 2014; Biasino et al., 2018). However, results generated using these culture-based approaches allow for the study of a small fraction of the bacterial communities present. The introduction of high-throughput sequencing technologies has profoundly contributed to the advancement of knowledge in this field, permitting the simultaneous detection of hundreds of bacterial genera for which culturing is not always suitable. A limited number of studies describing the microbiota of pig carcasses using high-throughput sequencing technologies have already revealed the usefulness of this approach and have shed new light on the diversity of the bacterial populations present on pig carcasses. For example, a study by Jakobsen et al. (2019) described the bacterial community transfer from pig tonsils to carcass surface and identified specific bacterial groups that were indicative of a bacterial transfer between these two anatomical regions. Another study by Peruzy et al. (2021) showed that various areas (ham, belly, back and jowl) of the surface of different pig carcasses were dominated by the same bacterial communities, though this microbiota varied between slaughterhouses.

To the best of our knowledge, no study has looked at the influence of the animal origin on the carcass surface microbiota following primary processing of pigs slaughtered in commercial conditions on the same day. Indeed, intestinal microbiota greatly varies between animals originating from different farms (Yang et al., 2018), and on-farm intestinal microbiota manipulation is being applied to increase animal health and lower the incidence of foodborne pathogens. It is therefore important to assess if the farm origin impacts carcass microbiota to gain insight into whether different on-farm interventions might also affect the microbiota of pork products. Therefore, the main objective of this study was to observe if pig carcass surface microbiota could be associated with the origin (batch) of animals. A comparison of carcass surface microbiota between samples of top and bottom areas from different batches was therefore carried out using both culture-dependent and high-throughput sequencing.

MATERIALS AND METHODS

Slaughterhouse

Carcass surface samples were collected at a pig slaughterhouse in Québec, Canada. Before slaughtering, pigs were kept in separate lairage pens for 3 h. The animals were rendered unconscious by carbon dioxide stunning. After bleeding, the carcasses were scalded for 7 min in hot water (temperature varied between 59.5°C and 64°C). Carcasses were then scraped, dehaired, and buckled before being pre-washed with clean water and inspected by a veterinarian. Carcasses were then eviscerated and washed again before cooling (24 h). The processing time between stunning and the start of cooling was 32 min. The number of pigs slaughtered per hour was 649.

Sampling and Sample Preparation

A total of 26 pig carcasses from 6 different batches were sampled at the end of carcass dressing, just before the final wash that precedes cooling. All steps until this final wash are considered in this study as primary processing. Throughout the manuscript, the term batch refers to animals raised in the same infrastructure (farm) at the same geographical address, of the same age and slaughtered at the same time. Being raised under a same integrated company, pigs from each farm were similar in terms of genetics and productivity and had been fed a similar diet. Batches were numbered 1–6, though pigs sampled in this study were not slaughtered in this order. For each batch, 4 carcasses were sampled, except for the first 2 batches where 5 carcasses were sampled. All samples were collected on the same day over 5 consecutive hours of production. During primary processing, pigs were hung by the back feet. For each carcass, one sample was collected from the top (thighs/buttock) and one sample from the bottom (neck/chest/shoulder). A total of 26 top carcass surface areas and 26 bottom carcass surface areas were sampled. Sterile wipes were first humidified with 10 mL of sterile PBS (ThermoFisher scientific, Ottawa, Canada) prior to sampling and put in a sterile bag. For each sample, an approximate 600 cm^2 area was firmly swabbed 10 times horizontally and 10 times vertically. Between each area samples, handler's glove were changed. Three negative controls (wipes exposed to the atmosphere of the slaughterhouse without touching any surfaces) and three external slaughterhouse controls (wipes humidified with 10 ml of sterile PBS and used to swab the floor of the slaughterhouse) were also included during sampling. All samples were kept on ice until processing at the laboratory. Wipes were homogenized in 50 mL of cold buffer (1 mM EDTA, 10 mM Tris–Hcl, 8.5 g NaCl pH 8), homogenized using a stomacher Smasher^{TM} AESAP1064 (Bioméreux, United States) for 1 min and kept on ice. From this volume, 5 mL was immediately used for bacterial enumeration while 30 mL was centrifuged for 20 min at 2,800 *g* (Sorvall legend X1R centrifuge, ThermoFisher scientific, United States). The supernatant was discarded and the dry pellet stored at −80°C until DNA extraction.

Bacterial Enumeration and Confirmation

For each microbial indicator, 100 μl from each sample were directly plated using a spiral seeder (Spiral Interscience, ThermoFisher scientific, United States) on the appropriate culture medium (**Table 1**).

Presumptive *Pseudomonas* colonies were confirmed by a positive oxidase test (Gordon-McLeod Reagent, Sigma-Aldrich). In parallel, control strains of each microbial indicator such as *Escherichia coli ATCC 25922*, *Pseudomonas aeruginosa ATCC 27853*, and *Lactobacillus salivarius ATCC 1174* were used to validate the method and culture conditions used.

Detection of *Salmonella*

The detection of *Salmonella* was based on a previous study conducted by our group (Larivière-Gauthier et al., 2019, ISO 6579-1:2017). Briefly, 1 mL of samples were pre-enriched in buffered peptone water (1:10 w:v, 24 h, 37°C). Three drops of 100 μL were selectively enriched on Modified Semi-Solid Rappaport-Vassiliadis Agar (MSRV) (Biokar diagnostic, Beauvais, France) (48 h, 42°C). Two selective media—Brilliant Green Sulfa (BGS) (BD Difco, Franklin Lakes, NJ, United States) and Xylose-Lysine-Desoxycholate (XLD) (Biokar diagnostic)—were inoculated from the migration front of each positive MSRV and incubated for 24 h at 37°C. When possible, two typical colonies from each culture medium were confirmed using triple sugar iron agar slants (BD Difco, Franklin Lakes, NJ, United States), urea agar slants, and seroagglutination with *Salmonella* O antiserum Poly A-I C Vi performed on

TABLE 1 | Culture conditions for the enumeration of mesophilic aerobic bacteria, Enterobacteria, *Escherichia coli*, lactic bacteria, and *Pseudomonas*.

Microorganisms	Culture media	Culture conditions	According to the procedure derived from
Mesophilic aerobic bacteria	Trypticase Soy Agar (BD Difco, Franklin Lakes, NJ, United States)	30°C, 48 h	ISO 4833-2:2013
Enterobacteria	Violet Red Bile Glucose Agar (BD Difco)	37°C, 48 h	ISO 21528-2:2017
Escherichia coli	MacConkey Agar (BD Difco)	37°C, 48 h	(MAPAQ, 2019)
Lactic bacteria	Man, Rogosa, Sharpe medium (BD Difco)	37°C, 48 h (GazpakAnaeroGen Thermo Scientific^{TM} Oxoid R)	(, 2019)
Pseudomonas	Cephalosporin-Fucidine-Cetrimide (Biokar diagnostic)	25°C, 48 h	ISO 13720:2010

colonies previously purified on blood agar (Statens Serum Institute, Denmark).

DNA Extraction

Total DNA was extracted from the pellets and kept at −80°C using a mechanical and chemical lysis followed by phenol/chloroform purification. For each sample, 500 mL of lysis buffer [500 mM Tris–HCl pH 8, 2 mM EDTA pH 8, 100 mM NaCl, and 1% SDS (w/v)] containing 1 g of 0.1 mm glass beads was added to each sample. Cells were mechanically lysed using FastPrep-24^{TM} (MP Biomedicals, Santa Ana, CA, United States) for 40 s at 6 m/s and kept on ice. Lysates were centrifuged for 15 min at 18,000 × *g* (Compact Micro Centrifuges, VWR International, United States) to remove beads and cell debris. For each sample, 300 μL of supernatant was mixed by inversion for 5 min with 300 μL of phenol/chloroform/isoamyl alcohol (25:24:1). After centrifugation (18,000 × *g*, 5 min) (VWR International), the aqueous phase was kept and added to 500 μL of chloroform/isoamyl alcohol (24:1). After centrifugation (18,000 × *g*, 10 min), 350 μL of supernatant was added to 117 μL ammonium acetate (0.1 g/mL) and 934 μL of 90% ethanol. DNA was precipitated overnight at −20°C. After centrifugation (18,000 × *g*, for 15 min), 250 μL of 70% ethanol was added to the pellet, samples were centrifuged, and the supernatant was removed. The DNA pellet was air dried for 30 min and 40 μL of dissolution solution (1 mM Tris–HCl pH 8, 0.1 mM EDTA pH 8) was added. Purified DNA samples were stored at −80°C. After extraction, the final DNA concentration was measured using the Qubit 3.0 broad range assay (Fisher Scientific, Ottawa, ON, Canada) on a Denovix (Wilmington, DE, United States) fluorometer. The purity of DNA was assessed using a Nanodrop (ThermoFisher Scientific, United States). In addition, a negative control was included during DNA extraction (water instead of a carcass swab sample) to assess potential cross-contamination during the extraction step.

16S rRNA Sequencing

A 291 pb fragment of the V4 region of the 16S rRNA gene was amplified by PCR using the V4-reverse 5′ACACTGACGAACTGGTTCTACAAGTGCAGCMGCCGCG GTAA 3′ and V4-forward 5′TACGGTAGCAGAGACTTGG TCTGGACTACHUGGGTWTCTAAT 3′ (Caporaso et al., 2012) primers. The PCR reaction mix (30 μl) contained 5X SuperfiBuffer, 5X SuperfiGCenhancer, 2 U/μL Platinum Superfi DNA Polymerase (Invitrogen, Burlington, ON, Canada), 10 mM dNTPmix (Bio Basic Inc., Markham, ON, Canada), 20 μM of primer (Invitrogen), 20 mg/mL BSA (ThermoFisher scientific, Ottawa, Canada), 12.5 ng of DNA, and sterile water to reach final volume. The amplification was carried out for 25 cycles and included a denaturation step at 95°C for 30 s, an annealing step at 55°C for 30 s, and an elongation step at 72°C for 60 s in a Mastercycler® Nexus (Eppendorf AG, Hamburg, Germany). These cycles were preceded by an initial denaturation of 5 min at 95°C and followed by a final elongation of 10 min at 72°C. A positive control containing DNA from eight known bacterial DNA with different 16S rRNA gene abundance (theorical composition based on 16S sequencing: 18.4% *Lactobacillus*, 17.4% *Bacillus*, 15.5% *Staphylococcus*, 14.1% *Listeria*, 10.4% *Salmonella*, 10.1% *Escherichia*, 9.9% *Enterococcus*, and 4.2% *Pseudomonas*) was included (ZymoBIOMICS Microbial Community DNA Standard, Zymo Research, Irvine, CA, United States). A PCR negative control (water instead of DNA) was included to assess potential cross-contamination during the PCR step. PCR amplification was confirmed by migration on a 1.5% agarose gel. Amplicons were sent for sequencing (Illumina MiSeq, PE 250) to McGill University and Genome Québec Innovation Center, Montréal, Canada.

Sequencing Data Processing and Analysis

Raw sequence reads were cleaned and analyzed using Mothur (Schloss et al., 2009) version 1.43 following the MiSeq standard operational procedure[1] with some modifications, as described in Kozich et al. (2013). The Deblur algorithm was used to complete preclustering. Sequences were aligned against SILVA 132 reference database[2]. Chimeras were removed using VSEARCH (Rognes et al., 2016). The resulting sequences were classified against the SILVA 132 reference database (formatted for Mothur). Sequences were clustered into Operational Taxonomic Units with a unique method, therefore Amplicon Sequence Variants (ASV) were used for analysis. Taxonomic assignation was performed with the Ribosomal Database Project (RDP) trainset 16 database[3]. Alpha diversity indexes (Observed ASV, Shannon and Inverse Simpson) were calculated using the R package "phyloseq" (McMurdie and Holmes, 2013). Beta diversity was analyzed using Jaccard and Bray-Curtis dissimilarity indexes, and the microbiota structure was visualized using non-metric multidimensional scaling (NMDS) graphs in RStudio (version 1.4.1103).

Statistical Analysis

Total bacterial counts were converted into $\log_{10}$ CFU/600 cm^2 values. Statistical analysis was performed using Rstudio (version 1.4.1103). A *t*-test was used for the comparison of the mesophilic aerobic bacteria mean count between the top and bottom areas of the sampled carcass surfaces. An ANOVA test was conducted for the comparison of mesophilic aerobic counts between the six different batches for samples recovered from both the top and bottom of the carcasses sampled. Graphs were generated with GraphPad Prism 8.0.2. Since counts for some indicator microorganisms were below the detection limit for most samples, the number of positive carcasses for Enterobacteria, *Escherichia coli*, and lactic bacteria between the different batches was analyzed by Fisher exact test. *Salmonella* prevalence was also analyzed by Fisher exact test. A $p < 0.05$ value was considered statistically significant.

Statistical tests related to sequencing data were performed in Rstudio according to our in-house Standard Operating Procedure (SOP) (Larivière-Gauthier et al., 2017). Data were first normalized according to the number of total counts in each sample using the lowest number of sequences found in a carcass

[1]https://mothur.org/wiki/miseq_sop/

[2]https://www.mothur.org/wiki/Silva_reference_files

[3]https://www.mothur.org/wiki/RDP_reference_files

sample. Alpha and beta diversity analyses were performed. A *t*-test was used to compare the alpha diversity measures identified from the top and bottom of the sampled carcasses. An Kruskal–Wallis test was performed to compare the alpha diversity measures from all top and bottom surfaces between the six batches sampled.

PERMANOVA tests were conducted using the ADONIS function in the vegan package (Oksanen et al., 2018) for the analysis of the microbiota structure to compare top and bottom carcass surfaces and to assess differences for each type of samples according to batch. Species abundance was compared between groups using Multivariate Association with Linear Models 2 (MaAsLin2) (Mallick et al., 2021) to identify biomarkers associated with sample type (top or bottom) or batches at the phylum, family, and genus levels [using the tax_glom function in the phyloseq package (McMurdie and Holmes, 2013)] after grouping of the 6 batches without rarefaction of the sequences. All analysis in MaAslin2 were performed using default options and an association was considered significant at a p-value < 0.05 and q-value < 0.25.

RESULTS

Enumeration of Bacterial Populations

The bacterial communities present on pig carcass surfaces were investigated for 26 carcasses from 6 different batches. Both the top (thighs/buttock) and bottom (neck/chest/shoulder) areas of the pig carcass were sampled. The mean mesophilic aerobic bacteria concentration for the top and bottom areas of the carcass, regardless of the batch was 4.6 $\log_{10}$ CFU/600 cm^2 and 5.3 $\log_{10}$ CFU/600 cm^2, respectively. When comparing these same mesophilic aerobic bacteria counts obtained from the sampling of the bottom area to those of the top part of the carcass for each batch separately, a significant difference between top and bottom areas was found for 5 out of the 6 batches sampled ($p < 0.05$) (**Figure 1**). When conducting this comparison at the batch level, no statistically significant difference could be observed for the mesophilic aerobic bacteria counts for both the top and bottom carcass surface areas among the six batches sampled ($p > 0.05$).

Bacterial counts for Enterobacteria, *Escherichia coli*, lactic acid bacteria, and *Pseudomonas* are presented in **Table 2**. For all of these bacterial populations, the percentage of positive carcasses sampled recovered from the bottom area was significantly higher than the proportion of positive samples identified for the top area of the carcass ($p < 0.05$) (**Table 2**).

Sequence Quality

The V4 region of the 16S rRNA gene was sequenced from sixty-one samples (26 bottom carcasses, 26 top carcasses, 2 negatives controls (extraction and PCR), 3 negative wipe controls, 3 external slaughterhouse controls, and 1 DNA community). After reads processing, a total of 1,449,466 sequences were retained and assigned to 27 phyla, 58 classes, 106 orders, 225 families, and 663 genera. The average number of sequences per sample was 79,421. The total number of ASV was 22,920. The mock community corresponding to the positive control was composed of 17.5% *Salmonella*, 15.9% *Escherichia/Shigella*, 14.6% *Bacillus*, 12.6% *Staphylococcus*, 12.3% *Lactobacillus*, 11.4% *Listeria*, 8.2% *Pseudomonas*, and 6.4% *Enterococcus*. The negative controls used for the DNA extraction and PCR steps contained 798 and 29,980 sequences respectively, and no band was visible on gel electrophoresis.

Carcass Surface Microbiota Description of the Top and Bottom Areas of Pig Carcasses

First, we compared the microbiota structure between the carcass samples (top or bottom areas), the mock community, the negative controls, and the external slaughterhouse controls to investigate dissimilarities between each sample type (**Figure 2**). Two samples (1 top sample and 1 bottom sample from different batches) were significantly different from the overall carcass samples collected and relatively close to the negative control samples

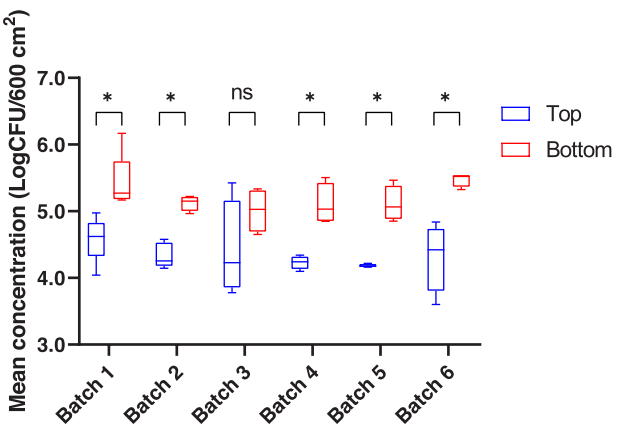


FIGURE 1 | Mean concentration (log CFU/600 cm^2) for mesophilic aerobic bacteria counts for the top and the bottom areas of the pig carcass surface. *$p < 0.05$. ns, non-significant.

TABLE 2 | Percentage of positive carcasses and mean bacterial concentrations for Enterobacteria *Escherichia coli*, lactic acid bacteria, *Pseudomonas*, and *Salmonella* for the top and bottom areas of the carcass surface samples analyzed.

	Positive samples (%)		Mean bacterial concentration ($\log_{10}$ CFU/600 cm^2 ± σ)	
	Top	Bottom	Top	Bottom
Enterobacteria	3.8^a	38.5^b	1.6 ± 0.2	2.4 ± 1.3
Escherichia coli	7.7^a	34.6^b	1.6 ± 0.4	2.5 ± 1.2
Lactic acid bacteria	58.0^a	96.1^b	3.4 ± 1.5	3.5 ± 0.3
Pseudomonas	0	0	<2.7	<2.7
Salmonella	0	0	<2.7	<2.7

Values with different superscripts in a row are significantly different ($p < 0.05$). For Pseudomonas and Salmonella, the mean bacterial concentrations (log10 CFU/600 cm^2) were below the detection threshold for the method used. σ, standard deviation. A sample is considered positive if a bacterial count could been performed.

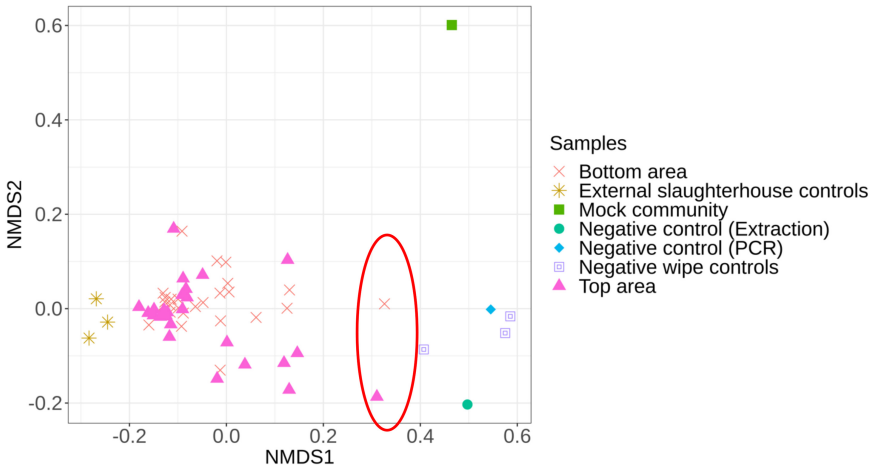


FIGURE 2 | Non-metric multidimensional scaling (NMDS) plot illustrating the microbiota structure between all samples. The red circle identifies the samples removed from the analysis due to their proximity to the negative controls.

(**Figure 2**). These two samples were therefore removed from the subsequent analysis.

The microbiota general composition at the phylum, family, and genus levels is illustrated on **Figure 3**. Overall, regardless of the batch, the major phyla, families, and genus were the same for samples recovered from both the top and bottom areas (**Figures 3B,C**). In total, 663 genera were assigned from all the sequences analyzed, but only 9 genus had a relative abundance greater than 5% (**Figure 3C**).

Multivariate association using linear models (MaAslin2) was performed in order to determine which microbial taxa were preferentially associated with the bottom or the top carcass surface, regardless of the batch. Indeed, any microbial taxa were associated with one of the six batches sampled. Samples recovered from the top carcass areas were positively associated with Proteobacteria. Firmicutes were positively associated with samples collected from the bottom area. At the family level, the microbiota of the top area of the carcass was significantly associated with *Bradyrhizobiaceae*, *Caulobacteraceae*, and *Planctomycetaceae*, while families of *Halomonadaceae*, *Corynebacteriaceae*, *Pasteurellaceae*, *and Aerococcaceae* were found to be significantly associated with its bottom counterpart. At the genus level, *Phenylobacterium* and *Bradyrhizobium* were found to be significantly associated with the top area of the carcass. The presence of *Halomonas*, *Lactococcus*, *Aerococcus*, and *Corynebacterium* was significantly associated with samples collected from the bottom area. All significant results ($p < 0.05$) are available in the **Supplementary Table 1**.

Microbiota Diversity of the Pig Carcass Surface

Alpha diversity analysis, which describes the bacterial richness and distribution within a sample, was used to compare the surface microbiota between the top and bottom areas of the pig carcasses sampled. Plots representing the alpha diversity measures obtained are shown in **Figure 4**. No significant differences were observed (t-test $p > 0.05$) between the 26 top surface microbiota and the 26 surface microbiota bottom areas samples.

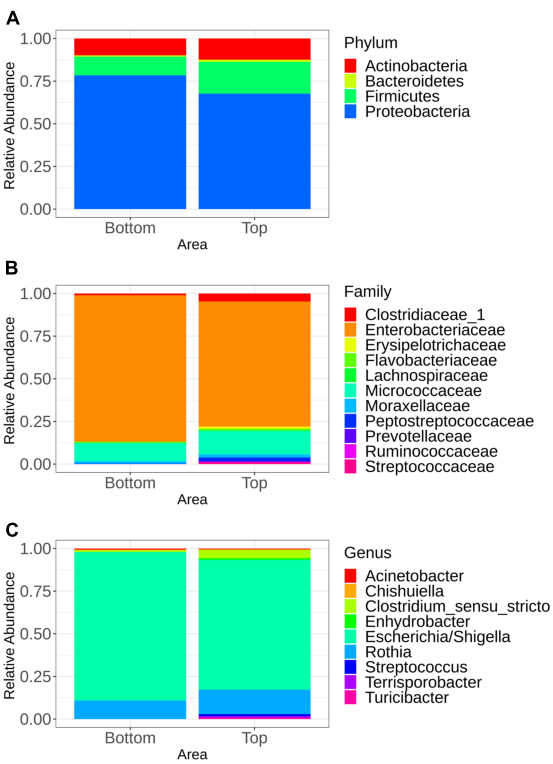


FIGURE 3 | Mean relative abundance of the major bacterial groups at the phylum level **(A)**, family level **(B)**, and genus level **(C)** identified in samples representing the top and bottom areas of the pig carcasses. Only bacterial communities representing at least 5% of carcass surface microbiota are shown.

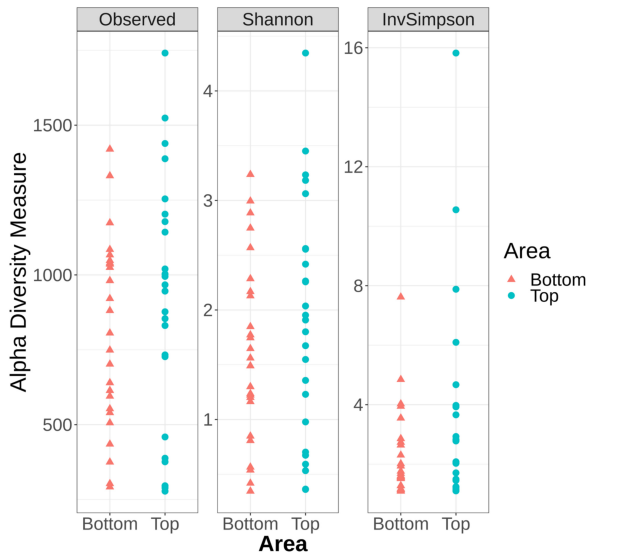


FIGURE 4 | Alpha diversity measures among the top and the bottom areas of the pig carcasses sampled using Observed, Shannon, and Inverse Simpson indices.

The alpha diversity measures from all top or bottom areas were compared between the six batches sampled and no significant differences between batches could be observed (Kruskal–Wallis test $p > 0.05$) (**Supplementary Figures 1,2**). These measures were also compared within each carcass of a same batch. Only the samples collected from the fourth batch sampled showed Observed number of ASVs (sequences of bacteria) and Inverse Simpson indices significantly higher for the top area of the carcass surface when compared to their bottom counterpart (**Supplementary Figure 2**).

Using Bray-Curtis and Jaccard distance matrices, results showed that the structure and membership of the bacterial community was similar (PERMANOVA $p > 0.05$) between the top and bottom areas of the pig carcasses sampled (**Figure 5A**). This same observation was made for the comparison of the top or bottom areas among the different batches sampled (**Figures 5B,C**).

DISCUSSION

In this study, the microbiota of 26 pig carcass surfaces were analyzed by culture-dependent methods and high-throughput sequencing in order to describe the carcass surface microbiota composition and to determine if an association could be observed between the carcass surface microbiota (top and bottom) and the animal's origin (batch).

Mesophilic total bacteria, usually regarded as an indicator of the hygiene conditions of the entire meat production process, showed counts that were similar to those reported in previous studies, with values ranging from 3 to 6.4 log CFU/cm^2 (Martínez et al., 2010; Bohaychuk et al., 2011; Piras et al., 2014). In our study, significant differences between counts for the top and bottom areas of the pig carcasses were identified on the same day of slaughter in the slaughterhouse, regardless of the batch. Similar findings were reported in the literature but the samples were collected over the course of several visits (Spescha et al., 2006; Biasino et al., 2018). The impact of animal batch has been reported for pathogenic bacteria like *Salmonella* and *Yersinia enterocolitica* (Gonzales-Barron et al., 2013; Vanantwerpen et al., 2015) but, to the best of our knowledge, this is the first report for mesophilic total bacteria.

Our results showed similar levels of Enterobacteria and *E. coli*—usually used to assess fecal contamination—to those reported by other studies, with mean levels of 2 log CFU/cm^2. In several other studies, the bottom parts of the carcasses were reported to be more contaminated than the top after evisceration (Spescha et al., 2006; Zweifel et al., 2008; Martínez et al., 2010; Wheatley et al., 2014).

Lactic acid bacteria and *Pseudomonas* are psychotropic bacteria responsible for meat spoilage (Salifou et al., 2013). A recent study that characterized the pig carcass surface microbiota according to different areas (Peruzy et al., 2021) observed levels of lactic acid bacteria similar to those observed in the present study, around 3.61 log CFU/cm^2. *Pseudomonas* was not detected in any of our samples and the probable growth of this psychrotrophic bacteria had perhaps not started before the cooling process. Indeed, in many studies, *Pseudomonas spp*. was mainly found on refrigerated pork products after cooling (EFSA Panel on Biological Hazards [BIOHAZ], 2016; Stellato et al., 2016).

In the present study, none of the carcasses sampled was contaminated by *Salmonella*, preventing us from making any association between the presence of the pathogen and the microbiota composition. Other studies are reporting different prevalence levels for *Salmonella* according to meat processing. For example, Piras et al. (2014) and Biasino et al. (2018) reported 64% and 18% of *Salmonella*-positive carcasses after evisceration and immediately after cooling respectively. The application of good slaughtering practices combined with the absence of *Salmonella* in the intestines of pigs slaughtered on the sampling day could explain this negative *Salmonella* status.

Microbiota analysis using high-throughput sequencing revealed similar bacterial communities between the top and bottom areas of the pig carcasses at the phylum level, while comparison at family and genus level showed significant differences in the relative abundance between these two areas. No information pertaining to the difference of phylum and family relative abundances between the top and bottom areas of pork carcasses seems to be available in the scientific literature. However, according to Peruzy et al. (2021), the bacterial community on four carcass areas (jowl, belly, back, ham) was dominated by the same bacterial genera that were also observed in the current study, i.e., *Escherichia* and *Rothia*. These similarities are not surprising as these genera reside in the oral and intestinal microbiota of pigs as well as in slaughterhouse environments.

In our study, *Terrisporobacter*, *Escherichia-Shigella*, *Turicibacter*, *Clostridium sensu stricto*, and *Streptococcus* represented the most abundant bacterial populations found on

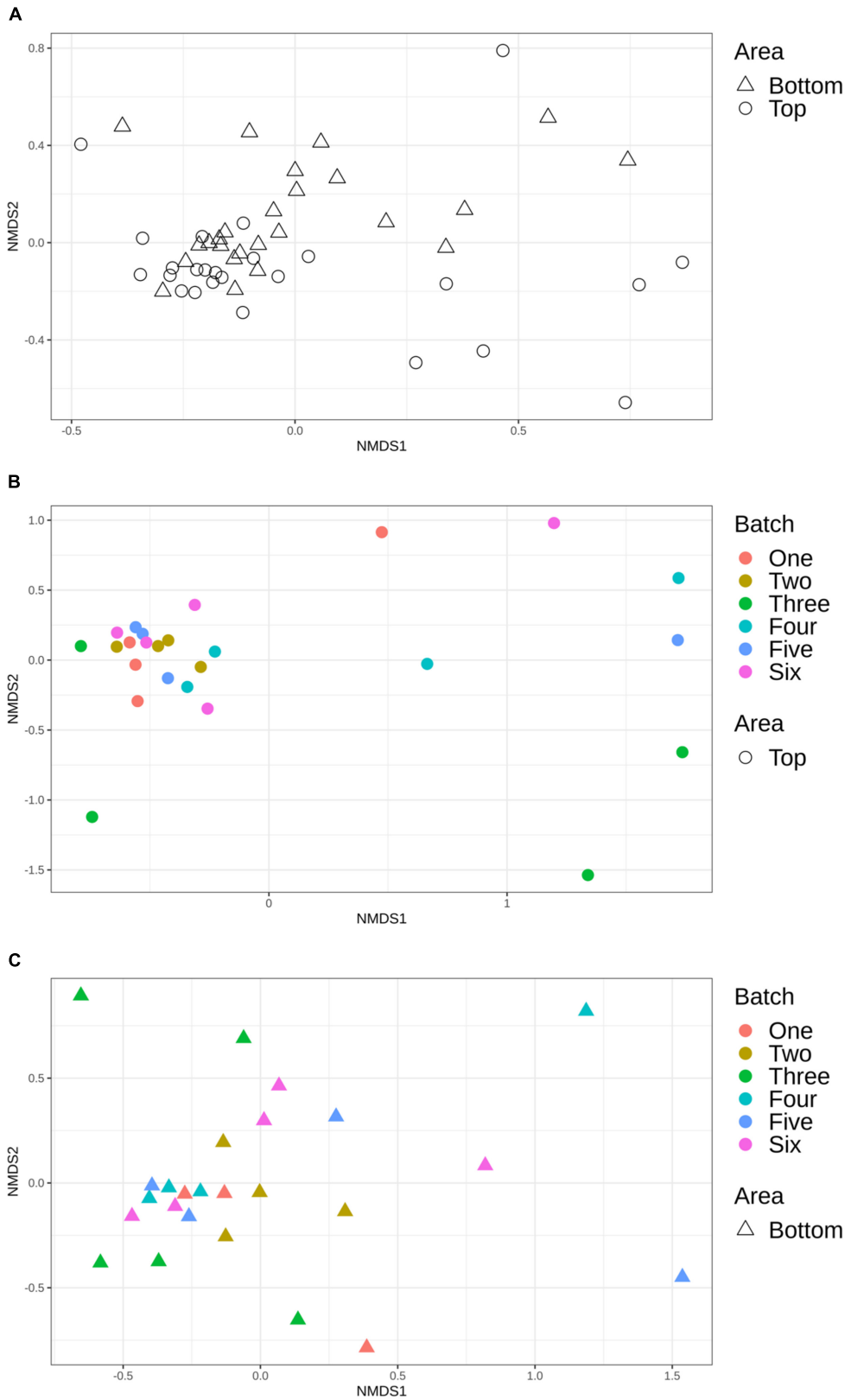


FIGURE 5 | Non-metric multidimensional scaling (NMDS) plot illustrating the microbiome structure of the pig carcass surface according to area **(A)**, top area according to batch **(B)**, and bottom area according to batch **(C)**.

carcass surfaces, representing more than 90% of the microbiota analyzed. It has been reported that these bacterial populations were present in the gut microbiota of pigs (Quan et al., 2018). Indeed, according to this study by Quan et al. (2018), *Escherichia-Shigella* (23.1%), *Terrisporobacter* (17.9%), and *Clostridium sensustricto1* (12.9%) were most prevalent in the pig ileum, and *Streptococcus* (8.0%) was one of the most prevalent genera in the colon of this animal species. *Turicibacter* was most present in the jejunum and ileum (Crespo-Piazuelo et al., 2018). In our study, *Terrisporobacter*, *Streptococcus*, and *Turicibacter* were only found on the top area of the sampled carcasses. It is worth noting that carcasses were sampled before washing and cooling in the current study. This could probably explain why fecal contaminants were found on the top surface area since washing usually involves water running down from the top to the bottom part of the carcass, creating a higher risk for contamination in this area (Bolton et al., 2002). Furthermore, several families (*Bradyrhizobiaceae*, *Caulobacteraceae*, *Planctomycetace*, *Halomonadaceae*, *Corynebacteriaceae*, *Pasteurellaceae*, *Aerococcaceae*) and genus (*Phenylobacterium*, *Bradyrhizobium*, *Halomonas*, *Lactococcus*, *Aerococcus*, *Corynebacterium*) have been observed and each of them were statistically associated with the top or the bottom carcass surface microbiota. These bacterial populations were all previously reported as members of the pig gut microbiota (Fan et al., 2017; Zwirzitz et al., 2019). The description of bacteria isolated from specific carcass areas allowed for a better understanding of the dispersion of bacteria on the carcass surface as well as for the identification of contamination origin.

The *Flavobacterium* family which is responsible for the occurrence of rancid odors causing meat spoilage—which has been suggested to originate from worker gloves during evisceration—was also present on the surface of pig carcasses sampled in the current study (Doulgeraki et al., 2012; Zwirzitz et al., 2020; Wu et al., 2021).

It is well known that the major source of contamination of the carcass surface is evisceration (Biasino et al., 2018), but the removal of tonsils, tongue, and gallbladder are other important contaminating steps during primary processing (Jakobsen et al., 2019). In the current study, the analysis of the carcass surface microbiota revealed the presence of bacteria such as *Acinetobacter*, *Enhydrobacter*, and *Rothia*. *Acinetobacter* was naturally found in the tonsils of pigs (Jakobsen et al., 2019) and was associated with meat spoilage (Iulietto et al., 2015). According to Bridier et al. (2019), *Acinetobacter* was also found in the slaughterhouse environment, especially at the dehairing step, and *Enhydrobacter* was present in the neck clipper environment. In another study, *Acinetobacter* and *Rothia* were observed to be the most dominant genera of the carcass surface microbiota (Peruzy et al., 2021; Wu et al., 2021) and were both common inhabitants of the oral microbiota of pigs.

Finally, in our study, the richness and the structure of carcass surface microbiota appeared similar between the top and bottom areas and between the six different batches sampled. According to the study by Peruzy et al. (2021), alpha diversity indices were not different between the ham (corresponding to the top area in our study) and the jowl (bottom area). Despite bacterial count results that show a difference of the mesophilic aerobic bacteria between the top and bottom areas of the pig carcasses sampled during the current study, the beta diversity also seemed to be comparable between batches. This indicates that the differences in the intestinal microbiota reported in the literature between animals originating from different farms were not replicated on the carcass, suggesting that the primary processing, until the end of carcass dressing, globally standardized the pig carcass surface in terms of microbial diversity. This is important for pig farming as the modulation of gut microbiota to improve feed efficiency is being explored at the farm level. Based on our observations however, even with optimal primary processing practices, it seems that these attempts of gut microbiota modifications may not have any profound effects on carcass microbiota. This is also important for the control of bacterial carcass surface contamination at slaughterhouse as sources of contamination other than intestinal contents, such as processing line, slaughterhouse equipment or worker hands also contributes to the surface microbiota.

The results presented are specific to this study, which was conducted in a single slaughterhouse, on the same day, and at a single sampling point. This limits our ability to draw conclusions regarding the effect of farm origin in a universal manner. Differences may appear later in the processing, which deserves further attention. For example, during and after the cooling period, psychrotrophic bacteria are recognized to grow (Zwirzitz et al., 2020) and more differences may be observed beyond this point. The stability and impact of pig origin on final meat cuts should therefore also be investigated.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA757709.

AUTHOR CONTRIBUTIONS

CB: methodology, software, visualization, writing—original draft, and writing—review and editing. PF: conceptualization, supervision, writing—review and editing, and funding acquisition. M-LG: writing—reviewing and editing the manuscript. GL-G: resources, software, and writing—review and editing. FS-B: resources, software, and writing—review and editing. JL: resources, project administration, and writing—review and editing. AT: conceptualization, methodology, validation, supervision, project administration, resources, and writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and F. Ménard, Olymel LP, financial partner in the Industrial Research

Chair in Meat Safety and the Swine and Poultry Infectious Diseases Research Center (CRIPA).

ACKNOWLEDGMENTS

We thank JL from F. Ménard, Olymel LP, for her collaboration and also the staff from F. Ménard for their support in conducting sampling. We also thank William P. Thériault and Jessika Beauchemin for their help at the slaughterhouse and in the laboratory. We also thank Kaarla Sundström for revising this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.849883/full#supplementary-material

REFERENCES

Biasino, W., De Zutter, L., Mattheus, W., Bertrand, S., Uyttendaele, M., and Van Damme, I. (2018). Correlation between slaughter practices and the distribution of *Salmonella* and hygiene indicator bacteria on pig carcasses during slaughter. *Food Microbiol.* 70, 192–199. doi: 10.1016/j.fm.2017.10.003

Bohaychuk, V. M., Gensler, G. E., and Barrios, P. R. (2011). Microbiological baseline study of beef and pork carcasses from provincially inspected abattoirs in Alberta, Canada. *Can. Vet. J.* 52, 1095–1100.

Bolton, D. J., Pearce, R. A., Sheridan, J. J., Blair, I. S., McDowell, D. A., and Harrington, D. (2002). Washing and chilling as critical control points in pork slaughter hazard analysis and critical control point (HACCP) systems. *J. Appl. Microbiol.* 92, 893–902. doi: 10.1046/j.1365-2672.2002.01599.x

Bonardi, S. (2017). *Salmonella* in the pork production chain and its impact on human health in the European Union. *Epidemiol. Infect.* 145, 1513–1526. doi: 10.1017/S095026881700036X

Borch, E., Nesbakken, T., and Christensen, H. (1996). Hazard identification in swine slaughter with respect to foodborne bacteria. *Int. J. Food Microbiol.* 30, 9–25. doi: 10.1016/0168-1605(96)00988-9

Bridier, A., Le Grandois, P., Moreau, M.-H., Prénom, C., Le Roux, A., Feurer, C., et al. (2019). Impact of cleaning and disinfection procedures on microbial ecology and *Salmonella* antimicrobial resistance in a pig slaughterhouse. *Sci. Rep.* 9:12947. doi: 10.1038/s41598-019-49464-8

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Casaburi, A., Piombino, P., Nychas, G.-J., Villani, F., and Ercolini, D. (2015). Bacterial populations and the volatilome associated to meat spoilage. *Food Microbiol.* 45, 83–102. doi: 10.1016/j.fm.2014.02.002

Crespo-Piazuelo, D., Estellé, J., Revilla, M., Criado-Mesas, L., Ramayo-Caldas, Y., Óvilo, C., et al. (2018). Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions. *Sci. Rep.* 8:12727. doi: 10.1038/s41598-018-30932-6

Doulgeraki, A. I., Ercolini, D., Villani, F., and Nychas, G.-J. E. (2012). Spoilage microbiota associated to the storage of raw meat in different conditions. *Int. J. Food Microbiol.* 157, 130–141. doi: 10.1016/j.ijfoodmicro.2012.05.020

EFSA Panel on Biological Hazards [BIOHAZ] (2016). Growth of spoilage bacteria during storage and transport of meat. *EFSA J.* 14:4523. doi: 10.2903/j.efsa.2016.4523

Fan, P., Liu, P., Song, P., Chen, X., and Ma, X. (2017). Moderate dietary protein restriction alters the composition of gut microbiota and improves ileal barrier function in adult pig model. *Sci. Rep.* 7:43412. doi: 10.1038/srep43412

Fouhse, J. M., Zijlstra, R. T., and Willing, B. P. (2016). The role of gut microbiota in the health and disease of pigs. *Anim. Front.* 6, 30–36. doi: 10.2527/af.2016-0031

Gonzales-Barron, U., Cadavez, V., Sheridan, J. J., and Butler, F. (2013). Modelling the effect of chilling on the occurrence of *Salmonella* on pig carcasses at study, abattoir and batch levels by meta-analysis. *Int. J. Food Microbiol.* 163, 101–113. doi: 10.1016/j.ijfoodmicro.2013.03.006

Gram, L., Ravn, L., Rasch, M., Bruhn, J. B., Christensen, A. B., and Givskov, M. (2002). Food spoilage—interactions between food spoilage bacteria. *Int. J. Food Microbiol.* 78, 79–97. doi: 10.1016/S0168-1605(02)00233-7

Hasan, N., and Yang, H. (2019). Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* 7:e7502. doi: 10.7717/peerj.7502

Iulietto, M. F., Sechi, P., Borgogni, E., and Cenci-Goga, B. T. (2015). Meat spoilage: a critical review of a neglected alteration due to ropy slime producing bacteria. *Ital. J. Anim. Sci.* 14:4011. doi: 10.4081/ijas.2015.4011

Jakobsen, A. M., Bahl, M. I., Buschhardt, T., Hansen, T. B., Al-Soud, W. A., Brejnrod, A. D., et al. (2019). Bacterial community analysis for investigating bacterial transfer from tonsils to the pig carcass. *Int. J. Food Microbiol.* 295, 8–18. doi: 10.1016/j.ijfoodmicro.2019.02.003

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13

Larivière-Gauthier, G., Thibodeau, A., and Fravalo, P. (2017). Reduction of *Salmonella* shedding by sows during gestation in relation to its fecal microbiome. *Front. Microbiol.* 8:2219. doi: 10.3389/fmicb.2017.02219

Larivière-Gauthier, G., Thibodeau, A., Letellier, A., Yergeau, É, and Fravalo, P. (2019). *Salmonella* shedding status of the sow affects the microbiota of their piglets at weaning. *J. Appl. Microbiol.* 126, 411–423. doi: 10.1111/jam.14139

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., et al. (2021). Multivariable association discovery in population-scale meta-omics studies. *bioRxiv* [Preprint]. doi: 10.1101/2021.01.20.427420

MAPAQ (2019). Plan de Surveillance des Contaminants dans les Produitsalimentaires Vendus Au Qu.Bec Bilan 2011-2016.

Martínez, B., Celda, M. F., Anastasio, B., García, I., and López-Mendoza, M. C. (2010). Microbiological sampling of carcasses by excision or swabbing with three types of sponge or gauze. *J. Food Prot.* 73, 81–87. doi: 10.4315/0362-028X-73.1.81

Martínez-Avilés, M., Garrido-Estepa, M., Álvarez, J., and de la Torre, A. (2019). *Salmonella* surveillance systems in swine and humans in Spain: a review. *Vet. Sci.* 6:20. doi: 10.3390/vetsci6010020

McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217

Oksanen, J., Guillaume, Blanchet, F., Friendly, M., Kindt, R., Legendre, P., et al. (2018). *Vegan: Community Ecology Package. R Package*.

Pala, C., Tedde, T., Salza, S., Uda, M. T., Lollai, S., Carboni, V., et al. (2019). Epidemiological survey on the prevalence of *Salmonella* spp. in the Sardinian pig production chain, using real-time PCR screening method. *Ital. J. Food Saf.* 8:7843. doi: 10.4081/ijfs.2019.7843

Peruzy, M. F., Houf, K., Joossens, M., Yu, Z., Proroga, Y. T. R., and Murru, N. (2021). Evaluation of microbial contamination of different pork carcass areas through culture-dependent and independent methods in small-scale slaughterhouses. *Int. J. Food Microbiol.* 336:108902. doi: 10.1016/j.ijfoodmicro.2020.108902

Piras, F., Fois, F., Mazza, R., Putzolu, M., Delogu, M. L., Lochi, P. G., et al. (2014). *Salmonella* prevalence and microbiological contamination of pig carcasses and slaughterhouse environment. *Ital. J. Food Saf.* 3:4581. doi: 10.4081/ijfs.2014.4581

Pothakos, V., Devlieghere, F., Villani, F., Björkroth, J., and Ercolini, D. (2015). Lactic acid bacteria and their controversial role in fresh meat spoilage. *Meat Sci.* 109, 66–74. doi: 10.1016/j.meatsci.2015.04.014

Quan, J., Cai, G., Ye, J., Yang, M., Ding, R., Wang, X., et al. (2018). A global comparison of the microbiome compositions of three gut locations in commercial pigs with extreme feed conversion ratios. *Sci. Rep.* 8:4536. doi: 10.1038/s41598-018-22692-0

Ritchie, H., and Roser, M. (2017). *Meat and Dairy Production. OurWorldInData.org*. Available online at: https://ourworldindata.org/meat-production

Rivas, T., Vizcaíno, J. A., and Herrera, F. J. (2000). Microbial contamination of carcasses and equipment from an Iberian pig Slaughterhouse. *J. Food Prot.* 63, 1670–1675. doi: 10.4315/0362-028X-63.12.1670

Regulation (EC) No 2073/2005 (2005). *Commission Regulation (EC) No 2073/2005 of 15 November 2005 on Microbiological Criteria for Foodstuffs (Text with EEA Relevance)*. Off. J. Eur. Union.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584

Salifou, C., Boko, K., Ahounou, G., Tougan, P., Kassa, S., Houaga, I., et al. (2013). Diversité de la microflore initiale de la viande et sécurité sanitaire des consommateurs. *Int. J. Biol. Chem. Sci.* 7, 1351–1369. doi: 10.4314/ijbcs.v7i3.41

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Self, J. L., Luna-Gierke, R. E., Fothergill, A., Holt, K. G., and Vieira, A. R. (2017). Outbreaks attributed to pork in the United States, 1998–2015. *Epidemiol. Infect.* 145, 2980–2990. doi: 10.1017/S0950268817002114

Spescha, C., Stephan, R., and Zweifel, C. (2006). Microbiological contamination of pig carcasses at different stages of slaughter in two European Union–approved abattoirs. *J. Food Prot.* 69, 2568–2575. doi: 10.4315/0362-028X-69.11.2568

Stellato, G., La Storia, A., De Filippis, F., Borriello, G., Villani, F., and Ercolini, D. (2016). Overlap of spoilage-associated microbiota between meat and the meat processing environment in small-scale and large-scale retail distributions. *Appl. Environ. Microbiol.* 82, 4045–4054. doi: 10.1128/AEM.00793-16

Vanantwerpen, G., Berkvens, D., Van Damme, I., De Zutter, L., and Houf, K. (2015). Assessment of risk factors for a high within-batch prevalence of *Yersinia enterocolitica* in pigs based on microbiological analysis at slaughter. *Foodborne Pathog. Dis.* 12, 571–575. doi: 10.1089/fpd.2014.1897

Wang, X., Tsai, T., Deng, F., Wei, X., Chai, J., Knapp, J., et al. (2019). Longitudinal investigation of the swine gut microbiome from birth to market reveals stage and growth performance associated bacteria. *Microbiome* 7:109. doi: 10.1186/s40168-019-0721-7

Wheatley, P., Giotis, E. S., and McKevitt, A. I. (2014). Effects of slaughtering operations on carcass contamination in an Irish pork production plant. *Ir. Vet. J.* 67:1. doi: 10.1186/2046-0481-67-1

Wu, J., Li, R., Zhang, M., Shan, K., Jia, X., Zhao, D., et al. (2021). Microbiota changes on the surface of pig carcasses during refrigerated transportation and marketing. *Food Mater. Res.* 1:4. doi: 10.48130/FMR-2021-0004

Yang, H., Xiao, Y., Wang, J., Xiang, Y., Gong, Y., Wen, X., et al. (2018). Core gut microbiota in Jinhua pigs and its correlation with strain, farm and weaning age. *J. Microbiol.* 56, 346–355. doi: 10.1007/s12275-018-7486-8

Zweifel, C., Fischer, R., and Stephan, R. (2008). Microbiological contamination of pig and cattle carcasses in different small-scale Swiss abattoirs. *Meat Sci.* 78, 225–231. doi: 10.1016/j.meatsci.2007.06.025

Zwirzitz, B., Pinior, B., Metzler-Zebeli, B., Handler, M., Gense, K., Knecht, C., et al. (2019). Microbiota of the gut-lymph node axis: depletion of mucosa-associated segmented filamentous bacteria and enrichment of *Methanobrevibacter* by colistin sulfate and Linco-Spectin in pigs. *Front. Microbiol.* 10:599. doi: 10.3389/fmicb.2019.00599

Zwirzitz, B., Wetzels, S. U., Dixon, E. D., Stessl, B., Zaiser, A., Rabanser, I., et al. (2020). The sources and transmission routes of microbial populations throughout a meat processing facility. *NPJ Biofilms Microbiomes* 6:26. doi: 10.1038/s41522-020-0136-z

Conflict of Interest: JL was employed by F. Ménard, Olymel LP.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fmicb.2022.884721

Genomic Characterization of *Cronobacter* spp. and *Salmonella* spp. Strains Isolated From Powdered Infant Formula in Chile

OPEN ACCESS

Edited by:
Jerome Combrisson,
Mars, United States

Reviewed by:
Luxin Wang,
University of California,
Davis, United States
Andrea Isabel Moreno Switt,
Pontificia Universidad Católica de
Chile, Chile

***Correspondence:**
Julio Parra-Flores
juparra@ubiobio.cl
Stephen Forsythe
sforsythe4j@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: 26 February 2022
Accepted: 04 May 2022
Published: 02 June 2022

Citation:
Parra-Flores J, Holý O, Acuña S, Lepuschitz S, Pietzka A, Contreras-Fernández A, Chavarría-Sepulveda P, Cruz-Córdova A, Xicohtencatl-Cortes J, Mancilla-Rojano J, Castillo A, Ruppitsch W and Forsythe S (2022) Genomic Characterization of Cronobacter spp. and Salmonella spp. Strains Isolated From Powdered Infant Formula in Chile. Front. Microbiol. 13:884721. doi: 10.3389/fmicb.2022.884721

Julio Parra-Flores[1*†], Ondřej Holý[2†], Sergio Acuña[3], Sarah Lepuschitz[4†], Ariane Pietzka[4†], Alejandra Contreras-Fernández[5], Pamela Chavarría-Sepulveda[1], Ariadnna Cruz-Córdova[6†], Juan Xicohtencatl-Cortes[6†], Jetsi Mancilla-Rojano[6,7†], Alejandro Castillo[8†], Werner Ruppitsch[4†] and Stephen Forsythe[9*†]

[1]Department of Nutrition and Public Health, Universidad del Bío-Bío, Chillán, Chile, [2]Science and Research Centre, Faculty of Health Sciences, Palacký University Olomouc, Olomouc, Czechia, [3]Department of Food Engineering, Universidad del Bío-Bío, Chillán, Chile, [4]Austrian Agency for Health and Food Safety, Institute for Medical Microbiology and Hygiene, Vienna, Austria, [5]Food Quality Testing and Certification Laboratory, Universidad del Bío-Bío, Chillán, Chile, [6]Intestinal Bacteriology Research Laboratory, Hospital Infantil de México Federico Gómez, Mexico City, Mexico, [7]Faculty of Medicine, Biological Sciences Graduate Program, Universidad Nacional Autónoma de México, Mexico City, Mexico, [8]Department of Nutrition and Food Science, Texas A&M University, College Station, TX, United States, [9]Foodmicrobe.com, Nottingham, United Kingdom

This study characterized five *Cronobacter* spp. and six *Salmonella* spp. strains that had been isolated from 155 samples of powdered infant formula (PIF) sold in Chile and manufactured in Chile and Mexico in 2018–2020. Two strains of *Cronobacter sakazakii* sequence type (ST) ST1 and ST31 (serotypes O:1 and O:2) and one strain of *Cronobacter malonaticus* ST60 (O:1) were identified. All *Salmonella* strains were identified as *Salmonella* Typhimurium ST19 (serotype O:4) by average nucleotide identity, ribosomal multilocus sequence typing (rMLST), and core genome MLST (cgMLST). The *C. sakazakii* and *C. malonaticus* isolates were resistant to cephalothin, whereas the *Salmonella* isolates were resistant to oxacillin and ampicillin. Nineteen antibiotic resistance genes were detected in the *C. sakazakii* and *C. malonaticus* isolates; the most prevalent were mcr-9.1, bla_{CSA}, and bla_{CMA}. In *Salmonella*, 30 genes encoding for aminoglycoside and cephalosporin resistance were identified, including *aac(6′)-Iaa*, β-*lactamases ampH*, *ampC1*, and *marA*. In the *Cronobacter* isolates, 32 virulence-associated genes were detected by WGS and clustered as flagellar proteins, outer membrane proteins, chemotaxis, hemolysins, invasion, plasminogen activator, colonization, transcriptional regulator, survival in macrophages, use of sialic acid, and toxin-antitoxin genes. In the *Salmonella* strains, 120 virulence associated genes were detected, adherence, magnesium uptake, resistance to antimicrobial peptides, secretion system, stress protein, toxin, resistance to complement killing, and eight pathogenicity islands. The *C. sakazakii* and *C. malonaticus* strains harbored I-E and I-F CRISPR-Cas systems and carried Col(pHHAD28) and IncFIB(pCTU1) plasmids, respectively. The *Salmonella* strains harbored type I-E CRISPR-Cas systems and carried IncFII(S) plasmids. The presence of *C. sakazakii* and *Salmonella* in PIF is a

health risk for infants aged less than 6 months. For this reason, sanitary practices should be reinforced for its production and retail surveillance.

Keywords: *Cronobacter sakazakii*, *Cronobacter malonaticus*, *Salmonella* Typhimurium, powdered infant formula, virulence, resistance genes, whole-genome sequencing, CRISPR-Cas

INTRODUCTION

The need to ensure the safety of powdered infant formula (PIF) led the FAO/WHO to establish the microbiological or epidemiological relationship of microbial agents found in PIF with infant infection. They identified three categories of microorganisms based on evidence of a causal relationship between the presence of these microorganisms and the disease they cause. The first category of microorganisms with clear causality were identified as *Cronobacter* spp. and *Salmonella enterica*. The second consists of microorganisms for which causality is possible but has not yet been demonstrated. These were mainly from the *Enterobacteriaceae* family, but also included *Acinetobacter*. Finally, the third involves microorganisms for which causality is less likely or has not yet been shown and has not been identified in PIF. Based on this the FAO/WHO recommended the absence of *Cronobacter* spp. and *Salmonella* in PIF for target age less than 6 months (FAO/WHO, 2004, 2006; Forsythe, 2018).

Cronobacter is a genus of bacterial pathogens consisting of seven species: *C. sakazakii*, *C. malonaticus*, *C. universalis*, *C. turicensis*, *C. muytjensii*, *C. dublinensis*, and *C. condimenti* (Iversen et al., 2008; Joseph et al., 2012; Stephan et al., 2014). The species with the greatest clinical significance are *C. sakazakii* and *C. malonaticus* and have been reported in cases and outbreaks associated with PIF in infants (Forsythe, 2018; Parra-Flores et al., 2021a). At present in the United States, three cases of illness caused by PIF contaminated with *C. sakazakii*, resulting in one fatality, and one case of *Salmonella* Newport are being investigated; this has prompted an international voluntary recall of these PIFs by the manufacture (U.S FDA, 2022). The severity of the clinical condition has been associated with the presence of virulence factors encoded on plasmids (Shi et al., 2018; Aly et al., 2019), adherence and invasion traits (Cruz et al., 2011; Parra-Flores et al., 2018a; Holý et al., 2021), and various other genes such as *aut*, *cpA*, *fliC*, *hly*, *ompA*, *sip*, *plas*, and *inv* (Cruz et al., 2011; Franco et al., 2011; Aldubyan et al., 2017; Holý et al., 2019). Other factors are the use of sialic acid as a carbon source, capsule composition and the presence of its capsule, and endotoxin production (Ogrodzki and Forsythe, 2015). Another important aspect is the resistance to β-lactam antibiotics such as cephalothin, cefotaxime, ceftazidime, and ampicillin in addition to the presence of resistance genes such as *marA*, *glpT*, *ampH*, *blaCSA*, and *mcr* (Flores et al., 2011; Lee et al., 2012; Fei et al., 2017; Holý et al., 2021).

Salmonella enterica is a gram-negative, rod-shaped, facultative anaerobic genus. More than 2,600 serotypes belonging to *S. enterica* have been described worldwide que incluyen, which can cause diseases in humans and animals (Mezal et al., 2014). *Salmonella* is the most widely studied microbial pathogen, and can be isolated from a variety of foods, including PIF associated with disease outbreaks in infants (Angulo et al., 2008; Carrasco et al., 2012; Jourdan-da Silva et al., 2018). Gastroenteric *Salmonella* infections usual develop as self-limiting gastroenteritis, and antibiotic treatment is necessary only in severe cases more often associated with immunocompromised patients or those at the extremes of age such as infants (de Toro et al., 2014). Therefore, the emergence of strains that are resistant to β-lactams and cephalosporins is a relevant public health and food safety problem (Güerri et al., 2004; de Toro et al., 2011; Wang et al., 2019). In addition, *Salmonella* exhibits virulence factors that play a decisive role in systemic infections, such as pathogenicity islands (PAIs), invasion and adherence genes, and enterotoxin coding (Murugkar et al., 2003; Nayak et al., 2004; Huehn et al., 2010; Thung et al., 2018).

Salmonella and *Cronobacter* species are known persist in low-moisture foods such as milk powder and powdered infant formula for up to 2 years (Caubilla-Barron and Forsythe, 2007). Consequently, outbreaks due to the consumption of contaminated products have been reported (Forsythe, 2018; Jones et al., 2019). *Cronobacter* can survive spray-drying and persist in the manufacturing environment as biofilms. Genotyping has shown the persistence of specific *Salmonella* and *Cronobacter* strains within production facilities for many years (Craven et al., 2010; Jones et al., 2019). Contamination may occur post-pasteurization due to the addition of contaminated ingredients (FAO/WHO, 2008).

Whole-genome sequencing (WGS) has facilitated the in-depth study of pathogenic organisms by generating extensive information that helps to determine relationships and taxonomic differences between them (Leopold et al., 2014). It is not only used for isolate identification, but also extensive profiling and genotyping; such as conventional 7-loci multilocus sequence typing (MLST), core genome MLST (cgMLST) and/or single nucleotide polymorphism (SNP) analysis, molecular serotyping, CRISPR-Cas array profiling, and detection of genes associated with antibiotic resistance and virulence. Consequently, more precise epidemiological links can be established (Marraffini, 2013; Leopold et al., 2014; Uelze et al., 2020). Therefore, the analysis of the complete genomes and their comparison enables a more complete analysis of the pathogenesis process of *C. sakazakii* (Lehner et al., 2018).

In 2017, a recall of powdered formula samples contaminated with *Cronobacter* occurred in Chile (Parra-Flores et al., 2018b). This situation led to the incorporation of microbial criteria ($n = 30$; $c = 0$) for *Cronobacter* spp. in PIF intended for consumption by infants aged less than 12 months into the Chilean Food Sanitary Regulations (RSA; Parra-Flores et al., 2018b), given that the microbiological criteria in the RSA for

Salmonella was already defined ($n = 10$; $c = 0$). This study considers the safety of PIF from 2018 to 2020 with the objective of performing a genomic characterization of five *Cronobacter* spp. and six *Salmonella* strains isolated from PIF sold in Chile. These PIF had been and manufactured in Chile and Mexico.

MATERIALS AND METHODS

Sampling

A total of 155 PIF samples from two commercial brands whose main ingredient was casein and whey were analyzed. Of these, 80 PIF samples were made in Chile and 75 PIF samples were made in Mexico. The experimental units, milk cans, were obtained monthly from supermarkets and pharmacies because these products are replenished monthly. In addition, this allowed obtaining greater variability in terms of the origin of the production batch.

Isolation and Identification

Cronobacter were isolated according to the method described by Iversen and Forsythe (2004). For each sample, 225 ml buffered peptone water (BPW) was added to 25 g PIF, homogenized in a stomacher at a mean velocity for 60 s, and incubated at 37°C. For *Cronobacter* spp., 10 ml of each sample was inoculated after incubation at 37°C for 24 h in 90 ml *Enterobacteriaceae* enrichment broth (BD Difco, Sparks, MD, United States). A loop was extracted from the culture suspension and striated in Brilliance Chromogenic Agar CM 1035 (OxoidTermo-Fisher, Hampshire, United Kingdom) at 37°C for 20 h. Five strains, presumed to be colonies of *Cronobacter* spp. (green or blue), were striated in trypticase soy agar (BD Difco, Sparks, MD, United States) to verify their purity prior to future analyses. The isolated strains were maintained in a strain collection and stored at −80°C.

The official method (Instituto Nacional de Normalización, 2002) NCh 2675- ISO 6572-2 rev 2017 for *Salmonella* in Chile was used. From the initial incubated sample of 25 g PIF with 225 ml BPW for 24 h, 0.1 ml was inoculated in 10 ml of Rappaport-Vassiliadis with soya broth (RVS, Oxoid, Hampshire, United Kingdom) and 1 ml in Muller-Kauffmann-Tetrathionate-Novobiocin broth (MKTTn, Merck, Darmstadt, Germany) incubated for 24 ± 3 h at 41.5°C and 37°C, respectively. The colonies were then isolated in Xylose Lysine Deoxycoholate (XLD, Merck, Darmstadt, Germany) and Salmonella chromogenic agar incubated at 37°C for 24 h. The typical colonies were confirmed by biochemical tests. Both pathogens were identified by Matrix-Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry (MALDI-TOF MS; Bruker, Billerica, MA, United States) and with the MBT Compass IVD software 4.1.60 (Bruker) described by Lepuschitz et al. (2017).

Whole-Genome Sequencing

Before WGS, all the *Cronobacter* spp. and *Salmonella* spp. strains were cultured in Columbia blood agar plates (bioMérieux, Marcy-l'Étoile, France) at 37°C for 24 h. DNA was isolated from bacterial cultures with the MagAttract HMW DNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The amount of DNA was quantified on a Lunatic instrument (Unchained Labs, Pleasanton, CA, United States). Nextera XT chemistry (Illumina Inc., San Diego, CA, United States) was used to prepare sequencing libraries for a 2 × 300 bp paired-end sequencing run on an Illumina MiSeq sequencer. Samples were sequenced to achieve a minimum of 80-fold coverage using standard protocols by Illumina. The resulting FASTQ files were quality trimmed and *de novo* assembled with the SPAdes version 3.9.0. Contigs were filtered for a minimum of 5-fold coverage and 200 bp minimum length with Ridom SeqSphere+ software v. 7.8.0 (Ridom, Münster, Germany; Jünemann et al., 2013).

Sequence Type and Core Genome Multilocus Sequence Typing of *Cronobacter* spp. and *Salmonella* spp.

A total of 3,678 targets were used to establish the core genome multilocus sequence typing (cgMLST) scheme of *Cronobacter* spp. using strain ATCC BAA-894 as a reference using Ridom SeqSphere+ software v. 7.8.0 (Ridom, Münster, Germany; Jünemann et al., 2013). For *Salmonella*, the cgMLST scheme was performed based on the profile of 2,969 *S. enterica* target gene loci task template of the Ridom SeqSphere+ software v. 7.8.0 (Ridom, Münster, Germany). According to the cgMLST scheme, isolates were visualized with a minimum spanning tree (MST) to establish their genotypic relationships (Lepuschitz et al., 2019). In addition, the sequences of the seven housekeeping genes of the conventional MLST for *Cronobacter* spp. and *Salmonella* were extracted and cross-checked against the *Cronobacter* MLST database[1] (Baldwin et al., 2009) and *Salmonella* MLST[2] (Achtman et al., 2012), respectively. The *Cronobacter* strains are ID 3409–3413 in the *Cronobacter* PubMLST database and *Salmonella* are ID RID389119–RID389124 in the cgMLST database.

Determination of Serotypes

The *gnd* and *galF* genes that are specific to the *Cronobacter* serotype O region was determined by WGS sequence analysis with the BIGSdb tool available in the PubMLST database[3] and CroTrait WGS analysis (Wang L. et al., 2021). For *Salmonella*, the SeqSero 1.2 tool available at https://cge.cbs.dtu.dk/services/SeqSero/ was used (Zhang et al., 2015).

Antibiotic Resistance Profile

The disk diffusion method was used in accordance with the recommendations of the Clinical and Laboratory Standards Institute (CLSI, 2020). The commercial disks that were used consist of ampicillin (10 μg), amikacin (30 μg), cephalothin (30 μg), chloramphenicol (30 μg), ceftriaxone (30 μg), cefotaxime (30 μg), cefepime (30 μg), gentamicin (10 μg), levofloxacin (5 μg), netilmicin (30 μg), oxacillin (1 μg), and sulfamethoxazole-trimethoprim (1.25/23.75 μg). The characterization of the resistance/susceptibility

[1]https://pubmlst.org/organisms/cronobacter-spp/
[2]http://enterobase.warwick.ac.uk/species/index/senterica
[3]http://pubmlst.org/cronobacter/

profiles was determined according to the CLSI guidelines. The *Escherichia coli* ATCC 25922 and *Pseudomonas aeruginosa* ATCC 27853 strains were used as references.

Detection of Antibiotic Resistance and Virulence Genes

The existence of virulence genes was confirmed by applying the task template function in SeqSphere+ for the WGS data and the ResFinder tool from the Center of Genomic Epidemiology (CGE).[4] Thresholds for the target scanning procedure were set with a required identity of ≥90% to the reference sequence and an aligned reference sequence ≥99%. The Comprehensive Antibiotic Resistance Database (CARD) with the "perfect" and "strict" default settings for sequence analysis (Jia et al., 2017), the Task Template AMRFinderPlus 3.2.3 available in Ridom SeqSphere+ v. 7.8.0 software using the EXACT method at 100%, and BLAST alignment for protein identification available in the AMRFinderPlus database were used for antimicrobial resistance genes.

Detection of Plasmids and Mobile Genetic Elements

The PlasmidFinder 2.1 and MobileElementFinder 1.0 tools were used to detect plasmids and mobile genetic elements (MGEs). The selected minimum identity was 95% and 90%, respectively (Carattoli et al., 2014; Johansson et al., 2021).[4]

Profiling of CRISPR-Cas Loci Profiling

The search and characterization of CRISPR arrays and their association with Cas proteins was determined with CRISPR Detect and CRISPRminer (available at http://crispr.otago.ac.nz/CRISPRDetect/predict_crispr_array.html and http://www.microbiome-bigdata.com/CRISPRminer; Biswas et al., 2016; Zhang et al., 2018). The following parameters were applied: 18–55 pb repeated sequence length, 25–60 pb spacer length, 0.6–2.5 spacer sequence size as a function of repeated sequence size, and 60% maximum percentage similarity between spacers. The PHASTER program (available at https://phaster.ca/) was used to identify sequences associated with prophages within the study genomes, and the phages associated with the spacer sequences were determined with the CRISPRminer program. The types of CRISPR systems were determined with the CRISPRmap program (Lange et al., 2013). The CRISPRTarget program was used to determine the PAM (protospacer adjacent motif) sequences associated with each of the repeated sequences of the identified arrays.

RESULTS

Identification, Genotyping, and Antibiotic Resistance Profiles of *Cronobacter* and *Salmonella* Isolates

Overall positivity for the *Cronobacter* spp. samples was 6.25% (5/80) and 2.7% (2/75) for *Salmonella*. Of the five *Cronobacter* spp. presumptive strains, four were identified as *C. sakazakii* and one as *C. malonaticus*. All strains were isolated from different PIF batches from the same manufacturer and country (Chile). The six *Salmonella* strains were identified as *Salmonella* Typhimurium. These were from two different batches and tins but same manufacturer and country (Mexico; **Table 1**).

Three strains of *C. sakazakii* ST1 (CC1) and ST31, CC31 (serotypes *Csak*: O:1 and O:2, respectively), and one strain of *C. malonaticus* ST60, CC60 (O:1) were identified by average nucleotide identity, rMLST, and cgMLST (**Figure 1**; **Table 1**).

All *Salmonella* strains were identified as *S.* Typhimurium ST19 (CC19; serotype O:4). Two *Salmonella* strains (510539-21 and 510540-21) were identified as potential monophasic variants of *S.* Typhimurium according to SeqSero analysis (**Figure 2**; **Table 1**).

All the *C. sakazakii* and *C. malonaticus* strains were susceptible to 10 of the 12 evaluated antibiotics. However, 100% of the *Cronobacter* strains were resistant to cephalothin and 40% to ampicillin. Meanwhile, 100% of the *Salmonella* isolates were resistant to oxacillin, 83% to ampicillin, 66.6% to cephalothin, and 16.6% to gentamicin (**Table 2**).

Detection of Antibiotic Resistance and Virulence Genes

A total of 19 antibiotic resistance genes were detected in the *C. sakazakii* and *C. malonaticus* isolates. All the *C. sakazakii* exhibited $bla_{\text{CSA-1}}$ and the *C. malonaticus* strain showed $bla_{\text{CMA-1}}$, conferring resistance to cephalosporins. Both *C. sakazakii* ST1 strains harbored the *mcr-9.1* gene, conferring resistance to colistin. All the *C. sakazakii* and *C. malonaticus* strains exhibited the same efflux genes (*adeF*, *H-NS*, *msbA*, *marA*, *kpnF*, *kpnE*, *emrR*, *emrB*, *rsmA*, and *CRP*), antibiotic inactivation gene (*ampH*), and four antibiotic target alteration genes (*pBP3*, *glpT*, *eF-Tu*, and *marR*; **Table 3**).

In *Salmonella*, 30 genes that encode for aminoglycoside and cephalosporin resistance were identified, including *aac(6′)-Iaa*, *ampH*, *ampC1*, and *marA*. All the strains exhibited the same efflux genes (*acrAB*, *golS*, *mdsA*, *adeF*, *marA*, *kpnF*, *kpnE*, *emrRB*, *rsmA*, *baeR*, *H-NS*, *sdiA*, *mdfA*, *mdtK*, and *kdpE*), three antibiotic resistance genes (*aac(6′)-Iaa*, β-*lactamase ampH*, and *ampC1*), and nine antibiotic target alteration genes (*bacA*, *pmrF*, *uhpT*, *glpT*, *PBP3*, *EF-Tu*, *soxS*, *soxR*, and *marR*; **Table 4**).

Cronobacter sakazakii isolates showed 32 virulence genes that were detected by WGS and clustered as flagellar proteins, outer membrane proteins, chemotaxis, hemolysins, invasion, plasminogen activator (*cpa*), colonization, transcriptional regulator, survival in macrophages, utilization of sialic acid (*nanA,K,T*), desiccation tolerance (*cheB*, *wzzB*), and toxin-antitoxin genes (*fic*, *relB*). In the *C. malonaticus* strain, the same virulence genes were detected as found in *C. sakazakii*, except for the *cpa* and *nanAK,T* genes (**Table 5**).

In the *Salmonella* strains, 120 virulence genes and eight pathogenicity islands were detected. The virulence genes clustered as adherence, magnesium uptake, resistance to antimicrobial peptides, secretion system, stress protein, toxin, resistance to complement killing. The *shdA* gene associated with persistence of the bacteria in the intestine was only present in the 510535-21

[4]http://www.genomicepidemiology.org

TABLE 1 | Identification of *Cronobacter* spp. and *Salmonella* spp. strains isolated from powdered infant formula by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) and whole-genome sequencing (WGS).

Sample ID (MLST database)	Country	MALDI-TOF MS	WGS	ST	CC	Serotype
510197-19 (*3409)	Chile	*C. sakazakii*	*C. sakazakii*	1	1	O-1
510199-19 (*3410)	Chile	*C. sakazakii*	*C. sakazakii*	1	1	O-1
510290-19 (*3411)	Chile	*C. sakazakii*	*C. sakazakii*	1	1	O-1
510556-19 (*3412)	Chile	*C. sakazakii*	*C. sakazakii*	31	31	O-2
510557-19 (*3413)	Chile	*C. malonaticus*	*C. malonaticus*	60	60	O-1
510535-21 (**RD389119)	Mexico	*S.* Typhimurium	*S.* Typhimurium	19	19	O-4:-
510536-21 (**RD389120)	Mexico	*S.* Typhimurium	*S.* Typhimurium	19	19	O-4:-:-
510537-21 (**RD389121)	Mexico	*S.* Typhimurium	*S.* Typhimurium	19	19	O-4:i:1,2
510538-21 (**RD389122)	Mexico	*S.* Typhimurium	*S.* Typhimurium	19	19	O:4:i:1,2
510539-21 (**RD389123)	Mexico	*S.* Typhimurium	*S.* Typhimurium	19	19	*O-4:i:-
510540-21 (**RD389124)	Mexico	*S.* Typhimurium	*S.* Typhimurium	19	19	*O-4:i:-

ST, sequence type and CC, clonal complex.
**MLST database ID.*
***cgMLST database ID. *Potential monophasic variant of S. Typhimurium.*

strain. The *gogB*, *pipB*, *ssaCTU*, *ssel/srfH*, *sseL*, *sspH2*, *shdA*, *sopD2*, and *sseK1,2* genes were not found in the 510536-21, 510537-21, 510538-21, 510539-21, and 510540-21 strains associated with the secretion system, effector proteins, adherence, and host survival (**Table 6**).

Detection of Plasmids and Mobile Genetics Elements

The Col(pHHAD28) plasmids and seven MGEs (IS903, IS26, ISEsa2, IS5075, ISEsa1, ISPpu12, and IS102) were detected in only three *C. sakazakii* strains. The IncFIB(pCTU1) plasmid and one MGE (IS481) were detected in the *C. malonaticus* strain.

All the *Salmonella* strains exhibited the IncFII(S) plasmids and five similar MGEs (ISSen7, ISSty2, ISEcI10, MITEEcl, and ISSen1; **Table 7**).

CRISPR-Cas Loci Profiling

Genome analysis showed CRISPR-Cas systems in all of the genomes. In *Cronobacter* spp., 80% ($n = 4/5$) of the genomes revealed the presence of up to three arrays, which were characterized by the same repeated sequences but at different positions in the genome (**Table 8**). In the case of *Salmonella* spp. isolates, 100% ($n = 6/6$) of the genomes showed two arrays associated with the CRISPR-Cas systems in different positions but characterized by the same number of repeated sequences and spacers, with up to 28 repeated sequences and 27 spacers.

Using the CRISPRmap program, the repeated sequences and associated *cas* genes allowed us to determine that the CRISPR systems identified in *Cronobacter* spp. genomes belonged to type I-E and I-F in which 60% ($n = 3/5$) were characterized by the presence of both types of CRISPR-Cas systems. However, the opposite was observed in the *Salmonella* genomes, which were associated with the presence of type I-E systems. As for the associated *cas* genes, type I-E systems of the *Salmonella* genomes showed a larger number of genes associated with the CRISPR arrays (**Figure 3**).

The analysis of the spacers with CRISPRminer revealed bacteriophage sequences with bacteriophages that are characteristic of *Salmonella* and *Enterobacteriaceae* in three *Cronobacter* spp. genomes (**Supplementary Table 1**). However, when searching for the PAM sequences, the spacers of the CRISPR arrays identified in the *Cronobacter* genomes were also characterized by phages associated with the *Klebsiella*, *Streptococcus*, and *Acinetobacter* genera. The spacers of the *Salmonella* spp. genomes were associated with the *crAss* and *Acanthamoeba polyphaga moumouvirus* phages and bacteriophages that are characteristic of *Salmonella* and *E. coli* (**Supplementary Table 2**). These results are correlated with the phages that were identified in the study genomes (**Supplementary Table 3**).

DISCUSSION

WGS enables the molecular typing of bacterial strains on a routine basis for use in epidemiology and real-time infection control. Its usefulness has also been demonstrated in identifying antimicrobial resistance markers, virulence, and the genetic prediction of antibiotic susceptibility test results (Lepuschitz et al., 2020).

In the present study, we identified and characterized *C. sakazakii* ST1 and ST31, *C. malonaticus* ST60, and *S.* Typhimurium ST19 strains from commercially available PIF. The three strains of the pathovar *C. sakazakii* ST1 were isolated from different batches of PIF which came same manufacturer. This may reflect either the widespread occurrence of this sequence type in PIF manufacturing environment or

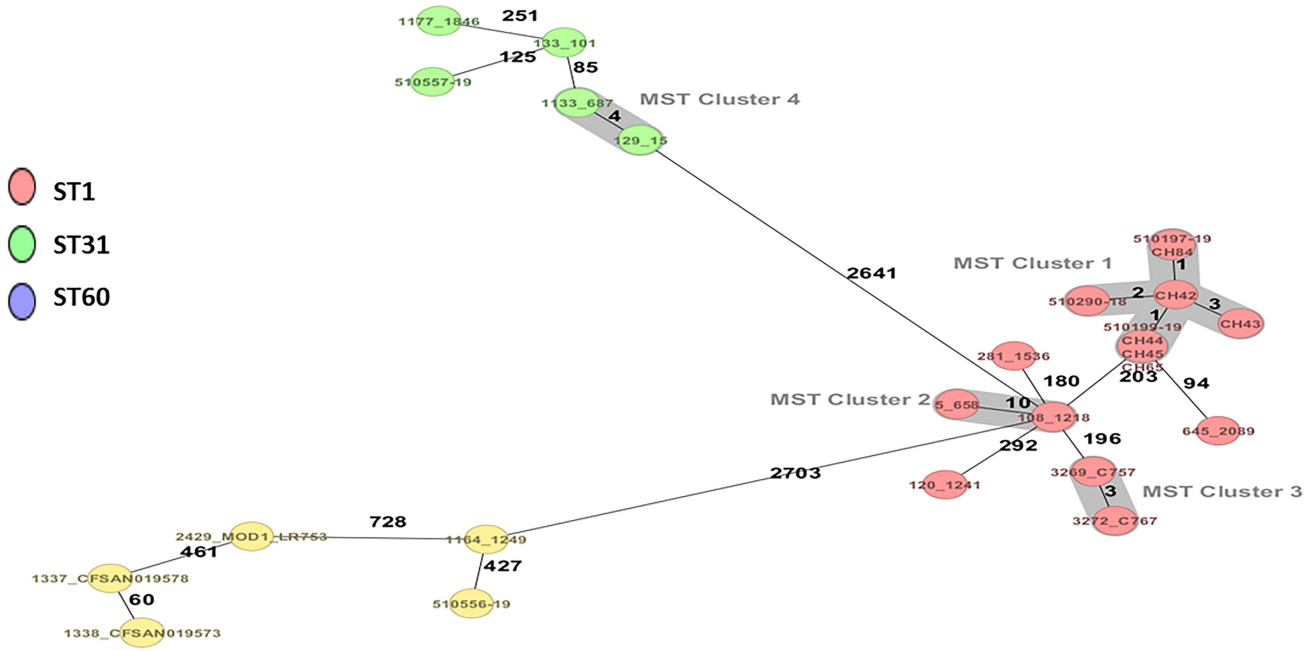


FIGURE 1 | Minimum spanning tree (MST) of five strains of *Cronobacter sakazakii* and one of *Cronobacter malonaticus* from powdered infant formula manufactured in Chile, complemented with strains of *C. sakazakii* and *C. malonaticus* ST1, ST31, and ST60 of clinical and food origin. Calculation of the MST was based on the defined cgMLST scheme comprising 3,678 target genes for *C. sakazakii* and *C. malonaticus*. Isolates are represented as colored circles according to the classical MLST. Black numbers according to the allelic differences between isolates. Isolates with closely related genotypes are marked as Cluster.

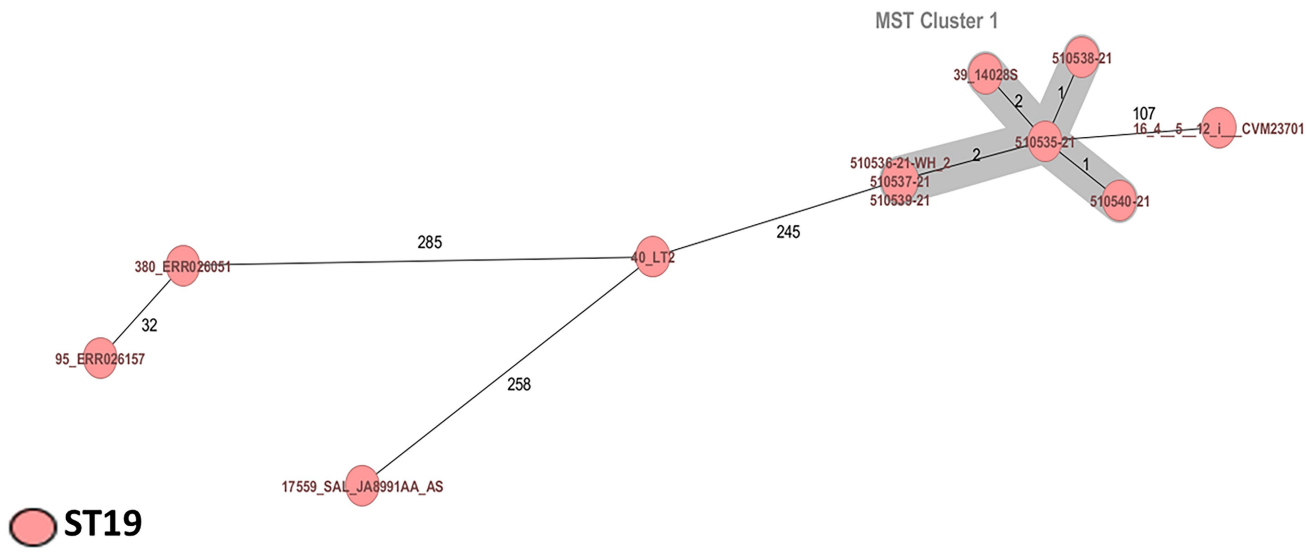


FIGURE 2 | Minimum spanning tree of six *Salmonella* Typhimurium strains from powdered infant formula manufactured in Mexico and supplemented with other *S.* Typhimurium ST19 strains of clinical and food origin. Calculation of the MST was based on the defined cgMLST scheme comprising 3,002 target genes for *Salmonella*. Isolates are represented as colored circles according to the classical MLST. Black numbers according to the allelic differences between isolates. Isolates with closely related genotypes are marked as Cluster.

that a common ingredient was contaminated with the same strain (Sonbol et al., 2013). *Cronobacter sakazakii* ST1 has frequently been found in PIF commercialized in different countries, in the PIF processing environment, and in invasive clinical cases such as fatal meningitis and septicemia (Joseph and Forsythe, 2012; Sonbol et al., 2013; Fei et al., 2015; Csorba et al., 2021; Holý et al., 2021; Parra-Flores et al., 2021b). A survey of *Cronobacter* in the Americas showed the majority of reports isolations were from North America (57.4%, $n = 465$) and Brazil (42.6%, $n = 465$). There were a total of 75 sequence

TABLE 2 | Antibiotic resistance profile of *Cronobacter* spp. and *Salmonella* spp. strains.

	Strain ID	Species	Antibiotics										
			AM (10 μg)	AK (30 μg)	CL (30 μg)	CRO (30 μg)	CTX (30 μg)	FEP (30 μg)	GE (10 μg)	KF (30 μg)	LEV (5 μg)	NET (30 μg)	OX (1 μg)	SXT (25 μg)
Cronobacter spp.	510197-19	*C. sakazakii*	**R**	S	S	S	S	S	S	**R**	S	S	S	S
	510199-19	*C. sakazakii*	S	S	S	S	S	S	S	**R**	S	S	S	S
	510290-19	*C. sakazakii*	S	**S**	S	S	S	S	S	**R**	S	S	S	S
	510556-19	*C. sakazakii*	**I**	**S**	S	S	S	S	S	**R**	S	S	S	S
	510557-19	*C. malonaticus*	**R**	S	S	S	S	S	S	**R**	S	S	S	S
Salmonella spp.	510535-21	S. Typhimurium	**I**	S	S	S	S	S	**I**	S	S	S	**R**	S
	510536-21	S. Typhimurium	**R**	**S**	S	S	**I**	S	**R**	S	S	S	**R**	S
	510537-21	S. Typhimurium	**R**	S	S	S	S	S	S	**R**	S	S	**R**	S
	510538-21	S. Typhimurium	**R**	S	S	S	S	S	S	**R**	S	S	**R**	S
	510539-21	S. Typhimurium	**R**	S	S	S	S	S	S	**R**	S	S	**R**	S
	510540-21	S. Typhimurium	**R**	S	S	S	S	S	S	**R**	S	S	**R**	S

AM, ampicillin; AK, amikacin; CL, chloramphenicol; CRO, ceftriaxone; CTX, cefotaxime; FEP, cefepime; GE, gentamicin; KF, cephalothin; LEV, levofloxacin; NET, netilmicin; OX, oxacillin; SXT, trimethoprim/sulfamethoxazole; **R**, *resistance; and* **S**, *susceptibility;* **I**, *intermediate. The values in parentheses in bold correspond to the concentrations of antibiotics.*

types, with the most frequently reported being the *C. sakazakii* pathovars ST4 (CC4) and ST1 (CC1; Costa et al., 2021).

The cgMLST scheme analysis clustered the three *C. sakazakii* ST1 strains closely to the *C. sakazakii* ST1 strains isolated from the food alert in Chile on 2017 with one to three alleles differences (Parra-Flores et al., 2018b). When analyzing the isolated strains of human infections in a European multicenter study using cgMLST, eight *C. sakazakii* ST1 isolates were found among all the *C. sakazakii* strains. Of these eight ST1 isolates, two strains were analyzed from an outbreak affecting two newborns suffering from necrotizing enterocolitis in Austria in 2009, which differed by only one allele. In addition, three ST1 isolates from Austria and one ST1 from Denmark differed by 203 alleles with the ATCC BAA-894 strain, which was isolated from PIF associated with a fatal case of an infant in the United States in 2001. *Cronobacter sakazakii* ST31 has also been isolated from clinical cases with fatal outcomes; however, it has been less prevalent in PIF and the environment (Sonbol et al., 2013; Fei et al., 2015; Ogrodzki and Forsythe, 2015; Lepuschitz et al., 2019). In this context, *C. sakazakii* ST1, ST8, and especially ST4 are the STs with the highest risk of causing disease in infants (Joseph and Forsythe, 2011; Forsythe, 2018; Lachowska et al., 2021). *Cronobacter malonaticus* ST60 has also been isolated in powdered milk, food, and the environment, but with less significant clinical cases than *C. malonaticus* ST7 and *C. sakazakii* ST1 (Forsythe, 2018; Costa et al., 2021).

Salmonella enterica is the most frequently identified cause of food poisoning in the European Union; serotype Typhimurium ST19 is most often associated with disease and death (Carroll et al., 2017; de Frutos et al., 2018) and commonly identified in human clinical samples, animals, food, and the environment (Panzenhagen et al., 2018; Monte et al., 2020). The cgMLST scheme analysis revealed a cluster of six *S.* Typhimurium ST19 with one to two allele differences. In this context, the present study is the first to identify *S.* Typhimurium ST19 in PIF, whereas recent reports of *Salmonella* outbreaks in PIF have involved *S.* Agona (Brouard et al., 2007; Jourdan-da Silva et al., 2018). We also presumptively identified two *Salmonella* monophasic variant strains. Monophasic *Salmonella* has been identified in different human cases in the United States, Spain, Brazil, and Thailand and characterization of these strains revealed resistance to multiple antibiotics (Mossong et al., 2007).

All five *C. sakazakii* strains analyzed in the present study were resistant to cephalothin and only two strains to ampicillin. The resistance of *C. sakazakii* to cephalothin and ampicillin has been reported in several studies, also suggesting an almost intrinsic resistance of the *Cronobacter* spp. genus to cephalothin (Kim et al., 2008; Molloy et al., 2009; Flores et al., 2011; Chon et al., 2012). Parra-Flores et al. (2020) found that 100% of the *C. sakazakii* strains isolated in powdered infant formula distributed in Latin America were resistant to cefotaxime and ampicillin, 60% to cefepime, 40% to amikacin, and 20% to cephalothin. Furthermore, one of the strains was resistant to six of the 12 evaluated antibiotics and another strain was resistant to five antibiotics. Multiple drug resistance (MDR) is a cause for concern; in a case of neonatal meningitis caused

TABLE 3 | Antibiotic-resistant genes of *Cronobacter* spp. strains identified by Comprehensive Antibiotic Resistance Database (CARD).

Best hits antibiotic resistance ontology (ARO)	Drug class	Resistance mechanism	510197-19 (ST1)	510199-19 (ST1)	510290-18 (ST1)	510556-19 (ST31)	510557-19 (ST60)
MCR-9.1	Peptide antibiotic	Antibiotic target alteration	+	+	+	–	–
CSA-1	Cephalosporin	Antibiotic inactivation	+	+	+	+	–
CMA-1	Cephalosporin	Antibiotic inactivation	–	–	–	–	+
pBP3	Cephalosporin, cephamycin, and penam	Antibiotic target alteration	+	+	+	+	+
glpT	Fosfomycin	Antibiotic target alteration	+	+	+	+	+
eF-Tu	Elfamycin antibiotic	Antibiotic target alteration	+	+	+	+	+
marR	Fluoroquinolone antibiotic, triclosan, rifamycin antibiotic, penam, phenicol antibiotic, glycylcycline, tetracycline antibiotic, and cephalosporin	Antibiotic target alteration	+	+	+	+	+
adeF	Fluoroquinolone antibiotic and tetracycline antibiotic	Antibiotic efflux	+	+	+	+	+
H-NS	Macrolide antibiotic, fluoroquinolone antibiotic, cephalosporin, cephamycin, penam, and tetracycline antibiotic	Antibiotic efflux	+	+	+	+	+
msbA	Nitroimidazole antibiotic	Antibiotic efflux	+	+	+	+	+
marA	Fluoroquinolone antibiotic, monobactam, carbapenem, cephalosporin, glycylcycline, cephamycin, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, triclosan, and penem	Antibiotic efflux	+	+	+	+	+
pnF	Macrolide antibiotic, aminoglycoside antibiotic, cephalosporin, tetracycline antibiotic, peptide antibiotic, and rifamycin antibiotic	Antibiotic efflux	+	+	+	+	+
kpnE	Macrolide antibiotic, aminoglycoside antibiotic, cephalosporin, tetracycline antibiotic, peptide antibiotic, and rifamycin antibiotic	Antibiotic efflux	+	+	+	+	+
emrR	Fluoroquinolone antibiotic	Antibiotic efflux	+	+	+	+	+
emrB	Fluoroquinolone antibiotic	Antibiotic efflux	+	+	+	+	+
rsmA	Fluoroquinolone antibiotic, diaminopyrimidine antibiotic, and phenicol antibiotic	Antibiotic efflux	+	+	+	+	+
cRP	Fluoroquinolone antibiotic, macrolide antibiotic, and penam	Antibiotic efflux	+	+	+	+	+
kpnH	Macrolide antibiotic, fluoroquinolone antibiotic, aminoglycoside antibiotic, carbapenem, cephalosporin, penam, peptide antibiotic, and penem	Antibiotic efflux	–	–	–	–	–
ampH ampC-type β-lactamase	Cephalosporin and penam	Antibiotic inactivation	+	+	+	+	+

+, presence and –, absence.

TABLE 4 | Antibiotic-resistant genes of *S.* Typhimurium strains identified by CARD.

Best hits antibiotic resistance ontology (ARO)	Drug class	Resistance mechanism	510535-21	510536-21	510537-21	510538-21	510539-21	510540-21
AAC(6')-Iaa	Aminoglycoside antibiotic	Antibiotic inactivation	+	+	+	+	+	+
ampH β-*lactamase*	Cephalosporin and penam	Antibiotic inactivation	+	+	+	+	+	+
ampC1 β-*lactamase*	Cephalosporin and penam	Antibiotic inactivation	+	+	+	+	+	+
bacA	Peptide antibiotic	Antibiotic target alteration	+	+	+	+	+	+
pmrF	Peptide antibiotic	Antibiotic target alteration	+	+	+	+	+	+
uhpT	Fosfomycin	Antibiotic target alteration	+	+	+	+	+	+
glpT	Fosfomycin	Antibiotic target alteration	+	+	+	+	+	+
PBP3	Cephalosporin, cephamycin, and penam	Antibiotic target alteration	+	+	+	+	+	+
EF-Tu	Elfamycin antibiotic	Antibiotic target alteration, antibiotic efflux	+	+	+	+	+	+
soxS	Fluoroquinolone antibiotic, monobactam, carbapenem, cephalosporin, glycylcycline, cephamycin, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, triclosan, and penem	Antibiotic target alteration, antibiotic efflux, reduced permeability to antibiotic	+	+	+	+	+	+
soxR	Fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, and triclosan	Antibiotic target alteration, antibiotic efflux, and reduced permeability to antibiotic	+	+	+	+	+	+
marR	Fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, and triclosan	Antibiotic target alteration, antibiotic efflux	+	+	+	+	+	+
acrA	Fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, and triclosan	Antibiotic efflux	+	+	+	+	+	+
AcrB	Fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, and triclosan	Antibiotic efflux	+	+	+	+	+	+
golS	Monobactam, carbapenem, cephalosporin, cephamycin, penam, phenicol antibiotic, and penem	Antibiotic efflux	+	+	+	+	+	+
MdsA	Monobactam, carbapenem, cephalosporin, cephamycin, penam, phenicol antibiotic, and penem	Antibiotic efflux	+	+	+	+	+	+
adeF	Fluoroquinolone antibiotic, and tetracycline antibiotic	Antibiotic efflux	+	+	+	+	+	+
marA	Fluoroquinolone antibiotic, monobactam, carbapenem, cephalosporin, glycylcycline, cephamycin, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, triclosan, and penem	Antibiotic efflux, reduced permeability to antibiotic	+	+	+	+	+	+
kpnE	Macrolide antibiotic, aminoglycoside antibiotic, cephalosporin, tetracycline antibiotic, peptide antibiotic, and rifamycin antibiotic	Antibiotic efflux	+	+	+	+	+	+
kpnF	Macrolide antibiotic, aminoglycoside antibiotic, cephalosporin, tetracycline antibiotic, peptide antibiotic, and rifamycin antibiotic	Antibiotic efflux	+	+	+	+	+	+
emrR	Fluoroquinolone antibiotic	Antibiotic efflux	+	–	+	+	+	+
emrB	Fluoroquinolone antibiotic	Antibiotic efflux	+	+	+	+	+	+
rsmA	Fluoroquinolone antibiotic, diaminopyrimidine antibiotic, and phenicol antibiotic	Antibiotic efflux	+	+	+	+	+	+
baeR	Aminoglycoside antibiotic and aminocoumarin antibiotic	Antibiotic efflux	+	+	+	+	+	+
H-NS	Macrolide antibiotic, fluoroquinolone antibiotic, cephalosporin, cephamycin, penam, and tetracycline antibiotic	Antibiotic efflux	+	+	+	+	+	+
sdiA	Fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rifamycin antibiotic, phenicol antibiotic, and triclosan	Antibiotic efflux	+	+	+	+	+	+
mdfA	Fluoroquinolone antibiotic, macrolide antibiotic, and penam	Antibiotic efflux	+	+	+	+	+	+
MdtK	Fluoroquinolone antibiotic	Antibiotic efflux	+	+	+	+	+	+
CRP	Macrolide antibiotic, fluoroquinolone antibiotic, and penam	Antibiotic efflux	–	+	+	+	+	+
kdpE	Aminoglycoside antibiotic	Antibiotic efflux	+	+	+	+	+	+

+, presence and –, absence.

TABLE 5 | Putative virulence and distribution of other genes in seven strains of *Cronobacter* spp. by WGS.

Virulence gene	Function	510197-19 (ST1)	510199-19 (ST1)	510290-18 (ST1)	510556-19 (ST31)	510557-19 (ST60)	*C. sakazakii* BAA-894 (ST1)	*C. malonaticus* LMG23826T (ST7)	ES15 control (ST125)
flgB	Motility	+	+	+	+	+	+	+	+
flgK	Flagellar hook-associated protein 1	+	+	+	+	+	+	+	+
flgL	Flagellar hook-associated protein 3	+	+	+	+	+	+	+	–
flgM	Negative regulator of flagellin synthesis	+	+	+	+	+	+	+	+
flgN	Flagellar synthesis FlgN protein	+	+	+	+	+	+	+	+
flhD	Flagellar hook-associated protein 2	+	+	+	+	+	+	+	+
fliA	Flagellar operon FliA	+	+	+	+	+	+	+	+
fliC	Flagellin	+	+	+	+	+	–	+	–
fliD	Flagellar hook-associated protein 2	+	+	+	+	+	+	+	+
fliR	Flagellar biosynthetic FliR protein	+	+	+	+	+	+	+	+
fliT	Flagellar FliT protein	+	+	+	+	+	+	+	+
fliZ	FliZ protein	+	+	+	+	+	+	+	+
lolA	Outer membrane lipoprotein carrier protein	+	+	+	+	+	+	+	+
motB	Chemotaxis MotA protein	+	+	+	+	+	+	+	+
sdiA	LuxR family transcriptional regulator	+	+	+	+	+	+	+	+
slyB	Outer membrane lipoprotein SlyB	+	+	+	+	+	+	+	+
tolC	Outer membrane channel protein	+	+	+	+	+	+	+	+
msbA	Survival in macrophage	+	+	+	+	+	+	–	+
mviN	Protective immunity and colonization	+	+	+	+	+	+	+	+
cpa	Plasminogen activator	+	+	+	+	–	+	–	–
hly	Hemolysin	+	+	+	+	+	+	–	+
ompA	Adhesion cell, biofilm formation	+	+	+	+	+	+	+	+
ompX	Adhesion cell	+	+	+	+	+	+	+	+
cheR	Chemotaxis protein methyltransferase	+	+	–	+	+	–	+	–
cheY	Response regulator of chemotaxis family	+	+	+	+	+	+	+	+
cheB	Desiccation tolerance	+	+	+	+	+	+	+	+
lpxA	Epithelial cell invasion and lipid A production	+	+	+	+	+	+	+	+
nanA,K,T	Exogenous sialic acid utilization	+	+	+	+	–	+	–	+
ibpA	Small heat shock protein	+	+	+	+	+	+	+	+
wzzB	Desiccation tolerance	+	+	+	+	+	+	+	+
fic	Cell filamentation protein	+	+	+	+	+	+	+	+
relB	RelE antitoxin	+	+	+	+	+	+	–	+

+, presence and –, absence.

TABLE 6 | Putative virulence and distribution of other genes in six strains of *S.* Typhimurium by WGS.

	Genes				510535	510536	510537	510538	510539	510540
Virulence	*avrA*	*bcfA*	*bcfB*	*bcfC*	+	+	+	+	+	+
	bcfD	*bcfE*	*bcfF*	*bcfG*						
	csgA	*csgB*	*csgC*	*csgD*						
	csgE	*csgF*	*csgG*	*fimC*						
	fimD	*fimF*	*fimH*	*fimI*						
	grvA	*invA*	*invB*	*invC*						
	invE	*invF*	*invG*	*invH*						
	invI	*invJ*	*lpfA*	*lpfB*						
	lpfC	*lpfD*	*lpfE*	*mgtB*						
	mgtC	*mig-14*	*misL*	*orgA*						
	orgB	*orgC*	*pefA*	*pefB*						
	pefC	*pefD*	*pipB2*	*prgH*						
	prgI	*prgJ*	*prgK*	*ratB*						
	rck	*sicA*	*sicP*	*sifA*						
	sifB	*sinH*	*sipA/sspA*	*sipB/sspB*						
	sipC/sspC	*sipD*	*slrP*	*sodCI*						
	sopA	*sopB/sigD*	*sopD*	*sopE2*						
	spaO	*spaP*	*spaQ*	*spaR*						
	spaS	*spiC/ssaB*	*sptP*	*spvB*						
	spvC	*spvR*	*ssaD*	*ssaE*						
	ssaG	*ssaH*	*ssaI*	*ssaJ*						
	ssaK	*ssaL*	*ssaM*	*ssaN*						
	ssaO	*ssaP*	*ssaQ*	*ssaR*						
	ssaS	*ssaV*	*sscA*	*sscB*						
	sseA	*sseB*	*sseC*	*sseD*						
	sseE	*sseF*	*sseG*	*sseJ*						
	sseL	*sspH1*	*steA*	*steB*						
	steC									
	gogB	*pipB*	*ssaCTU*	*ssel/srfH*	+	−	+	+	+	+
	sseK2	*sseL*	*sspH2*							
	shdA				+	−	−	+	−	−
	sopD2				+	−	−	+	+	+
	sseK1				+	−	+	+	−	+
	TTSS(Type III secretion system)				+	+	+	+	+	+
Pathogenicity islands	*SPI-1*				+	+	+	+	+	+
	SPI-2				+	+	+	+	+	+
	SPI-3				+	+	+	+	+	+
	SPI-4				+	+	−	+	+	−
	SPI-5				+	+	+	+	+	+
	SPI-9				+	+	+	+	+	+
	SPI-13				+	+	+	+	+	+
	SPI-14				+	+	+	+	+	+

+, presence and –, absence.

by *C. sakazakii* in China, the isolated strains were resistant to eight antibiotics (Zeng et al., 2018).

For the antibiotic resistance genes, all *C. sakazakii* strains and the *C. malonaticus* strain exhibited the same efflux, antibiotic inactivation, and antibiotic target alteration genes that confer antibiotic resistance to β-lactams, fluoroquinolones, aminoglycosides, imidazoles, and disinfectants such as triclosan. The *marA* gene, whose transcription function regulates multidrug efflux, modulates membrane permeability and activates the transcription of the AcrAB-TolC efflux pump that plays an important role in antibiotic resistance (Wang X. et al., 2021). Several authors have detected the *msbA*, *emrR*, *H-NS*, *emrB*, *marA*, *CRP*, *PBP3*, *H-NS*, and *msrB* genes that are associated with antibiotic resistance by using efflux pumps, regulatory systems, and antibiotic target protection genes (Aly et al., 2019; Lepuschitz et al., 2019; Parra-Flores et al., 2021b). All strains exhibited the *glpT* gene that encodes fosfomycin resistance. This is relevant because fosfomycin is considered effective in patients with MDR bacterial infections (Falagas et al., 2019). We also identified the *mcr-9.1* gene that confers resistance to colistin and the bla_{CSA} and bla_{CMA} genes that confer resistance to cephalothin. The *mcr-9.1* gene can generate resistance to colistin in several enteropathogens and can silently circulate undetected unless induced by colistin (Carroll et al., 2019; Kieffer et al., 2019). The presence of mobile genes resistant to colistin (*mcr*) is a worldwide public concern because colistin is considered as a last resort to treat infections caused by multidrug-resistant *Enterobacteriaceae* (Borowiak et al., 2020). The bla_{CSA} and bla_{CMA} genes were first described in 2014 and are associated with β-lactamase class C resistance; they are not inducible and are regarded as cephalosporinases (Müller et al., 2014). Jang et al. (2020) found variants of class C *bla*

TABLE 7 | Plasmids and mobile genetic elements of *Cronobacter* spp. and *S.* Typhimurium.

Bacteria	ID strain	Plasmid	Plasmids accession number	Mobile genetic elements
Cronobacter sakazakii	510197-19	Col(pHDA28)	KU674895	IS903, IS26, ISEsa2, IS5075, ISEsa1, ISPpu12, IS102
	510199-19	Col(pHDA28)	KU674895	IS903, IS26. ISEsa2, IS5075, ISEsa1, ISPpu12, IS102
	510290-18	Col(pHDA28)	KU674895	ISEsa2, IS5075, ISEsa1, ISPpu12, IS102
	510556	----		ISEsa1
Cronobacter malonaticus	510557-19	IncFIB(pCTU1)	FN543094	IS481
Salmonella Typhimurium	510535-21	IncFII(S)	FN543094	ISSen7, ISSty2,ISEcl10, MITEEcl, ISSen1
	510536-21	IncFIB(S)	FN432031	ISSen7, ISSen1, MITEEc1, ISEcl10
		INCFII(S)	CP000858	
	510537-21	IncFII(S)	FN543094	ISSen7, MITEEcl, ISSen1, ISEcl10, ISSty2
	510538-21	IncFII(S)	FN543094	ISSen7, MITEEcl, ISEcl10, ISSen1
	510539-21	IncFII(S)	FN543094	ISSen7, MITEEcl, ISSen1, ISSty2, ISEcl10
	510540-21	IncFII(S)	FN543094	ISSen7, MITEEcl, ISSen1, ISEcl10

resistance genes identified as *CSA-2* or *CSA-1* and *CMA* in all the analyzed strains. Holý et al. (2021) encountered bla_{CSA} genes in all the *C. sakazakii* strains isolated in powdered milk produced in the Czech Republic between 2010 and 2014.

In the present study, all six *S.* Typhimurium ST19 strains were resistant to oxacillin, five to ampicillin, four to cephalothin, and one to gentamicin. *S.* Typhimurium ST19 has shown extensive resistance to a variety of critically important antimicrobials (Monte et al., 2020). Jain et al. (2018) reported that all the evaluated *S.* Typhimurium ST19 strains were resistant to 7 of the 15 tested antibiotics and encountered only strains that were susceptible to ampicillin and gentamicin; this contrasts with our study in which strains were resistant to these two antibiotics.

The *Salmonella* strains carried many genes such as *aac(6′)-Iaa* that encodes resistance to aminoglycosides (e.g., gentamicin), which have been found in multidrug-resistant *S.* Typhimurium strains and caused a foodborne outbreak at a banquet in China in 2017 (Xiang et al., 2020). Wei et al. (2019) reported the presence of β-*lactamase ampC* genes the same as in our study; in addition to finding β-*lactamase ampC-1*, we found that it co-harbored with β-*lactamase ampH*. The β-*lactamase ampC* is chromosomally encoded in several gram-negative bacteria, including *Enterobacter* spp., *Citrobacter freundii*, or *Serratia marcescens*. Furthermore, plasmid-encoded *ampC* genes can be horizontally transferred to other *Enterobacteriaceae* without the presence of chromosomally encoded *ampC*, such as *Klebsiella* and *Salmonella*, resulting in a highly effective and dynamic dissemination mechanism (Ingram et al., 2011). We also found the *uhpT* and *glpT* genes that encode fosfomycin resistance. The UhpT and GlpT transporters facilitate fosfomycin incorporation into bacterial cells (Silver, 2017). The reduced expression or introduction of *glpT* or *uhpT* mutations and the efflux pump can decrease fosfomycin uptake and thus lower antibiotic susceptibility (Garallah and Al-Jubori, 2020). Several studies have now shown that the overuse of antibiotics in the food chain and the presence of several antibiotic resistance operons promote multiresistance to these drugs (Ferri et al., 2017).

Among the 30 virulence genes detected in *C. sakazakii*, we identified *C. sakazakii ompA* and *ompX*, which encode proteins for basolateral adherence in cell lines and a possible involvement in the blood–brain barrier penetration (Mange et al., 2006; Kim et al., 2010). The Cpa protein is related to *C. sakazakii* serum resistance and invasion. Recent studies have suggested that the *cpa* locus could be considered specific for *C. sakazakii* and *C. universalis* but is absent in *C. malonaticus* (Franco et al., 2011). Jang et al. (2020) noted that highly virulent *C. sakazakii* ST8 clinical strains that carry the pESA3 plasmid do not possess the *cpa* gene; this indicates that the disease can be associated with other virulence factors. Hemolysins (Hly) are outer membrane proteins occurring in various pathogens belonging to the *Enterobacteriaceae* family that have hemolytic ability, such as *E. coli*, *Klebsiella*, and some gram-positive pathogens (Mare et al., 2020; Mazzantini et al., 2020). Other important genes of the *C. sakazakii* strains isolated from PIF are *nanAKT* that encode for exogenous sialic acid utilization. Sialic acid is naturally present in breast milk and supplemented in PIF because of its association with brain development (Forsythe, 2018). Sialic acid also regulates the expression of enzymes, such as sialidase and adhesins, or inhibits transcription factors of the *fimB* gene that mediates epithelial cell adherence and invasion. *Escherichia coli* K1 use this compound to modify their cell surface (Severi et al., 2007; Sohanpal et al., 2007).

In our study, we found the *fic* toxin encoding gene and the *relB* gene that encodes for the *relE* antitoxin. The *fic* gene encodes for the toxin component of the toxin-antitoxin bicistronic operon. The toxin-antitoxin (TA) systems are small genetic elements found in plasmids, phage genomes, and in the chromosomes of different bacterial species. The

TABLE 8 | CRISPR-Cas systems identified in the *Cronobacter* spp. and *Salmonella* spp. genomes.

Strains	Operon structure type	Position	Maximum number of spacers per strain	Number of CRISPR arrays per strain	Repeat consensus	*cas* genes
510197-*Cronobacter* spp.	Type I-F CAS	77362-76641	12	13	GTTCACTGCCGTACAGGCAGCTTAGAAA	DEDDh,*cas3,cas8e,cse2gr11,cas7,cas5,cas6e,cas1,cas2*
	Type I-E CAS	171199-172847	27	28	CTGTTCCCCGCGCGAGCGGGGATAAACCG	
		199092-200862	29	30	GTGTTCCCCGCGCGAGCGGGGATAAACCG	
510199-*Cronobacter* spp.	Type I-E CAS	482009-482670	11	12	GTTCACTGCCGTACAGGCAGCTTAGAAA	*cas2,cas1,cas6e,cas5,cas7,cse2gr11,cas8e,cas3,*DEDDh
		77461-79109	27	28	CGGTTTATCCCCGCTCGCGCGGGGAA	
		105476-107002	25	26	CGGTTTATCCCCGCTCGCGCGGGGAACAG	
510290-*Cronobacter* spp.	Type I-F CAS	480714-481435	12	13	GTTCACTGCCGTACAGGCAGCTTAGAAA	DEDDh, *csa3,cas3,cas8e,cse2gr11,cas7,DinG,cas6,cas1*
	-Type I-E CAS	173384-175032	27	28	CTGTTCCCCGCGCGAGCGGGGATAAACCG	
		201277-203047	29	30	GTGTTCCCCGCGCGAGCGGGGATAAACCG	
510556-*Cronobacter* spp.	Type I-F CAS	161191-162091	9	10	CTGTTCCCCGCGCGAGCGGGGATAAACCG	*cas2*, *csa3,DinG,cas3,DEDDh,csa3,cas2,cas1,cas6e,cas5,cas7,cse2gr11,cas8e*
	-Type I-E CAS	7728-8277	15	16	GTGTTCCCCGCGCGAGCGGGGATAAACCG	
		199129-200044	15	16	GTTCACTGCCGTACAGGCAGCTTAGAAA	
510557-*Cronobacter* spp.	-Type I-E CAS	6074-7051	17	18	GTGTTCCCCGCGCGAGCGGGGATAAACCG	*cas8e,cas3,cse2gr11,cas7,cas5,cas6e,cas1,cas2,csa3,WYL,csa3,DEDDh*, *DinG*
		165212-166676	25	26	CTGTTCCCCGCGCGAGCGGGGATAAACCG	
510535-*Salmonella* spp.	Type I-E CAS	241410-242857	24	25	GTGTTCCCCGCGCCAGCGGGGATAAACCG	*DinG,c2c9_V-U4,cas3,cas14j,cas3,PD-DExK,cas2,cas1,cas6e,cas5,cas7,cse2gr11,cas8e*
		259017-260604	27	28	GTGTTCCCCGCGCCAGCGGGGATAAACCG	
510536-*Salmonella* spp.	Type I-E CAS	7220-8807	27	28	GTGTTCCCCGCGCCAGCGGGGATAAACCG	*c2c9_V-U4,cas14j,csa3 DinG,DEDDh,cas3,cas2,cas1,cas6e,cas5,cas7,cse2gr11,cas8e,PD-DExK*
		27933-29380	24	25	GTGTTCCCCGCGCCAGCGGGGATAAACCG	
510537-*Salmonella* spp.	Type I-E CAS	5320-6907	27	28	GTGTTCCCCGCGCCAGCGGGGATAAACCG	*c2c9_V-U4,cas14j,DinG,c2c9_V-U4,DEDDh,cas3,DEDDh,DinG,cas3,cas8e,cse2gr11,cas7,cas5,cas6e,cas1,cas2,cas14j*
		23067-24514	24	25	GTGTTCCCCGCGCCAGCGGGGATAAACCG	
510538-*Salmonella* spp.	Type I-E CAS	166775-165328	24	25	GTGTTCCCCGCGCCAGCGGGGATAAACCG	*c2c9_V-U4,cas14j,c2c9_V-U4,DinG,DEDDh,cas3,PD-DExK,cas2,cas1,cas6e,cas5,cas7,cse2gr11,cas8e*
		184522-182935	27	28	GTGTTCCCCGCGCCAGCGGGGATAAACCG	
510539-*Salmonella* spp.	Type I-E CAS	5479-7066	24	25	GTGTTCCCCGCGCCAGCGGGGATAAACCG	*c2c9_V-U4,c2c9_V-U4,csa3,DinG,cas14j,cas3,cas8e,DnG,cse2gr11,cas7,cas5,cas6e,cas1,cas2,DEDDh,cas3*
		8963-10410	27	28	GTGTTCCCCGCGCCAGCGGGGATAAACCG	
510540-*Salmonella* spp.	-Type I-E CAS	161008-162455	24	25	GTGTTCCCCGCGCCAGCGGGGATAAACCG	*c2c9_V-U4,c2c9_V-U4,DinG,DEDDh,cas3,DEDDh,cas14j,cas3,PD-DExK,cas2,cas1,cas6e,cas5,cas7,cse2gr11,cas8e,cas3*
		178615-180202	27	28	GTGTTCCCCGCGCCAGCGGGGATAAACCG	

Characteristic repeated sequences of the identified CRISPR arrays and their position in the genome.

TA genes play a fundamental role in the physiology of bacterial stress response, such as in stabilizing horizontally acquired mobile genetic elements and participating in a persistence phenotype in some species, including *E. coli* and *Salmonella* (Deter et al., 2017; Walling and Butler, 2019). Finkelstein et al. (2019) noted in preliminary studies with *C. sakazakii* isolates that 2 typical toxin genes, *fic* and *hipA*, followed the evolutionary lines of the species and that *C. sakazakii* ST1 strains were the only strains containing the 22 TA homologs.

In the *Salmonella* strains, 110 of the 121 detected virulence genes were similar in all the strains, including the *invA*, *ssaAR ssrAB*, *sipAC*, *sopBE*, *spvBC*, and *rck* genes. Genes such as *invA*, *sipA*, *sopB*, and *sopE* are associated with adherence in

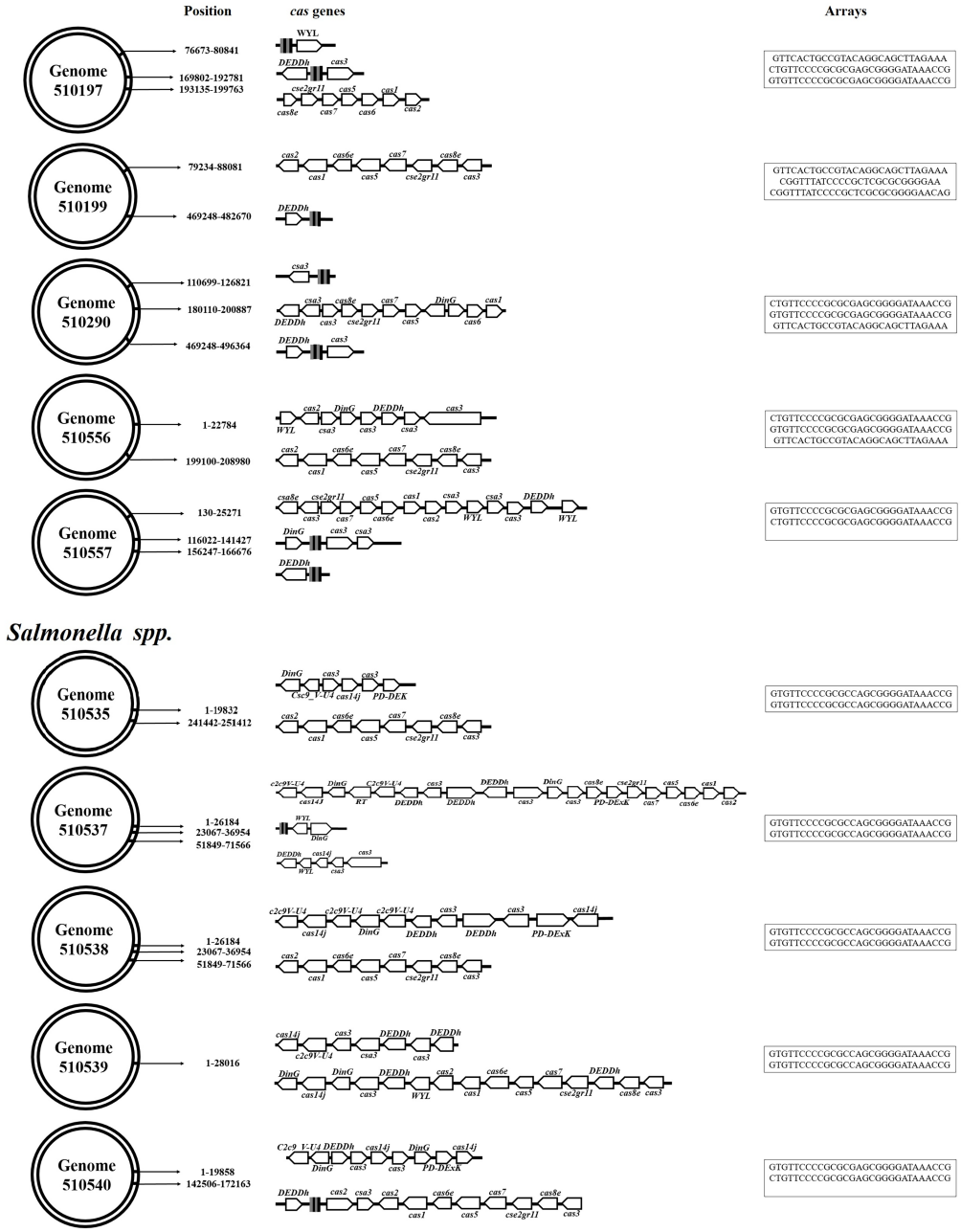


FIGURE 3 | CRISPR-Cas systems identified in *Cronobacter* spp. and *Salmonella* spp. genomes. The identified systems belong to the I-E and I-F CRISPR-Cas systems.

epithelial cells and phagocyte invasion. *Salmonella* pathogenicity islands (SPI) are virulence gene clusters acquired by horizontal transfer that promote virulence in *Salmonella* spp. The SPI-2 gene encodes genes such as *ssaR* and *ssrA* that promote survival and replication within the macrophages and their dissemination in the host (Barilli et al., 2018). The *spvC* gene enables the survival and rapid growth of *Salmonella* in the host (Hu et al., 2021), while the *spvB* gene is known as an intracellular toxin that encodes an enzyme that ADP-ribosylates actin and destabilizes the cytoskeleton of eukaryotic cells (Lesnick et al., 2001).

The *rck* gene is defined as an invasin that generates a colonization niche or as a cyclomodulin with genotoxic activity (Mambu et al., 2020). To date, 24 SPIs have been identified in *Salmonella* of which SPI-1 and SPI-2 are the most important for virulence traits. SPI-1 is involved in the entire *Salmonella* infection process, including invasion, macrophage proliferation, and host responses. Both SPI-1 and SPI-2 encode different

secretion systems, such as T3SS, that transport effector proteins to the host cells (Cheng et al., 2019; Lou et al., 2019).

Only the *C. sakazakii* ST1 strains carried the COL(pHHAD28) plasmids, which are associated with antibiotic resistance genes; these have previously been detected in *C. sakazakii* strains isolated from dairy products in Chile (Ramsamy et al., 2020; Khezri et al., 2021; Parra-Flores et al., 2021b). The *C. malonaticus* strain harbored a plasmid homologous to pESA3 called IncFIB (pCTU1), which encodes iron acquisition virulence genes necessary for pathogen survival, but not an external protease known as plasminogen activator (*cpa*) that enhances pathogen propagation in the host (Franco et al., 2011). Meanwhile, all the *Salmonella* strains showed IncFII(S) plasmid incompatibility, which is associated with several antibiotic resistance genes, such as *aac(6′)-Iaa* and *mcr-9*, and β-lactamase, such as *ampC* (Cha et al., 2020; Richter et al., 2020).

Bacterial CRISPR-Cas systems are considered as mechanisms of acquired immunity because they provide them with the ability to avoid bacteriophage infection and the acquisition of mobile genetic material from plasmids. This immunity is due to the information stored in the matrices of these systems, specifically in the spacer sequences. It was possible to determine that repeated sequences and *cas* genes in *Cronobacter* spp. were associated with two types of systems previously identified as I-E and I-F (Ogrodzki and Forsythe, 2016; Zeng et al., 2017), including cases in which the same strain showed both systems. The opposite was observed in *Salmonella* spp. genomes that were characterized only by the presence of type I-E (Louwen et al., 2014). Although both genera carry systems of the same type, there are differences between the associated *cas* gene operons because the *Salmonella* spp. systems show a larger number of genes. However, these systems are characterized by the presence of the *cas1* and *cas2* genes, which are involved in integrating and processing the information that is integrated into functional CRISPR arrays. *Salmonella enterica* and *E. coli* are closely related and harbor the IE-type CRISPR-Cas system in a similar manner (Makarova et al., 2015). The array sizes of the *Cronobacter* spp. genomes are smaller. However, when showing more than two arrays, they tend to have a larger number of spacer, suggesting that they have been exposed to a larger number of invasive elements than *S.* Typhimurium. Although it was not possible to determine that all the spacers of the systems are similar to bacteriophage sequences, there are cases in which they are specific to these genera. Of particular interest, some of the *S.* Typhimurium spacer were associated with *CrAssphages* and other phages associated with the human intestinal microbiota (Guerin et al., 2018). It is known that *Salmonella* is recognized as an intestinal pathogen; therefore, this is an ecosystem where bacteria can acquire external genetic material and integrate it into their genome by various horizontal transfer mechanisms and by other means such as CRISPR-Cas systems (Louwen et al., 2014). The *cas1* and *cas2* genes are indispensable for integrating and processing the information acquired by the bacteria; in their absence, the system loses the ability to acquire information (Parra-Flores et al., 2021b). The search for prophages in the studied genomes showed that they may have few intact phages. *Cronobacter* spp. tend to show a larger number of intact and incomplete prophages, which are not only characteristic of this genus but also of *Salmonella*. Therefore, our data suggest that these genomes carry the necessary information to prevent these bacteriophages from infecting both *Cronobacter* spp. and *Salmonella* strains. In addition, this information can be useful in the future when using gene therapy as a therapeutic option for infections caused by these pathogens that are difficult to treat (Gordillo Altamirano and Barr, 2019).

In the present study, WGS allowed us to determine multiple virulence and antibiotic resistance genes in bacterial pathogens isolated from PIF intended for consumption by infants aged less than 6 months and distributed throughout Latin America. In our opinion, the identification of *C. sakazakii* and *S.* Typhimurium in PIFs not only violates current health regulations but also endanger the health of infants consuming these products. It is therefore necessary that health authorities conduct more preventive control activities related to these products and carry out campaigns that emphasize the use of rehydration water at 70°C for infant formula. This recommendation by the World Health Organization highlights that the 70°C temperature has a proven effect in significantly decreasing the risk of disease by pathogens in reconstituted PIF (FAO/WHO, 2004, 2006). Control of *Salmonella* and *Cronobacter* during the production of milk powder and PIF is through microbiological sampling according to Codex Alimentarius Commission (2007) of finished product, as well as ingredients and intermediate products (International Standards for Organization (ISO), 2017; Podolak and Black, 2017). The use of zoning production facilities to focus environmental sampling according to risk is used (Cordier, 2008; FDA, 2019).

CONCLUSION

C. sakazakii, *C. malonaticus*, and *S.* Typhimurium strains isolated from PIF exhibit antibiotic resistance and various virulence genes and resistance to β-lactam antibiotics. Continuous monitoring of these powdered infant formulas is necessary due to the risk associated with pathogen contamination of the product and consumption by the immunologically vulnerable child population.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

JP-F, OH, SA, SL, AC-C, JM-R, AC, and SF conceived the experiments and prepared the manuscript. JP-F, AC-F, PC-S,

AC-C, AP, SL, and WR conducted the laboratory work. JP-F, OH, SA, AC-C, JX-C, JM-R, AC, WR, and SF drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Research Directorate of the Universidad del Bío-Bío, Projects 191520 4/R and GI 195420/EF.

ACKNOWLEDGMENTS

We wish to thank the Research Directorate of the Universidad del Bío-Bío.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.884721/full#supplementary-material

REFERENCES

Achtman, M., Wain, J., Weill, F. X., Nair, S., Zhou, Z., Sangal, V., et al. (2012). Multilocus sequence typing as a replacement for serotyping in *Salmonella* enterica. *PLoS Pathog.* 8:e1002776. doi: 10.1371/journal.ppat.1002776

Aldubyan, M., Almami, I., Benslimane, F., Alsonosi, A., and Forsythe, S. (2017). Comparative outer membrane protein analysis of high and low-invasive strains of *Cronobacter malonaticus*. *Front. Microbiol.* 8:2268. doi: 10.3389/fmicb.2017.02268

Aly, M. A., Domig, K. J., Kneifel, W., and Reimhult, E. (2019). Whole genome sequencing-based comparison of food isolates of *Cronobacter sakazakii*. *Front. Microbiol.* 10:1464. doi: 10.3389/fmicb.2019.01464

Angulo, F., Cahill, S., Wachsmuth, K., Costarrica, M. L., and Ben Embarek, P. (2008). Powdered infant formula as a source of *Salmonella* infection in infants. *Clin. Infect. Dis.* 46, 268–273. doi: 10.1086/524737

Baldwin, A., Loughlin, M., Caubilla-Barron, J., Kucerova, E., Manning, G., Dowson, C., et al. (2009). Multilocus sequence typing of *Cronobacter sakazakii* and *Cronobacter malonaticus* reveals stable clonal structures with clinical significance which do not correlate with biotypes. *BMC Microbiol.* 9:223. doi: 10.1186/1471-2180-9-223

Barilli, E., Bacci, C., StellaVilla, Z., Merialdi, G., D'Incau, M., Brindani, F., et al. (2018). Antimicrobial resistance, biofilm synthesis and virulence genes in *Salmonella* isolated from pigs bred on intensive farms. *Ital. J. Food Saf.* 7:7223. doi: 10.4081/ijfs.2018.7223

Biswas, A., Staals, R., Morales, S., Fineran, P., and Brown, C. (2016). CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics* 17:356. doi: 10.1186/s12864-016-2627-0

Borowiak, M., Baumann, B., Fischer, J., Thomas, K., Deneke, C., Hammerl, J. A., et al. (2020). Development of a novel *mcr-6* to *mcr-9* multiplex PCR and assessment of *mcr-1* to *mcr-9* occurrence in colistin-resistant *Salmonella* enterica isolates from environment, feed, animals and food (2011–2018) in Germany. *Front. Microbiol.* 11:80. doi: 10.3389/fmicb.2020.00080

Brouard, C., Espié, E., Weill, F. X., Kérouanton, A., Brisabois, A., Forgue, A. M., et al. (2007). Two consecutive large outbreaks of *Salmonella* enterica serotype Agona infections in infants linked to the consumption of powdered infant formula. *Pediatr. Infect. Dis. J.* 26, 148–152. doi: 10.1097/01.inf.0000253219.06258.23

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). In silico detection and typing of plasmids using plasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Carrasco, E., Morales-Rueda, A., and García-Gimeno, R. M. (2012). Cross-contamination and recontamination by *salmonella* in foods: a review. *Food Res. Int.* 45, 545–556. doi: 10.1016/j.foodres.2011.11.004

Carroll, L., Gaballa, A., Guldimann, C., Sullivan, G., Henderson, L., and Wiedmann, M. (2019). Identification of novel mobilized colistin resistance gene *mcr-9* in a multidrug-resistant, colistin-susceptible *Salmonella* enterica serotype Typhimurium isolate. *MBio* 10, e00853–e00919. doi: 10.1128/mBio.00853-19

Carroll, L. M., Wiedmann, M., den Bakker, H., Siler, J., Warchocki, S., Kent, D., et al. (2017). Whole-genome sequencing of drug-resistant *Salmonella* enterica isolates from dairy cattle and humans in New York and Washington States reveals source and geographic associations. *Appl. Environ. Microbiol.* 83, e00140–e00217. doi: 10.1128/AEM.00140-17

Caubilla-Barron, J., and Forsythe, S. (2007). Dry stress and survival time of *Enterobacter sakazakii* and other Enterobacteriaceae in dehydrated powdered infant formula. *J. Food Prot.* 70, 2111–2117. doi: 10.4315/0362-028x-70.9.2111

Cha, M. H., Woo, G. J., Lee, W., Kim, S. H., Woo, J. H., Kim, J., et al. (2020). Emergence of transferable mcr-9 gene-carrying colistin-resistant *Salmonella* enterica Dessau ST14 isolated from retail chicken meat in Korea. *Foodborne Pathog. Dis.* 17, 720–727. doi: 10.1089/fpd.2020.2810

Cheng, R. A., Eade, C. R., and Wiedmann, M. (2019). Embracing diversity: differences in virulence mechanisms, disease severity, and host adaptations contribute to the success of Nontyphoidal *Salmonella* as a foodborne pathogen. *Front. Microbiol.* 10:1368. doi: 10.3389/fmicb.2019.01368

Chon, J., Song, K., Kim, S., Hyeon, J., and Seo, K. (2012). Isolation and characterization of *Cronobacter* from desiccated foods in Korea. *J. Food Sci.* 77, 354–358. doi: 10.1111/j.1750-3841.2012.02750.x

CLSI. (2020). *Performance Standards for Antimicrobial Susceptibility Testing. 30th Edn.* Wayne, PA: CLSI.

Codex Alimentarius Commission (2007). Standard for Infant Formula and Formulas for Special Medical Purposes Intended for Infants. CXS 72-1981 Revised 2007, 119–11.

Cordier, J.-L. (2008). "Chapter 6: Production of powdered infant formulae and microbiological control measures," in *Enterobacter sakazakii*. eds. J. Farber and S. J. Forsythe (Washington, D.C.: ASM Press).

Costa, P. V., Vasconcellos, L., Forsythe, S. J., and Brandão, M. L. L. (2021). Diversity of *Cronobacter* genus isolated between 1970 and 2019 on the American continent and genotyped using multi-locus sequence typing. *FEMS Microbiol. Lett.* 368, 1–9. doi: 10.1093/femsle/fnab027

Craven, H. H., McAuley, C. M., Duffy, L. L., and Fegan, N. (2010). Distribution, prevalence and persistence of Cronobacter (Enterobacter sakazakii) in the non-processing and processing environments of five milk powder factories. *J. Appl. Microbiol.* 109, 1044–1052. doi: 10.1111/j.1365-2672.2010.04733.x

Cruz, A., Xicohtencatl, J., Gonzalez, B., Bobadilla, M., Eslava, C., and Rosas, I. (2011). Virulence traits in *Cronobacter* species isolated from different sources. *Can. J. Microbiol.* 57, 735–744. doi: 10.1139/w11-063

Csorba, C., Pajić, M., Blagojević, B., Forsythe, S., Radinović, M., and Velebit, B. (2021). Prevalence, characterization, and antibiotic susceptibility of *Cronobacter* spp. in a milk powder processing environment: the first reported case in Serbia. *Food Sci. Nutr.* 10, 554–563. doi: 10.1002/fsn3.2681

de Frutos, M., López-Urrutia, L., Berbel, C., Allue, M., Herrera, S., Azcona, J. M., et al. (2018). Brote de *Salmonella* Typhimurium monofásica asociada al consumo de carne asada de cerdo [Monophasic Salmonella Typhimurium outbreak due to the consumption of roast pork meat]. *Rev. Esp. Quimioter.* 31, 156–159.

de Toro, M., Sáenz, Y., Cercenado, E., Rojo-Bezares, B., García-Campello, M., Undabeitia, E., et al. (2011). Genetic characterization of the mechanisms of resistance to amoxicillin/clavulanate and third-generation cephalosporins in *Salmonella* enterica from three Spanish hospitals. *Int. Microbiol.* 14, 173–181. doi: 10.2436/20.1501.01.146

de Toro, M., Seral, C., Rojo-Bezares, B., Torres, C., Castillo, F. J., and Sáenz, Y. (2014). Resistencia a antibióticos y factores de virulencia en aislados clínicos de *Salmonella* enterica. *Enferm. Infecc. Microbiol. Clin.* 32, 4–10. doi: 10.1016/j.eimc.2013.03.006

Deter, H. S., Jensen, R. V., Mather, W. H., and Butzin, N. C. (2017). Mechanisms for differential protein production in toxin–antitoxin systems. *Toxins* 9:211. doi: 10.3390/toxins9070211

Falagas, M., Athanasaki, F., Voulgaris, G., Triarides, N., and Vardakas, K. (2019). Resistance to fosfomycin: mechanisms, frequency and clinical consequences. *Int. J. Antimicrob. Agents* 53, 22–28. doi: 10.1016/j.ijantimicag.2018.09.013

FAO/WHO (2004). *Enterobacter sakazakii* and Other Microorganisms in Powdered Infant Formula: Meeting Report. Microbiological Risk Assessment. Series No. 6 Geneva: WHO.

FAO/WHO (2006). *Enterobacter Sakazakii* and *Salmonella* Powdered Infant Formula. Microbiological Risk Assessment. Series No. 10. Rome: FAO

FAO/WHO (2008). Enterobacter sakazakii and other microorganisms in powdered follow-up formula. Meeting report. Microbiological Risk Assessment. Series No. 15. Rome: FAO.

FDA (2019). Investigations Operation Manual. Available at: https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-references/investigations-operations-manual (Accessed March, 2021).

Fei, P., Jiang, Y., Feng, J., Forsythe, S. J., Li, R., Zhou, Y., et al. (2017). Antibiotic and desiccation resistance of *Cronobacter sakazakii* and *C. malonaticus* isolates from powdered infant formula and processing environments. *Front. Microbiol.* 8:316. doi: 10.3389/fmicb.2017.00316

Fei, P., Man, C., Lou, B., Forsythe, S., Chai, Y., Li, R., et al. (2015). Genotyping and source tracking of the *Cronobacter sakazakii* and *C. malonaticus* isolated from powdered infant formula and an infant formula production factory in China. *Appl. Environ. Microbiol.* 81, 5430–5439. doi: 10.1128/AEM.01390-15

Ferri, M., Ranucci, E., Romagnoli, P., and Giaccone, V. (2017). Antimicrobial resistance: a global emerging threat to public health systems. *Crit. Rev. Food Sci. Nutr.* 57, 2857–2876. doi: 10.1080/10408398.2015.1077192

Finkelstein, S., Negrete, F., Jang, H., Gangiredla, J., Mammel, M., Patel, I. R., et al. (2019). Prevalence, distribution, and phylogeny of type two toxin-antitoxin genes possessed by *Cronobacter* species where *C. sakazakii* homologs follow sequence type lineages. *Microorganisms* 7:554. doi: 10.3390/microorganisms7110554

Flores, J. P., Arvizu, S., Silva, J., and Fernández-Escartín, E. (2011). Two cases of hemorrhagic diarrhea caused by *Cronobacter sakazakii* in hospitalized nursing infants associated with the consumption of powdered infant formula. *J. Food Prot.* 74, 2177–2181. doi: 10.4315/0362-028X.JFP-11-257

Forsythe, S. J. (2018). Updates on the *Cronobacter* genus. *Annu. Rev. Food Sci. Technol.* 9, 23–44. doi: 10.1146/annurev-food-030117-012246

Franco, A. A., Kothary, M., Gopinath, G., Jarvis, K., Grim, C. J., Hu, L., et al. (2011). Cpa, the outer membrane protease of *Cronobacter sakazakii*, activates plasminogen and mediates resistance to serum bactericidal activity. *Infect. Immun.* 79, 1578–1587. doi: 10.1128/IAI.01165-10

Garallah, E. T., and Al-Jubori, S. (2020). Molecular detection of *glpT* and *uhpT* genes as fosfomycin pathways in UTI infection patients. *Gene Rep.* 21:100930. doi: 10.1016/j.genrep.2020.100930

Gordillo Altamirano, F. L., and Barr, J. J. (2019). Phage therapy in the postantibiotic era. *Clin. Microbiol. Rev.* 32, e00066–e00068. doi: 10.1128/CMR.00066-18

Guerin, E., Shkoporov, A., Stockdale, S. R., Clooney, A. G., Ryan, F. J., Sutton, T., et al. (2018). Biology and taxonomy of *crAss*-like bacteriophages, the Most abundant virus in the human gut. *Cell Host Microbe* 24, 653.e6–664.e6. doi: 10.1016/j.chom.2018

Güerri, M. L., Aladueña, A., Echeíta, A., and Rotger, R. (2004). Detection of integrons and antibiotic-resistance genes in *Salmonella* enterica serovar Typhimurium isolates with resistance to ampicillin and variable susceptibility to amoxicillin-clavulanate. *Int. J. Antimicrob. Agents* 24, 327–333. doi: 10.1016/j.ijantimicag.2004.04.009

Holý, O., Cruz-Cordova, A., Xicohtencatl-Cortés, J., Hochel, I., Parra-Flores, J., Petrzelova, J., et al. (2019). Occurrence of virulence factors in *Cronobacter sakazakii* and *Cronobacter malonaticus* originated from clinical samples. *Microb. Pathog.* 127, 250–256. doi: 10.1016/j.micpath.2018.12.011

Holý, O., Parra-Flores, J., Lepuschitz, S., Alarcón-Lavín, M. P., Cruz-Córdova, A., Xicohtencatl-Cortes, J., et al. (2021). Molecular characterization of *Cronobacter sakazakii* strains isolated from powdered milk. *Foods* 10:20. doi: 10.3390/foods10010020

Hu, L., Cao, G., Brown, E. W., Allard, M., Ma, L., and Zhang, G. (2021). Whole genome sequencing and protein structure analyses of target genes for the detection of *Salmonella*. *Sci. Rep.* 11:20887. doi: 10.1038/s41598-021-00224-7

Huehn, S., La Ragione, R. M., Anjum, M., Saunders, M., Woodward, M. J., Bunge, C., et al. (2010). Virulotyping and antimicrobial resistance typing of *Salmonella* enterica serovars relevant to human health in Europe. *Foodborne Pathog. Dis.* 7, 523–535. doi: 10.1089/fpd.2009.0447

Ingram, P., Inglis, T., Vanzetti, T., Henderson, B., Harnett, G., and Murray, R. (2011). Comparison of methods for *AmpC β-lactamase* detection in Enterobacteriaceae. *J. Med. Microbiol.* 60, 715–721. doi: 10.1099/jmm.0.029140-0.

Instituto Nacional de Normalización (2002). Chile. Norma Chilena NCh 2675 de detección de Salmonella. Available at: https://ecommerce.inn.cl/nch2675200244470 (Accessed October 10, 2020).

International Standards for Organization (ISO) (2017). ISO 22964:2017: microbiology of the food chain—horizontal method for the detection of *Cronobacter* spp. international standards Organization. Available at: https://www.iso.org/standard/64708.htm

Iversen, C., and Forsythe, S. (2004). Isolation of *Enterobacter sakazakii* and other *Enterobacteriaceae* from powdered infant formula milk and related products. *Food Microbiol.* 21, 771–777. doi: 10.1016/j.fm.2004.01.009

Iversen, C., Mullane, N., Mc Cardell, B., Tall, B. D., Lehner, A., Fanning, S., et al. (2008). *Cronobacter* gen. nov., a new genus to accommodate the biogroups of *Enterobacter sakazakii*, and proposal of *Cronobacter sakazakii* gen. Nov. comb. nov., *C. malonaticus* sp. nov., *C. turicensis* sp. nov., *C. muytjensii* sp. nov., *C. dublinensis* sp. nov., *Cronobacter* genomospecies 1, and of three subspecies, *C. dublinensis* sp. nov. subsp. *dublinensis* subsp. nov., *C. dublinensis* sp. nov. subsp. *lausannensis* subsp. nov., and *C. dublinensis* sp. nov. subsp. *lactaridi* subsp. nov. *Int. J. Syst. Evol. Microbiol.* 58, 1442–1447. doi: 10.1099/ijs.0.65577-0

Jain, P., Sudhanthirakodi, S., Chowdhury, G., Joshi, S., Anandan, S., Ray, U., et al. (2018). Antimicrobial resistance, plasmid, virulence, multilocus sequence typing and pulsed-field gel electrophoresis profiles of *Salmonella* enterica serovar Typhimurium clinical and environmental isolates from India. *PLoS One* 13:e0207954. doi: 10.1371/journal.pone.0207954

Jang, H., Chase, H. R., Gangiredla, J., Grim, C. J., Patel, I. R., Kothary, M. H., et al. (2020). Analysis of the molecular diversity among *Cronobacter* species isolated from filth flies using targeted PCR, pan genomic DNA microarray, and whole genome sequencing analyses. *Front. Microbiol.* 11:561204. doi: 10.3389/fmicb.2020.561204

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004

Johansson, M., Bortolaia, V., Tansirichaiya, S., Aarestrup, F. M., Roberts, A. P., and Petersen, T. N. (2021). Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella* enterica using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* 76, 101–109. doi: 10.1093/jac/dkaa390

Jones, G., Pardos de la Gandara, M., Herrera-Leon, L., Herrera-Leon, S., Varela Martinez, C., Hureaux-Roy, R., et al. (2019). Outbreak of *Salmonella* enterica serotype Poona in infants linked to persistent *Salmonella* contamination in an infant formula manufacturing facility, France, august 2018 to February 2019. *Euro Surveill.* 24:1900161. doi: 10.2807/1560-7917.ES.2019.24.13.1900161

Joseph, S., Cetinkaya, E., Drahovska, H., Levican, A., Figueras, M., and Forsythe, S. (2012). *Cronobacter condimenti* sp. nov., isolated from spiced meat, and *Cronobacter universalis* sp. nov., a species designation for *Cronobacter* sp. Geneomoespecies 1, recovered from a leg infection, water and food ingredients. *Int. J. Syst. Evol. Microbiol.* 62, 1277–1283. doi: 10.1099/ijs.0.032292-0

Joseph, S., and Forsythe, S. J. (2011). Predominance of *Cronobacter sakazakii* sequence type 4 in neonatal infections. *Emerg. Infect. Dis.* 17, 1713–1715. doi: 10.3201/eid1709.110260

Joseph, S., and Forsythe, S. (2012). Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Front. Microbiol.* 3:397. doi: 10.3389/fmicb.2012.00397

Jourdan-da Silva, N., Fabre, L., Robinson, E., Fournet, N., Nisavanh, A., Bruyand, M., et al. (2018). Ongoing nationwide outbreak of *Salmonella* Agona associated with internationally distributed infant milk products, France,

December 2017. *Euro Surveill.* 23, 17–00852. doi: 10.2807/1560-7917.ES.2018.23.2.17-00852

Jünemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., et al. (2013). Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* 31, 294–296. doi: 10.1038/nbt.2522

Khezri, A., Avershina, E., and Ahmad, R. (2021). Plasmid identification and plasmid-mediated antimicrobial gene detection in norwegian isolates. *Microorganisms* 9:52. doi: 10.3390/microorganisms9010052

Kieffer, N., Royer, G., Decousser, J. W., Bourrel, A. S., Palmieri, M., Ortiz De La Rosa, J. M., et al. (2019). *Mcr-9*, an inducible gene encoding an acquired phosphoethano-lamine transferase in *Escherichia coli*, and its origin. *Antimicrob. Agents Chemother.* 63, e00965–e001019. doi: 10.1128/AAC.00965-19

Kim, K., Jang, S., Kim, S., Park, J., Heu, S., and Ryu, S. (2008). Prevalence and genetic diversity of *Enterobacter sakazakii* in ingredients of infant foods. *Int. J. Food Microbiol.* 122, 196–203. doi: 10.1016/j.ijfoodmicro.2007.11.072

Kim, K., Kim, K., Choi, J., Lim-Jeong, A., Lee, J., Hwang, S., et al. (2010). Outer membrane proteins a (OmpA) and x (OmpX) are essential for basolateral invasion of *Cronobacter sakazakii*. *Appl. Environ. Microbiol.* 76, 5188–5198. doi: 10.1128/AEM.02498-09

Lachowska, M., Izdebski, R., Urbanowicz, P., Żabicka, D., and Królak-Olejnik, B. (2021). Infection of *Cronobacter sakazakii* ST1 producing SHV-12 in a premature infant born from triplet pregnancy. *Microorganisms* 9:1878. doi: 10.3390/microorganisms9091878

Lange, S., Alkhnbashi, O., Rose, D., Will, S., and Backofen, R. (2013). CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res.* 41, 8034–8044. doi: 10.1093/nar/gkt606

Lee, Y.-D., Park, J., and Chang, H. (2012). Detection, antibiotic susceptibility and biofilm formation of *Cronobacter* spp. from various foods in Korea. *Food Control* 24, 225–230. doi: 10.1016/j.foodcont.2011.09.023

Lehner, A., Tall, B. D., Fanning, S., and Srikumar, S. (2018). *Cronobacter* spp.—opportunistic foodborne pathogens: an update on evolution, osmotic adaptation and pathogenesis. *Curr. Clin. Microbiol. Rep.* 5, 97–105. doi: 10.1007/s40588-018-0089-7

Leopold, S., Goering, R., Witten, A., Harmsen, D., and Mellmann, A. (2014). Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J. Clin. Microbiol.* 52, 2365–2370. doi: 10.1128/JCM.00262-14

Lepuschitz, S., Ruppitsch, W., Pekard-Amenitsch, S., Forsythe, S. J., Cormican, M., Mach, R. L., et al. (2019). Multicenter study of *Cronobacter sakazakii* infections in humans, Europe, 2017. *Emerg. Infect. Dis.* 25, 515–522. doi: 10.3201/eid2503.181652

Lepuschitz, S., Sorschag, S., Springer, B., Allerberger, F., and Ruppitsch, W. (2017). Draft genome sequence of carbapenemase-producing *Serratia marcescens* isolated from a patient with chronic obstructive pulmonary disease. *Genome Announc.* 5, e01288–e01317. doi: 10.1128/genomeA.01288-1

Lepuschitz, S., Weinmaier, T., Mrazek, K., Beisken, S., Weinberger, J., and Posch, A. E. (2020). Analytical performance validation of next-generation sequencing based clinical microbiology assays using a K-mer analysis workflow. *Front. Microbiol.* 11:1883. doi: 10.3389/fmicb.2020.01883

Lesnick, M. L., Reiner, N. E., Fierer, J., and Guiney, D. G. (2001). The *Salmonella spvB* virulence gene encodes an enzyme that ADP-ribosylates actin and destabilizes the cytoskeleton of eukaryotic cells. *Mol. Microbiol.* 39, 1464–1470. doi: 10.1046/j.1365-2958.2001.02360.x

Lou, L., Zhang, P., Piao, R., and Wang, Y. (2019). *Salmonella* Pathogenicity Island 1 (SPI-1) and its complex regulatory network. *Front. Cell. Infect. Microbiol.* 9:270. doi: 10.3389/fcimb.2019.00270

Louwen, R., Staals, R. H., Endtz, H. P., van Baarlen, P., and van der Oost, J. (2014). The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 78, 74–88. doi: 10.1128/MMBR.00039-13

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569

Mambu, J., Barilleau, E., Fragnet-Trapp, L., Le Vern, Y., Olivier, M., Sadrin, G., et al. (2020). *Rck* of *Salmonella* Typhimurium delays the host cell cycle to facilitate bacterial invasion. *Front. Cell. Infect. Microbiol.* 10:586934. doi: 10.3389/fcimb.2020.586934

Mange, J. P., Stephan, R., Borel, L., Wild, P., Kim, K. S., Pospischil, A., et al. (2006). Adhesive properties of *Enterobacter sakazakii* to human epithelial and brain microvascular endothelial cells. *BMC Microbiol.* 6, 58–68. doi: 10.1186/1471-2180-6-58

Mare, A., Man, A., Toma, F., Ciurea, C. N., Coşeriu, R. L., Vintilă, C., et al. (2020). Hemolysin-producing strains among diarrheagenic *Escherichia coli* isolated from children under 2 years old with diarrheal disease. *Pathogens* 9:1022. doi: 10.3390/pathogens9121022

Marraffini, L. A. (2013). CRISPR-Cas immunity against phages: its effects on the evolution and survival of bacterial pathogens. *PLoS Pathog.* 9:e1003765. doi: 10.1371/journal.ppat.1003765

Mazzantini, D., Fonnesu, R., Celandroni, F., Calvigioni, M., Vecchione, A., Mrusek, D., et al. (2020). GTP-dependent FlhF homodimer supports secretion of a hemolysin in *Bacillus cereus*. *Front. Microbiol.* 11:879. doi: 10.3389/fmicb.2020.00879

Mezal, E. H., Sabol, A., Khan, M. A., Ali, N., Stefanova, R., and Khan, A. A. (2014). Isolation and molecular characterization of *salmonella* enterica serovar Enteritidis from poultry house and clinical samples during 2010. *Food Microbiol.* 38, 67–74. doi: 10.1016/j.fm.2013.08.003

Molloy, C., Cagney, C., O'Brien, S., Iversen, C., Fanning, S., and Duffy, G. (2009). Surveillance and characterization by pulsed-field gel electrophoresis of *Cronobacter* spp in farming and domestic environments, food production animals and retails foods. *Int. J. Food Microbiol.* 136, 198–203. doi: 10.1016/j.ijfoodmicro.2009.07.007

Monte, D. F. M., Sellera, F. P., Lopes, R., Keelara, S., Landgraf, M., Greene, S., et al. (2020). Class 1 integron-borne cassettes harboring blaCARB-2 gene in multidrug-resistant and virulent *Salmonella* Typhimurium ST19 strains recovered from clinical human stool samples, United States. *PLoS One* 15:e0240978. doi: 10.1371/journal.pone.0240978

Mossong, J., Marques, P., Ragimbeau, C., Huberty-Krau, P., Losch, S., Meyer, G., et al. (2007). Outbreaks of monophasic *Salmonella* enterica serovar 4, [5],12: i: - in Luxembourg, 2006. *Euro Surveill.* 12, E11–E12. doi: 10.2807/esm.12.06.00719-en

Müller, A., Hächler, H., Stephan, R., and Lehner, A. (2014). Presence of *AmpC beta-lactamases*, *CSA-1*, *CSA-2*, *CMA-1*, and *CMA-2* conferring an unusual resistance phenotype in *Cronobacter sakazakii* and *Cronobacter malonaticus*. *Microb. Drug Resist.* 20, 275–280. doi: 10.1089/mdr.2013.0188

Murugkar, H. V., Rahman, H., and Dutta, P. K. (2003). Distribution of virulence genes in *Salmonella* serovars isolated from man & animals. *Indian J. Med. Res.* 117, 66–70.

Nayak, R., Stewart, T., Wang, R. F., Lin, J., Cerniglia, C. E., and Kenney, P. B. (2004). Genetic diversity and virulence gene determinants of antibiotic-resistant *Salmonella* isolated from preharvest Turkey production sources. *Int. J. Food Microbiol.* 91, 51–62. doi: 10.1016/S0168-1605(03)00330-1

Ogrodzki, P., and Forsythe, S. (2015). Capsular profiling of the *Cronobacter* genus and the association of specific *Cronobacter sakazakii* and *C. malonaticus* capsule types with neonatal meningitis and necrotizing enterocolitis. *BMC Genomics* 16:758. doi: 10.1186/s12864-015-1960-z

Ogrodzki, P., and Forsythe, S. (2016). CRISPR–cas loci profiling of *Cronobacter sakazakii* pathovars. *Future Microbiol.* 11, 1507–1519. doi: 10.2217/fmb-2016-0070

Panzenhagen, P. H. N., Paul, N. C., Conte, C. A., Costa, R. G., Rodrigues, D. P., and Shah, D. H. (2018). Genetically distinct lineages of *Salmonella* Typhimurium ST313 and ST19 are present in Brazil. *Int. J. Med. Microbiol.* 308, 306–316. doi: 10.1016/j.ijmm.2018.01.005

Parra-Flores, J., Cruz-Córdova, A., Acuña, S., Riffo-Sepúlveda, F., Maury-Sintjago, E., Rodriguez-Fernández, A., et al. (2021a). "*Cronobacter* spp. in milk," in *Reference Module in Food Science*. eds. P. L. H. McSweeney and J. P. McNamara (Amsterdam, Netherlands: Elsevier), 1–9.

Parra-Flores, J., Aguirre, J., Juneja, V., Jackson, E., Cruz-Córdova, A., Silva-Sanchez, J., et al. (2018a). Virulence and antibiotic resistance profiles of *Cronobacter sakazakii* and *Enterobacter* spp. involved in the diarrheic hemorrhagic outbreak in Mexico. *Front. Microbiol.* 9:2206. doi: 10.3389/fmicb.2018.02206

Parra-Flores, J., Cerda-Leal, F., Contreras, A., Valenzuela-Riffo, N., Rodriguez, A., and Aguirre, J. (2018b). *Cronobacter sakazakii* and microbiological parameters in dairy formulas associated with a food alert in Chile. *Front. Microbiol.* 9:1708. doi: 10.3389/fmicb.2018.01708.

Parra-Flores, J., Holý, O., Riffo, F., Lepuschitz, S., Maury-Sintjago, E., Rodríguez-Fernández, A., et al. (2021b). Profiling the virulence and antibiotic resistance genes of *Cronobacter sakazakii* strains isolated from powdered

and dairy formulas by whole-genome sequencing. *Front. Microbiol.* 12:694922. doi: 10.3389/fmicb.2021.694922

Parra-Flores, J., Maury-Sintjago, E., Rodríguez-Fernández, A., Acuña, S., Cerda, F., Aguirre, J., et al. (2020). Microbiological quality of powdered infant formula in Latin America. *J. Food Prot.* 83, 534–541. doi: 10.4315/0362-028X.JFP-19-399

Podolak, R., and Black, D. G. (2017). "Introduction and overview," in *In Control of Salmonella and Other Bacterial Pathogens in Low-Moisture Foods. 1st Edn*. eds. R. Podolak and D. G. Black (West Sussex, UK: John Wiley and Sons), 1–27.

Ramsamy, Y., Mlisana, K. P., Allam, M., Amoako, D. G., Abia, A. L. K., Ismail, A., et al. (2020). Genomic analysis of carbapenemase-producing extensively drug-resistant *Klebsiella pneumoniae* isolates reveals the horizontal spread of p18-43_01 plasmid encoding blaNDM-1 in South Africa. *Microorganisms* 8:137. doi: 10.3390/microorganisms8010137

Richter, L., du Plessis, E. M., Duvenage, S., and Korsten, L. (2020). Occurrence, phenotypic and molecular characterization of extended-Spectrum- and *AmpC-β-lactamase* producing Enterobacteriaceae isolated from selected commercial spinach supply chains in South Africa. *Front. Microbiol.* 11:638. doi: 10.3389/fmicb.2020.00638

Severi, E., Hood, D. W., and Thomas, G. H. (2007). Sialic acid utilization by bacterial pathogens. *Microbiology* 153, 2817–2822. doi: 10.1099/mic.0.2007/009480-0

Shi, L., Liang, Q., Zhan, Z., Feng, J., Zhao, Y., Chen, Y., et al. (2018). Co-occurrence of 3 different resistance plasmids in a multi-drug resistant *Cronobacter sakazakii* isolate causing neonatal infections. *Virulence* 9, 110–120. doi: 10.1080/21505594.2017.1356537

Silver, L. L. (2017). Fosfomycin: mechanism and resistance. *Cold Spring Harb. Perspect. Med.* 7:a025262. doi: 10.1101/cshperspect.a025262

Sohanpal, B. K., Friar, S., Roobol, J., Plumbridge, J. A., and Blomfield, I. C. (2007). Multiple co-regulatory elements and IHF are necessary for the control of *fimB* expression in response to sialic acid and N-acetylglucosamine in *Escherichia coli* K-12. *Mol. Microbiol.* 63, 1223–1236. doi: 10.1111/j.1365-2958.2006.05583.x

Sonbol, H., Joseph, S., McAuley, C., Craven, H., and Forsythe, S. (2013). Multilocus sequence typing of *Cronobacter* spp. from powdered infant formula and milk powder production factories. *Int. Dairy J.* 30, 1–7. doi: 10.1016/j.idairyj.2012.11.004

Stephan, R., Grim, C., Gopinath, G., Mammel, M., Sathyamoorthy, V., Trach, L., et al. (2014). Re-examination of the taxonomic status of *Enterobacter helveticus*, *Enterobacter pulveris* and *Enterobacter turicensis* as members of the genus *Cronobacter* and their reclassification in the genera *Franconibacter* gen. nov. and *Siccibacter* gen. nov. as *Franconibacter helveticus* comb. nov., *Franconibacter pulveris* comb. nov. and *Siccibacter turicensis* comb. nov., respectively. *Int. J. Syst. Evol. Microbiol.* 64, 3402–3410. doi: 10.1099/ijs.0.059832-0

Thung, T. Y., Radu, S., Mahyudin, N. A., Rukayadi, Y., Zakaria, Z., Mazlan, N., et al. (2018). Prevalence, virulence genes and antimicrobial resistance profiles of *Salmonella* Serovars from retail beef in Selangor, Malaysia. *Front. Microbiol.* 8:2697. doi: 10.3389/fmicb.2017.02697

U.S FDA (2022). FDA Investigation of *Cronobacter* and *Salmonella* Complaints: Powdered Infant Formula (February 2022). Available at: https://www.fda.gov/food/outbreaks-foodborne-illness/fda-investigation-cronobacter-and-salmonella-complaints-powdered-infant-formula-february-2022 (Accessed February 21, 2022).

Uelze, L., Borowiak, M., Deneke, C., Szabó, I., Fischer, J., Tausch, S. H., et al. (2020). Performance and accuracy of four open-source tools for in silico serotyping of *salmonella* spp. based on whole-genome short-read sequencing data. *Appl. Environ. Microbiol.* 86, e02265–e02319. doi: 10.1128/AEM.02265-19

Walling, L. R., and Butler, J. S. (2019). Toxins targeting transfer RNAs: translation inhibition by bacterial toxin–antitoxin systems. *Wiley Interdiscip. Rev. RNA* 10:e1506. doi: 10.1002/wrna.1506

Wang, X., Biswas, S., Paudyal, N., Pan, H., Li, X., Fang, W., et al. (2019). Antibiotic resistance in *Salmonella* Typhimurium isolates recovered from the food chain through National Antimicrobial Resistance Monitoring System Between 1996 and 2016. *Front. Microbiol.* 10:985. doi: 10.3389/fmicb.2019.00985

Wang, X., Wang, Y., Ling, N., Shen, Y., Zhang, D., Liu, D., et al. (2021). Effects of *tolC* on tolerance to bile salts and biofilm formation in *Cronobacter malonaticus*. *J. Dairy Sci.* 104, 9521–9531. doi: 10.3168/jds.2021-20128

Wang, L., Zhu, W., Lu, G., Wu, P., Wei, Y., Su, Y., et al. (2021). In silico species identification and serotyping for *Cronobacter* isolates by use of whole-genome sequencing data. *Int. J. Food Microbiol.* 358:109405. doi: 10.1016/j.ijfoodmicro.2021.109405

Wei, Z., Xu, X., Yan, M., Chang, H., Li, Y., Kan, B., et al. (2019). *Salmonella* Typhimurium and *salmonella* Enteritidis infections in sporadic diarrhea in children: source tracing and resistance to third-generation Cephalosporins and ciprofloxacin. *Foodborne Pathog. Dis.* 16, 244–255. doi: 10.1089/fpd.2018.2557

Xiang, Y., Li, F., Dong, N., Tian, S., Zhang, H., Du, X., et al. (2020). Investigation of a salmonellosis outbreak caused by multidrug resistant *salmonella* Typhimurium in China. *Front. Microbiol.* 11:801. doi: 10.3389/fmicb.2020.00801

Zeng, H., Lei, T., He, W., Zhang, J., Liang, B., Li, C., et al. (2018). Novel multidrug resistant *Cronobacter sakazakii* causing meningitis in neonate, China, 2015. *Emerg. Infect. Dis.* 24, 2121–2124. doi: 10.3201/eid2411.180718

Zeng, H., Zhang, J., Li, C., Xie, T., Ling, N., Wu, Q., et al. (2017). The driving force of prophages and CRISPR-Cas system in the evolution of *Cronobacter sakazakii*. *Sci. Rep.* 7:40206. doi: 10.1038/srep46783

Zhang, S., Yin, Y., Jones, M. B., Zhang, Z., Deatherage Kaiser, B. L., Dinsmore, B. A., et al. (2015). *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J. Clin. Microbiol.* 53, 1685–1692. doi: 10.1128/JCM.00323-15

Zhang, F., Zhao, S., Ren, C., Zhu, Y., Lai, Y., Zhou, F., et al. (2018). CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Commun. Biol.* 1:180. doi: 10.1038/s42003-018-0184-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 June 2022
doi: 10.3389/fmicb.2022.914036

Novel Insights Into the Phylogeny and Biotechnological Potential of *Weissella* Species

Francesca Fanelli[1†], *Marco Montemurro*[2†], *Daniele Chieffi*[1], *Gyu-Sung Cho*[3], *Charles M. A. P. Franz*[3], *Anna Dell'Aquila*[1], *Carlo Giuseppe Rizzello*[4] and *Vincenzina Fusco*[1*]

[1] National Research Council, Institute of Sciences of Food Production (CNR-ISPA), Bari, Italy, [2] Department of Soil, Plant and Food Science, University of Bari Aldo Moro, Bari, Italy, [3] Department of Microbiology and Biotechnology, Max Rubner-Institut, Kiel, Germany, [4] Department of Environmental Biology, Sapienza University of Rome, Rome, Italy

OPEN ACCESS

Edited by:
Caroline Barretto,
Nestle, Switzerland

Reviewed by:
Laura Pérez-Través,
Institute of Agrochemistry and Food Technology (CSIC), Spain
Zhongjing Lu,
Kennesaw State University, United States

***Correspondence:**
Vincenzina Fusco
vincenzina.fusco@ispa.cnr.it

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: 06 April 2022
Accepted: 19 May 2022
Published: 22 June 2022

Citation:
Fanelli F, Montemurro M, Chieffi D, Cho G-S, Franz CMAP, Dell'Aquila A, Rizzello CG and Fusco V (2022) Novel Insights Into the Phylogeny and Biotechnological Potential of Weissella Species. Front. Microbiol. 13:914036. doi: 10.3389/fmicb.2022.914036

In this study, the genomes of the *Weissella* (*W.*) *beninensis*, *W. diestrammenae*, *W. fabalis*, *W. fabaria*, *W. ghanensis*, and *W. uvarum* type strains were sequenced and analyzed. Moreover, the ability of these strains to metabolize 95 carbohydrates was investigated, and the genetic determinants of such capability were searched within the sequenced genomes. 16S rRNA gene and genome-based-phylogeny of all the *Weissella* species described to date allowed a reassessment of the *Weissella* genus species groups. As a result, six distinct species groups within the genus, namely, *W. beninensis, W. kandleri, W. confusa, W. halotolerans, W. oryzae*, and *W. paramesenteroides* species groups, could be described. Phenotypic analyses provided further knowledge about the ability of the *W. beninensis*, *W. ghanensis*, *W. fabaria*, *W. fabalis*, *W. uvarum,* and *W. diestrammenae* type strains to metabolize certain carbohydrates and confirmed the interspecific diversity of the analyzed strains. Moreover, in many cases, the carbohydrate metabolism pathway and phylogenomic species group clustering overlapped. The novel insights provided in our study significantly improved the knowledge about the *Weissella* genus and allowed us to identify features that define the role of the analyzed type strains in fermentative processes and their biotechnological potential.

Keywords: *Weissella* spp., phylogenomics, genomics, carbohydrate metabolism, trehalose

INTRODUCTION

Weissella is a genus belonging to the phylum *Firmicutes,* order *Lactobacillales,* and family *Lactobacillaceae*. *Weissella* species are nonspore-forming, catalase-negative, and Gram-positive bacteria with either a coccoid or rod shape. They have a widespread occurrence, being isolated from various ecological niches such as soil (mainly *Weissella soli*) (Chen et al., 2005), sediments of coastal marsh and estuary (Sica et al., 2010), plants (Emerenini et al., 2014), and foods. Their presence in food matrices and many spontaneous fermentation processes of vegetables, dairy, and cereals (Wang et al., 2011; Zannini et al., 2013, 2018; Mun and Chang, 2020) reveals their role in

these fermentative processes, in which they exert an influence on the quality, and especially on the texture of foods. In addition, they have been isolated from the gut, oral cavity, breast milk (Martín et al., 2007; Albesharat et al., 2011), and urogenital and gastrointestinal tracts of humans (Lee et al., 2012), as well as from many animals, such as giant pandas and rainbow trout, and from the skin, milk, and gastrointestinal tracts of vertebrates (Xiong et al., 2019; Mortezaei et al., 2020; Wang et al., 2020).

The presence of these species as common inhabitants of the human intestine (Lee et al., 2012) and animal feces (Cai et al., 1998; Beasley et al., 2006; Muñoz-Atienza et al., 2013), as well as their occurrence in a variety of food matrices, has suggested their potential use as probiotics (Teixeira et al., 2021). This potential probiotic activity is supported by observations that some *Weissella* species can overcome the gastric barrier (Wang et al., 2008) and produce antimicrobial compounds such as bacteriocins (Srionnual et al., 2007; Masuda et al., 2012). For example, *W. cibaria* has been reported to inhibit biofilm formation *in vitro* and the proliferation of the main bacterial pathogens in dental caries and upper respiratory infections (Yeu et al., 2021), and has been studied for its potential anti-inflammatory activity against lipopolysaccharide stimulation (Yu et al., 2019), while *W. viridescens* was proposed as a potential probiotic for the skin (Espinoza-Monje et al., 2021). Furthermore, potential probiotic features of this genus include the reduction in depressive-like behavior (Sandes et al., 2020), the influence on gut permeability and intestinal epithelial regeneration (Prado et al., 2020), and the antagonistic activity against common human pathogens (Afolayan et al., 2017; Choi et al., 2018; Yu et al., 2019).

The capacity of this genus to grow at a wide range of temperatures, water activity levels, and pH (Ricciardi et al., 2009) increases its promising biotechnological potential, mainly attributed to the ability to produce exopolysaccharides (EPS) (Fusco et al., 2015). This has already promoted their use as starter cultures in dairy industries and sourdough fermentation (Galle et al., 2010), especially in the preparation of gluten-free baked products (Li et al., 2021). *W. cibaria* (Galli et al., 2020) and *W. confusa* have been largely studied for their ability to produce high amounts of dextrans, which enhance the softness of fresh bread (Wolter et al., 2014), and also for improving the textural properties of gluten-free bread (Montemurro et al., 2021).

The genus *Weissella*, originally described by Collins et al. (1993), to date comprises 21 validated species. The last taxonomic classification of *Weissella* species had been proposed by Fusco et al. (2015) who, updating the phylogenetic analysis performed by De Bruyne et al. (2010), clustered the species in 5 defined branches. These were designated as *W. kandleri* (including *W. kandleri*, *W. soli*, *W. diestrammenae*, *W. koreensis*, and *W. oryzae*), *W. confusa* (comprising *W. cibaria* and *W. confusa*), *W. halotolerans* (*W. halotolerans*, *W. ceti*, *W. viridescens*, *W. minor*, and *W. uvarum*), *W. paramesenteroides* (*W. thailandensis*, *W. hellenica*, and *W. paramesenteroides*), and *W. beninensis* (*W. beninensis*, *W. fabalis*, *W. fabaria*, and *W. ghanensis*) species groups, respectively.

Since 2015, seven additional *Weissella* species have been described, including *W. bombi* (Praet et al., 2015), *W. coleopterorum* (Hyun et al., 2021), *W. cryptocerci* (Heo et al., 2019), *W. muntiaci* (Lin et al., 2020), *W. sagaensis* (Li et al., 2020), and *W. jogaejeotgali* (Lee et al., 2015). According to Kwak et al. (2019), *W. jogaejeotgali* is a later heterotypic synonym of *W. thailandensis*, while Ennahar and Cai (2004) reported that *W. kimchii* (Choi et al., 2002) is a later heterotypic synonym of *W. cibaria* (Björkroth et al., 2002).

In this study, we sequenced the genomes of *W. beninensis*, *W. diestrammenae*, *W. fabalis*, *W. fabaria*, *W. ghanensis*, and *W. uvarum* type strains. These were the only *Weissella* species whose assemblies were not available at the time this study was performed. Based on our genomic analysis, we updated the taxonomic classification of *Weissella* species. We also phenotypically characterized the newly sequenced strains, evaluating their capability to metabolize different carbohydrates, while identifying the genetic determinants encoding these enzymatic activities.

MATERIALS AND METHODS

Strain Info and Culture Conditions

Weissella beninensis LMG 25373^T, *W. diestrammenae* DSM 27840^T, *W. fabalis* LMG 26217^T, *W. fabaria* LMG 24289^T, and *W. ghanensis* DSM 19935^T were purchased from DSMZ (website: http://www.dsmz.de) and BCCM/LMG (http://bccm.belspo.be/db/lmg_search_form.php), whereas the *W. uvarum* B18NM42 type strain (=DSM 28060^T) was kindly provided by Dr. Aspasia Nisiotou from the Institute of Technology of Agricultural Products, Hellenic Agricultural Organization Lykovrisi, Greece. The strains were grown as described by Fusco et al. (2011).

Whole-Genome Sequencing

A single colony of each strain was inoculated in 10 ml of MRS (de Man, Rogosa, and Sharpe) broth (Oxoid, Italy) and incubated overnight at 30°C. Two mL of each broth culture were washed in Tris-HCl (10 mM, pH 7.5) and resuspended in 500 μl of the same buffer. Genomic DNA was extracted using the peqGOLD bacterial DNA kit (Peqlab, Erlangen, Germany) according to the manufacturer's instructions. The integrity, purity, and quantity of DNA were assessed using agarose gel electrophoresis, Nanodrop photometer (Peqlab), and Qubit 3.0 fluorometer (Life Technologies). To prepare sequencing libraries, the Illumina TruSeq Nano DNA LT Library Prep Kit (MiSeq v3-kit) (Illumina, San Diego, USA) was used according to the manufacturer's instructions, and then sequenced on the Illumina MiSeq platform using the 2 × 250 pair procedure. Reads were then trimmed with the NxTrim (V2) (O'Connell et al., 2015) and the Trimmomatic (Bolger et al., 2014), and then *de novo* assembly was performed using SPAdes version 3.10.1 (Bankevich et al., 2012). The whole-genome shotgun projects were deposited at DDBJ/ENA/GenBank under the accessions JAGMVS000000000 for *W. beninensis* LMG 25373^T, JAGMVT000000000 for *W. diestrammenae* DSM 27840^T, JAGMVU000000000 for *W. fabalis* LMG 26217^T, JAGMVV000000000 for *W. fabaria* LMG 24289^T, JAGMVW000000000 for *W. ghanensis* DSM 19935^T, and JAGMVX000000000 for *W. uvarum* B18NM42^T. The versions described in this study are JAGMVS010000000

for *W. beninensis* LMG 25373^T, JAGMVT010000000 for *W. diestrammenae* DSM 27840^T, JAGMVU010000000 for *W. fabalis* LMG 26217^T, JAGMVV010000000 for *W. fabaria* LMG 24289^T, JAGMVW010000000 for *W. ghanensis* DSM 19935^T, and JAGMVX010000000 for *W. uvarum* B18NM42^T.

Bioinformatic Methods

The completeness of the assemblies was assessed by identifying a set of essential genes, which are typically present in a single copy in almost all prokaryotic genomes, by using MiGA (Rodriguez-R et al., 2018) and estimating the percentage of their presence in the genome. The quality scores were then calculated as the completeness percentage minus five times the contamination percentage.

Proteins were predicted by using the Prokaryotic Genome Annotation Pipeline (Tatusova et al., 2016) and PROKKA pipeline (Seemann, 2014) implemented in the Galaxy platform (Galaxy Version 1.14.6 + galaxy0; Afgan et al., 2018). Functional classification, subsystem prediction, and metabolic reconstruction comparison were performed using the RAST server (Aziz et al., 2008). All the protein sequences used in this study were retrieved from GenBank (NCBI). The homology-based relationship of *Weissella* strains toward selected proteins was determined using the BLASTP algorithm on the NCBI site (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Gene models were manually determined, and clustering and orientation were subsequently deduced for the closely linked genes. Genes were also retrieved by keyword search within the UniProtID entry list obtained by functional annotation and then manually curated.

Phylogenetic Analysis

A comparative genomic analysis was performed using strains and genomic assemblies listed in **Supplementary Table S1**. The 16S rRNA gene sequences were extracted from each genome by a BLASTn search and compared to the type strain sequences retrieved from available assemblies, whose accession numbers are indicated in **Supplementary Table S1**. An alignment was generated by MUSCLE 3.8.31 run mode using the phylogeny.fr platform (http://www.phylogeny.fr/index.cgi), configured as "A la Carte" Mode (Dereeper et al., 2008). The phylogenetic reconstruction was performed using the neighbor-joining method; the phylogenetic robustness was inferred by a bootstrapping procedure with 500 replications to obtain the confidence value for the aligned sequence dataset. The tree was graphically generated by iTOL version 5.5 (Letunic and Bork, 2019).

The genetic divergence was calculated using the ANI/AAI calculator (Goris et al., 2007; Rodriguez-R and Konstantinidis, 2016), which estimates the average nucleotide/amino acid identity (ANI/AAI) between genomic datasets using both best hits (one-way ANI) and reciprocal best hits (two-way ANI).

A genome-based phylogeny was reconstructed using the Phylogenetic Tree Building Service implemented in the Patric platform (www.patricbrc.org) with *Bifidobacterium bifidum* ATCC 29521^T as an outgroup according to Fusco et al. (2015) and the maximum likelihood method RAxML with progressive refinement (Stamatakis, 2014). For each strain, the genomic sequences deposited in GenBank were used for the analysis (**Supplementary Table S1**).

Comparative Genomic Analysis

A comparative analysis of carbohydrate-active enzymes (CAZy) was performed by executing a CAZyme annotation using the dbCAN meta server (http://bcb.unl.edu/dbCAN2/blast.php). The database was searched for preserved domains of all CAZy families, following the protocol proposed by dbCAN (Yin et al., 2012). A heatmap was manually constructed and visualized using the heatmapper web server (www.heatmapper.ca; Babicki et al., 2016), with average linkage as the clustering method and the Euclidean distance measurement method.

A comparative analysis of carbohydrate metabolism was performed by using the pathway reconstruction, annotating proteins with the Kyoto Encyclopedia of Genes and Genome (KEGG) Mapper (Kanehisa and Sato, 2020). A heatmap was again manually constructed based on the protein count in each pathway retrieved for individual genomes, and visualized by using the heatmapper web server (www.heatmapper.ca; Babicki et al., 2016) with average linkage as the clustering method and the Euclidean distance measurement method.

A comparative analysis of SEED subsystems was performed by uploading individual genome sequences to the SEED Viewer Server (Overbeek et al., 2014). Functional roles of RAST annotated genes were assigned and grouped in subsystem feature categories.

Phenotypic Characterization of *Weissella* Strains

Phenotypic Characterization

Phenotypic characterization was carried out as previously described by Fanelli et al. (2020), with some modifications. Biolog AN (Biolog, Inc., Hayward, CA, USA) plates were used to evaluate the consumption of 95 different carbon sources. Briefly, strains were grown in MRS broth (Oxoid, Italy) for 24 h. Cells were then harvested by centrifugation (10,000 rpm for 10 min) and washed two times with sterile phosphate buffer (50 mmol/l pH 7.0). Thereafter, cells were resuspended in sterile physiological saline (0.9 w/v NaCl). Each well of the plates was inoculated with 100 μl of bacterial suspension, adjusted to 65% transmittance. Plates were incubated in an anaerobic jar at 35°C for 24 h as recommended by the manufacturer. Positive reactions were automatically recorded using the MicroStation microplate reader (Biolog) with 590 nm and 750 nm wavelength filters.

Statistical Analysis

All analyses were carried out in triplicates. Metabolic fingerprints of *Weissella* strains as determined by the Biolog system were subjected to permutation analyses using the PermutMatrix as described by Fanelli et al. (2020).

RESULTS

In silico Analysis of *Weissella* Genomes

General Features of *Weissella* Genomes

The genomes of the six *Weissella* type strains were sequenced and assembled using the SPAdes software (version 3.10.1), obtaining a total of 76, 33, 39, 40, 106, and 26 contigs (>500 bp) for *W. beninensis* LMG 25373^T, *W. diestrammenae* DSM 27840^T, *W. fabalis* LMG 26217^T, *W. fabaria* LMG 24289^T, *W. ghanensis* DSM 19935^T, and *W. uvarum* B18NM42^T, respectively. The mol% GC content ranged from 35.45% (*W. beninensis* LMG 25373^T) to 39.53% (*W. ghanensis* DSM 19935^T), while the assembly length varied from 1,639 Mbp (*W. diestrammenae* DSM 27840^T) to 2,014 Mbp (*W. ghanensis* DSM 19935^T) (**Table 1**). The quality of the assemblies was excellent for all genomes, except *W. fabaria* LMG 24289^T, for which the assembly quality score was 76.5 (high). The contig N50 values were 57, 163, 166, 143, 54, and 246 kbp for *W. beninensis* LMG 25373^T, *W. diestrammenae* DSM 27840^T, *W. fabaria* LMG 24289^T, *W. fabalis* LMG 26217^T, *W. ghanensis* DSM 19935^T, and *W. uvarum* B18NM42^T, respectively.

Phylogenetic Analysis

The 16S rRNA gene sequence-based phylogeny is shown in **Figure 1**. According to the reconstructed tree, *Weissella* species can be clustered into six different species groups. The first group is constituted by *W. thailandensis*, *W. bombi*, *W. paramesenteroides*, *W. hellenica*, and *W. sagaensis*; a second group is formed by *W. cibaria* (*W. kimchii*) and *W. confusa*, which have a 16S rRNA gene sequence identity of 99.3%. The same percentage is shared by *W. oryzae* and *W. muntiaci*, which can be placed in a third species group. *W. soli* occurs close to these couples in the phylogenetic tree and has the highest 16S rRNA gene sequence identity with *W. muntiaci* (97.35%). *W. ceti, W. halotolerans, W. minor, W. uvarum,* and *W. viridescens* are clustered in a fourth species group with an average 16S rRNA gene sequence identity of 95.1%. *W. diestrammenae*, *W. kandleri*, W. *coleoptorum*, and *W. koreensis* clustered in the fifth group with a shared average 16S rRNA gene sequence identity of 96.8%. *W. cryptocerci*, *W. beninensis*, *W. ghanensis*, *W. fabalis*, and *W. fabaria* are located outside this large clade and share an average sequence identity of 96%, with *W. fabaria* and *W. fabalis* clustering at a value of 99.41%.

The ANI analysis results are reported in **Supplementary Table S2**. The species sharing the highest ANI values are *W. hellenica* and *W. sagaensis* (89.73%), followed by *W. ghanensis* and *W. fabalis* (85.63%), *W. beninensis* and *W. kandleri* (68.08%), while *W. halotolerans* with either *W. beninensis* (67.29%) or *W. ghanensis* (67.01%) are the most divergent species within the genus.

The matrix of AAI values is shown in **Supplementary Table S3**. Confirming the ANI analysis results, the most closely related species sharing the highest AAI values were *W. hellenica* and *W. sagaensis* (92.75%), followed by *W. hellenica* and *W. bombi* (89.45%). The lowest AAI value was shared between *W. muntiaci* and *W. cryptocerci* (51.91%).

A genome-based phylogeny was inferred from RAxML analysis. The results are shown in **Figure 2.** Partially confirming the clustering obtained by 16S rRNA gene sequence phylogeny, six species groups could be identified and named as *W. beninensis, W. kandleri, W. oryzae*, *W. confusa, W. paramesenteroides*, and *W. halotolerans* species groups. The *W. beninensis* group includes *W. beninensis*, *W. cryptocerci*, *W. fabalis*, *W. fabaria*, and *W. ghanensis*. The *W. kandleri* group comprises *W. kandleri*, *W. diestrammenae*, *W. koreensis*, *W. soli*, and *W. coleoptorum*. The *W. oryzae* group includes *W. oryzae* and *W. muntiaci*. The fourth species group, named the *W. confusa* species group, comprises *W. confusa* and *W. cibaria*. The fifth species group, designated as *W. paramestenteroides* by Fusco et al. (2015), includes *W. paramesenteroides*, *W. hellenica*, and *W. thailandensis*, with the addition of the recently described *W. sagaenisis* (Li et al., 2020) and *W. bombi* species (Praet et al., 2015). The sixth species group, already designated as *W. halotolerans* by Fusco et al. (2015), includes *W. halotolerans*, *W. ceti*, *W. uvarum*, *W. minor,* and *W. viridescens*.

Genomic and Phenotypic Characterization

Carbohydrate Metabolism Comparative Analysis

A comparative analysis of CAZy families present in the genomes of *Weissella* spp. is shown in **Figure 3**. The largest families are represented by the glycosyl transferases (GTs), which catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds, followed by the glycosyl hydrolases (GHs). Polysaccharide lyases (PLs) and carbohydrate esterase (CE) are poorly represented, with CE9 and CE1 being the most frequently occurring. The highest number of CAZymes, with 61 annotated proteins, was identified in the genome of the *W. cibaria* type strain, while the lowest number was found in the *W. ceti* type strain, for which only 13 CAZymes could be identified. The *W. bombi* and *W. cibaria* type strains have the highest numbers of GHs (28), while the *W. muntiaci* and *W. cibaria* type strains have the highest number of GTs (29). **Figure 3** also shows the clustering of *Weissella* species obtained by analyzing the occurrence and distribution of CAZymes among species.

A detailed analysis of the enzymes involved in the metabolism of carbohydrates of the six strains sequenced in this study is reported in **Table 2** and **Supplementary Table S4**. Regarding the hydrolytic enzymes, 13 GH proteins were annotated in the *W. ghanensis* type strain, while 12 were found in the *W. beninensis* type strain and *W. diestrammenae* type strain each. One FAD-binding protein with an auxiliary activity family 4 domain that catalyzes the conversion of a wide range of phenolic compounds was identified in the *W. ghanensis* type strain. Polysaccharide lyase enzymes were retrieved in the genomic sequences of the *W. ghanensis*, *W. fabaria*, and *W. fabalis* type strains. One carbohydrate-binding module CBM34 zyme was retrieved in the genomic sequence of the alpha-glycosidase (GH13_20) of the *W. diestrammenae* type strain. All strains harbor one N-acetylglucosamine 6-phosphate deacetylase belonging to the carbohydrate esterase of family 9 (CE9), while in the *W. diestrammenae* type strain, two additional CEs were annotated. Glycosyl transferases were the class with the highest count, with

TABLE 1 | Genomic features of *Weissella* spp. strains.

Features	*W. beninensis* LMG 25373^T	*W. diestrammenae* DSM 27940^T	*W. fabalis* LMG 26217^T	*W. fabaria* LMG 24289^T	*W. ghanensis* DSM 19935^T	*W. uvarum* B18NM42^T
Genome size (bp)	1,831,593	1,639,510	1,958,377	1,898,123	2,013,964	1,682,559
GC (%)	35.45	39.22	36.40	38.14	39.53	39.84
Number of contigs	76	33	39	40	106	26
Completeness % (essential genes found)	100% (106/106)	99.1% (105/106)	100% (106/106)	100% (106/106)	100% (106/106)	99.1% (105/106)
Quality	90.5 (excellent)	89.6 (excellent)	81.0 (excellent)	76.5 (high)	81.0 (excellent)	80.01 (excellent)
Contig N50 (bp)	57,467	162,981	143,119	166,541	53,895	246,285
Genes (total)	1,867	1,662	1,962	1,920	2,043	1,731
Genes (coding)	1,745	1,551	1,877	1,840	1,968	1,659
Coding density	85.40	87.07	86.28	87.31	87.15	87.59
RNA	61	57	58	66	55	60
rRNAs	1, 1, 2 (5S, 16S, 23S)	1, 1 (16S, 23S)	2, 2, 2 (5S, 16S, 23S)	2, 1, 2 (5S, 16S, 23S)	1, 1, 1 (5S, 16S, 23S)	1, 2, 1 (5S, 16S, 23S)
tRNAs	53	52	49	58	49	53
ncRNAs	4	3	3	3	3	3
Pseudo genes (total)	61	14	27	14	20	12

FIGURE 1 | 16S rRNA gene-based phylogeny of *Weissella* species. Phylogeny was performed using the neighbor-joining method; phylogenetic robustness was inferred by a bootstrapping procedure from 500 replications to obtain the confidence value for the aligned sequence dataset. *Bifidobacterium bifidum* ATCC 29521^T was used as an outgroup. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. Scaled circles are representative of bootstrap values.

a total of 21 enzymes in the *W. ghanensis* type strain, 19 in the *W. fabaria* type strain, 16 in the *W. fabalis* type strain, 15 in the *W. diestrammenae* and *W. uvarum* type strains, and 14 in the *W. beninensis* type strain.

A comparative analysis of carbohydrate metabolism pathways is depicted in **Figure 4.** The analysis shows that the highest count of enzymes within these pathways was retrieved in the categories "amino sugar and nucleotide sugar metabolism," "pyruvate metabolism," "glycolysis and gluconeogenesis," and "starch and glucose metabolism." The *W. cryptocerci* type strain has the highest enzyme count (172 in total), while the lowest occurred in the *W. kandleri* type strain (111).

A comparative analysis of the SEED subsystem is shown in **Figure 5**. The highest number of features was determined in the subsystem "protein metabolism," followed by "carbohydrates and nucleosides and nucleotides." The *W. cibaria*, *W. cryptocerci*, and *W. jogaejeotgali* type strains showed the highest total feature counts.

Substrate Consumption by *Weissella* Strains

Carbon source consumption was evaluated using Biolog AN plates, and the results are presented in **Table 3**. All *Weissella* strains used N-acetyl-D-glucosamine, D-fructose, α-D-glucose, D-mannose, palatinose, turanose, and maltotriose.

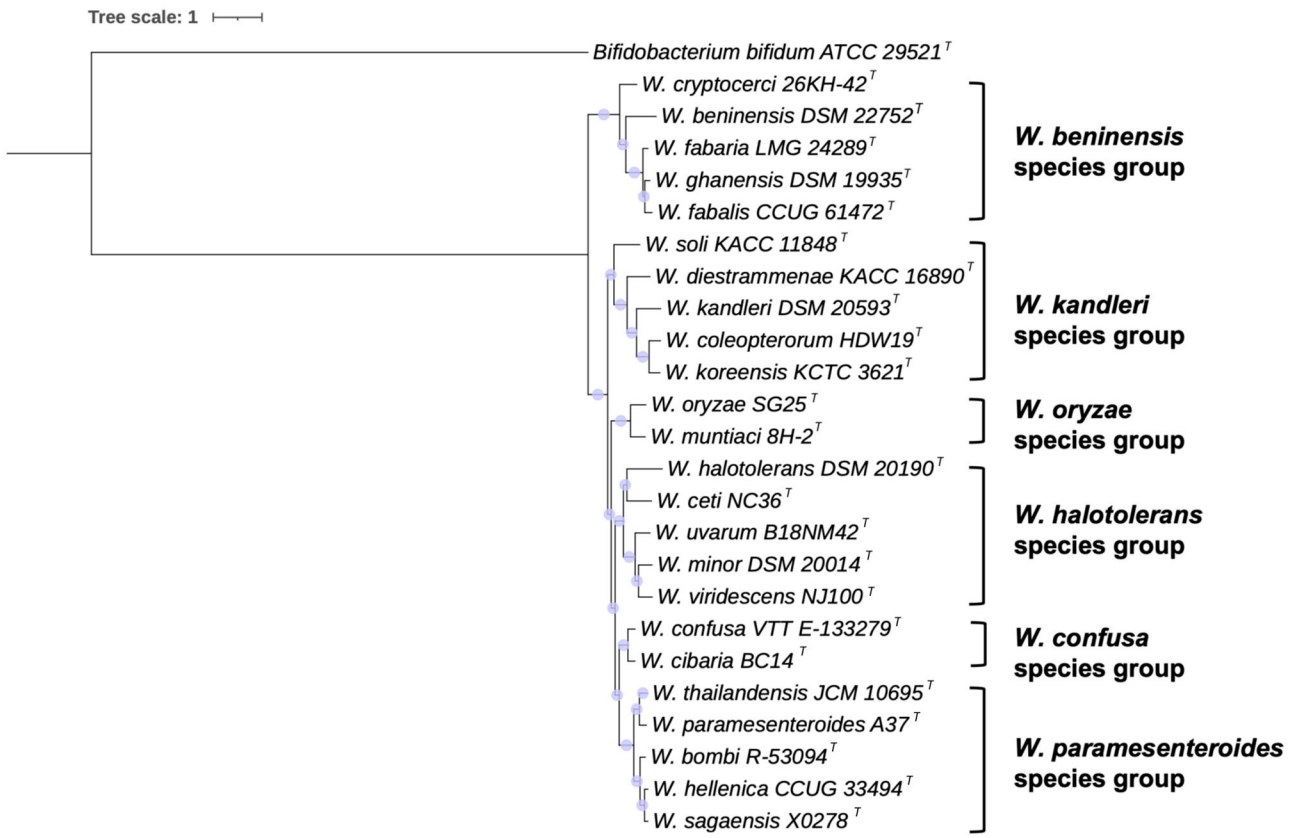


FIGURE 2 | Genome-based phylogeny of *Weissella* species. The tree was inferred by using the maximum likelihood method RAxML with progressive refinements. *Bifidobacterium bifidum* ATCC 29521^T was used as an outgroup. The tree is drawn to scale. Support values are represented by scaled circles at each node.

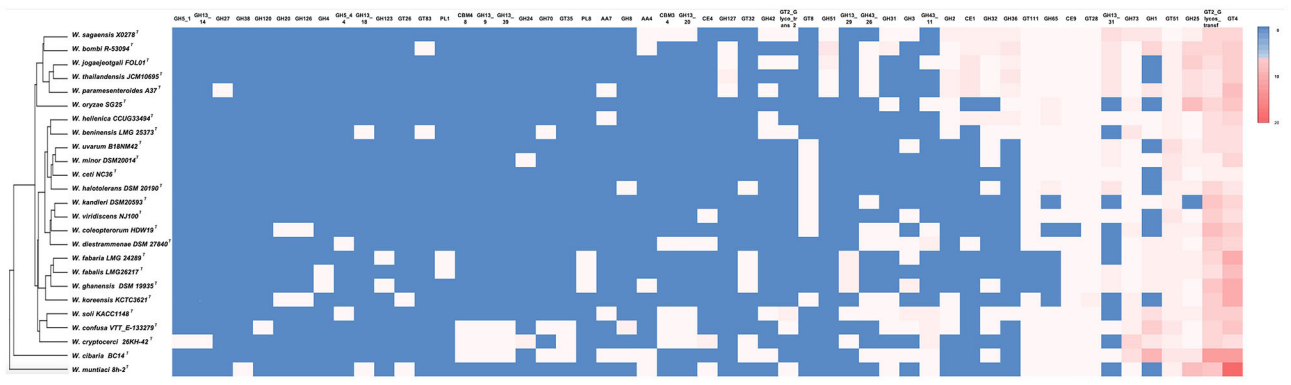


FIGURE 3 | Comparative analysis of CAZymes in *Weissella* species. The heatmap was manually constructed based on CAZymes count in each species and visualized with the average linkage as a clustering method and the Euclidean distance measurement method, providing the resulting dendrogram.

Furthermore, according to the comparative analysis of carbohydrate metabolism pathways and the cluster analysis based on carbon source oxidation (**Supplementary Figure S1**), the cluster including *W. beninensis*, *W. fabalis*, *W. fabaria*, and *W. ghanensis* type strains was characterized by the consumption of glycyl-L-methionine, α-ketobutyric acid, and pyruvic acids. Only *W. fabalis* and *W. fabaria* consumed i-erythritol, D-trehalose, and D-malic acid (**Table 3**). Moreover, formic acid was used only by the *W. fabaria* type strain, while α-cyclodextrin, fumaric acid, and glycyl-L-glutamine were utilized by the *W. fabalis* type strain. Among all the strains tested, the *W. beninensis* type strain showed the widest consumption of carbon sources.

TABLE 2 | Number of CAZymes present in sequenced *Weissella* strains.

	W. beninensis LMG 25373^T	*W. fabalis* LMG 26217^T	*W. uvarum* B18NM42^T	*W. fabaria* LMG 24289^T	*W. diestrammenae* DSM 27840^T	*W. ghanensis* DSM 19935^T
Auxiliary Activities (AAs)						
AA4	-	-	-	-	-	1
Carbohydrate Binding-Module (CBMs)						
CBM34	-	-	-	-	1	-
Carbohydrate Esterases (CEs)						
CE1	-	-	-	-	1	-
CE4	-	-	-	-	1	-
CE9	1	1	1	1	1	1
Glycoside Hydrolases (GHs)						
GH1	1	2	-	2	2	2
GH2	1	-	-	-	-	-
GH3	-	-	1	1	1	2
GH4	-	1	-	-	-	1
GH5_44	-	-	-	-	1	-
GH13_18	1	-	-	-	-	-
GH13_20	-	-	-	-	1	-
GH13_29	-	2	-	2	-	2
GH13_31	-	2	2	-	-	2
GH25	1	2	2	2	1	1
GH31	-	-	-	-	1	-
GH32	1	-	1	-	-	1
GH36	1	-	-	-	-	-
GH42	1	-	-	-	-	-
GH43_11	-	-	-	-	2	-
GH43_26	-	-	-	-	1	-
GH65	1	-	1	-	1	-
GH70	1	-	-	-	-	-
GH73	3	1	1	1	1	1
GH123	-	-	-	1	-	1
Glycosyl Transferases (GTs)						
GT2_Glyco_trans_2_3	1	-	-	-	-	-
GT2_Glycos_transf_2	4	4	4	5	6	7
GT4	4	8	4	10	5	10
GT8	-	-	1	-	-	-
GT28	1	1	1	1	1	1
GT32	-	1	-	1	-	1
GT51	2	2	4	2	2	2
GT83	1	-	-	-	-	-
GT111	1	-	1	-	1	-
Polysaccharide Lyases (PLs)						
PL1	-	1	-	1	-	-
PL8	-	1	-	1	-	1

In fact, only this strain was capable of using D-galactose, α-D-lactose, lactulose, D-melibiose, β-methyl-D-galactoside, D-raffinose, sucrose, pyruvic acid methyl ester, and uridine-5'-monophosphate. Conversely, D-arabitol and D-sorbitol were exclusively consumed by the *W. uvarum* type strain (**Table 3**).

DISCUSSION

Based on the results of ANI, AAI, 16S rRNA gene sequence identity, and 16S rRNA gene- and genome-based phylogeny, all *Weissella*-type strains investigated could be confirmed as representing individual and separate species. Based on our

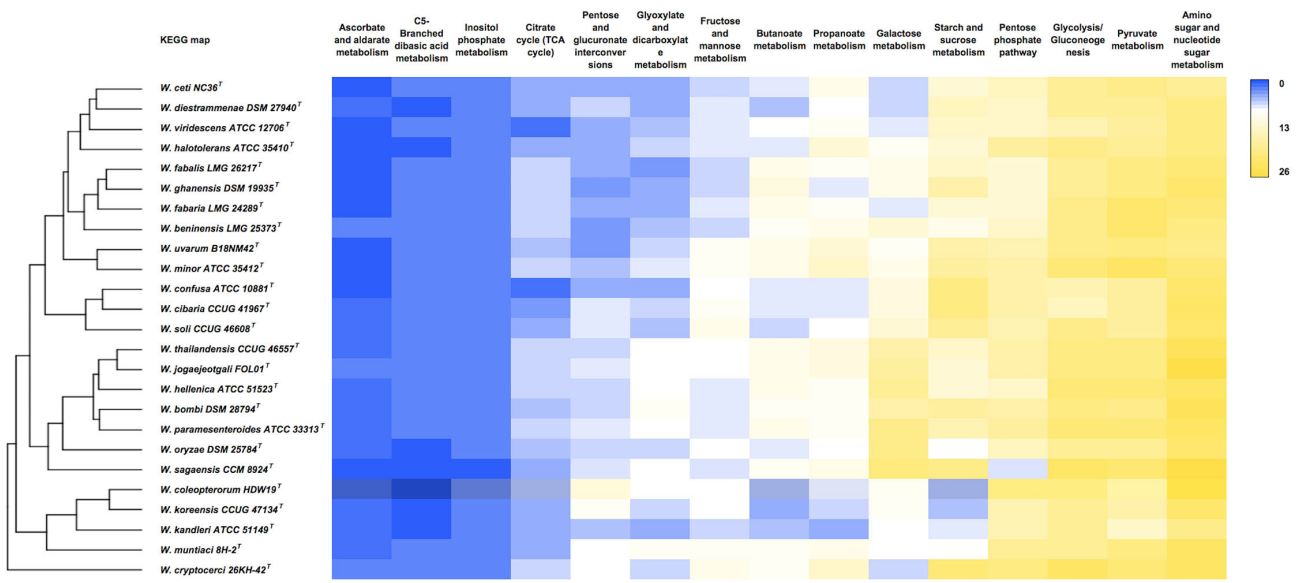


FIGURE 4 | Comparative analysis of carbohydrate metabolism in *Weissella* species. The heatmap was manually constructed based on the number of proteins associated with each KEGG pathway in each genome and visualized with the average linkage as a clustering method and the Euclidean distance measurement method, providing the resulting dendrogram.

ANI/AAI results, we confirmed that *W. jogaejeotgali,* first described by Lee et al. (2015) as a novel species, could be classified as a later heterotypic synonym of *W. thailandensis*. Furthermore, we also confirmed that *W. kimchii*, described as a novel species by Choi et al. (2002), is a later heterotypic synonym of *W. cibaria*.

Our data partially validate the taxonomic clustering previously proposed by De Bruyne et al. (2010) and Fusco et al. (2015), which were essentially based on the 16S rRNA gene and *pheS* gene sequence phylogenies, with the latter showing a higher discriminatory power as a marker gene. The description of novel species (Lee et al., 2015; Praet et al., 2015; Heo et al., 2019; Li et al., 2020; Lin et al., 2020; Hyun et al., 2021; this study) and the availability of additional genomic sequences (refer to **Supplementary Table S1**) allowed us to perform a comprehensive genomic-based phylogenetic analysis of all species in the genus *Weissella.* Our analysis also integrates genomic indexes, which are currently considered as minimal standards to perform taxonomic classification. Therefore, based on our phylogenomic reconstruction, we were able to define six distinct species groups within the genus *Weissella*, i.e., the *W. beninensis, W. kandleri, W. confusa, W. halotolerans, W. oryzae*, and *W. paramesenteroides* species groups.

The first divergent line within the genus is constituted by the *W. beninensis* group, which includes *W. beninensis*, *W. cryptocerci*, *W. fabalis*, *W. fabaria*, and *W. ghanensis*. In this case, the clustering recently updated by Fusco et al. (2015) is confirmed by the genome-based phylogeny and by genomic analysis: the average ANI value among these species is 81.2%, with *W. ghanensis* and *W. fabaria* being the most closely related, with an ANI value of 85.63%. Species within this group share 96% of 16S rRNA gene sequence identity; the percentage rises to 99.5% among *W. fabaria*, *W. fabalis*, and *W. ghanensis*. The second species group, designated as the *W. kandleri* group, comprises *W. kandleri*, *W. diestrammenae*, *W. koreensis*, *W. soli*, and *W. coleoptorum*. The latter was described in 2021 (Hyun et al., 2021) and was therefore not represented in the previous classification. The 16S rRNA gene sequence identity among these species is 97.1% on average, while the ANI value is 82%, with *W. coleoptorum* and *W. soli* being the closest related, sharing an ANI value of 83.91%. In contrast to the results previously presented by Fusco et al. (2015), *W. oryzae* was not included in this group, as the genomic analysis performed in this study placed it into a separate branch together with *W. muntiaci*. These two species share 99.32% of 16S rRNA gene sequence identity, 80.15% of ANI, and 75.64 of AAI, and can be assigned to a third species group, namely the *W. oryzae* species group. The fourth species group, namely the *W. confusa* species group, comprises *W. confusa* and *W. cibaria*, confirming the clustering previously proposed by Fusco et al. (2015). Based on the results of this study, we added the recently described *W. sagaenisis* (Li et al., 2020) and *W. bombi* species (Praet et al., 2015) to the fifth species group, designated by Fusco et al. (2015) as *W. paramestenteroides,* which also includes *W. paramesenteroides*, *W. hellenica*, and *W. thailandensis*. The average ANI value shared among these species is 82.4%, while 16S rRNA gene sequence identity is 98%. Notably, *W. sagaenisis* and *W. hellenica* 16S rRNA gene sequences are identical, while their shared ANI value is 89.7%, thus confirming the correct designation of *W. sagaenisis* as a novel species. In accordance with Fusco et al. (2015), the sixth species group is the *W. halotolerans* group, which includes *W. halotolerans*, *W. ceti*, *W. uvarum*, *W. minor,* and *W. viridescens*. Within these species, the 16S rRNA gene sequence identity reaches 95.1%, with the

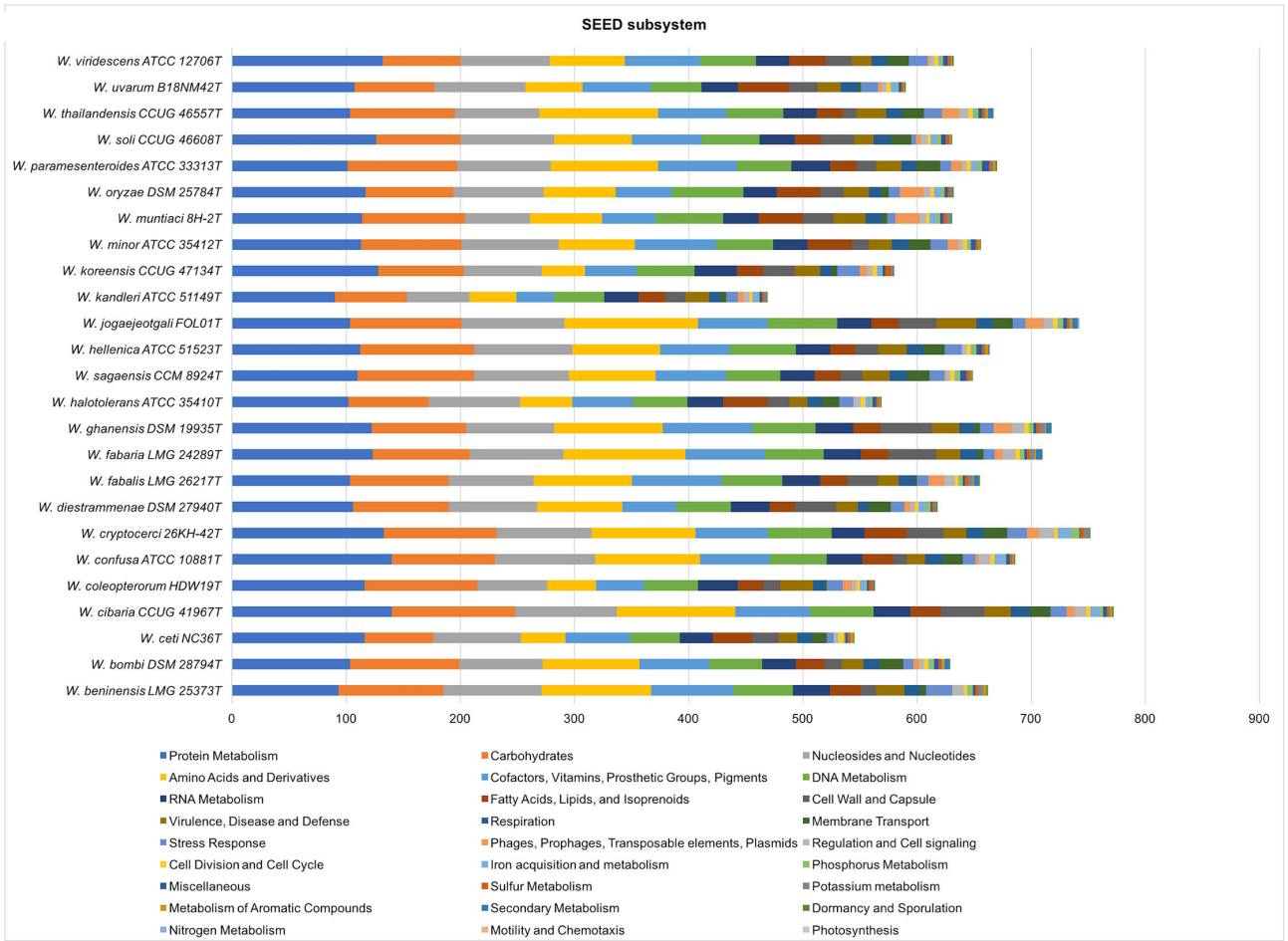


FIGURE 5 | Comparative analysis of SEED subsystem features in *Weissella* species. Genes annotated by RAST were assigned to functional categories and grouped into subsystems. Colored bars indicate the number of genes assigned to each category.

maximum value of 99.14% occurring between *W. minor* and *W. uvarum*, while the shared ANI value for this group is 79.06%.

According to CAZy and KEGG pathway analyses, species within this genus harbor a comprehensive carbohydrate utilization system, including sugar uptake, transporters, and metabolism-related genes, which confer them strong carbohydrate utilization capabilities, as shown by different carbohydrate utilization profiles. The carbohydrate metabolism pathway clustering follows the phylogenomic species group clustering in many cases: this occurred for *W. beninensis*, *W. fabalis*, *W. fabaria*, and *W. ghanensis;* for *W. minor* and *W. uvarum*; for *W. ceti* and *W. halotolerans*; for *W. bombi* and *W. sagaensis*; and for *W. thailandensis, W. paramesenteroides,* and *W. jogaejeotgali*.

The *W. fabalis*, *W. ghanensis*, and *W. uvarum* type strains were found to harbor the genetic determinants for the synthesis of oligo-1,6-glucosidases, which hydrolyze the α-1,6 linkage in starch, glycogen, and the derived oligosaccharides to produce sugars with an α-configuration. The *W. diestrammenae* type strain was found to harbor a gene coding for a maltogenic amylase, an enzyme that favors starch degradation in maltose. Therefore, these *Weissellas* may play an important role in the fermentation of sourdoughs due to the metabolism of starch and amylopectin, which are composed of α-1,4 glucose main chain and α-1,6 glucose side chain (van der Maarel et al., 2002; Wang et al., 2021).

All strains were able to use the D-mannose sugar. This sugar is a dextrorotatory hexose aldehyde/aldose monosaccharide found in certain bacteria, fungi, and plants, and is rarely present in nature as a free monosaccharide. Nevertheless, it is a constituent of numerous simple and complex polysaccharides and is mostly found in nature as a component of mannan, hemicellulose, and cellulose in dietary fiber (Hu et al., 2016). For example, it constitutes the basic molecule of mannans, the reserve polysaccharides of some plant species (e.g., palm), or is associated with galactose (mannogalactans) to form gummy mucilages that protect the seeds of some plants (e.g., carob); they are widely used as stabilizers of food products such as ice cream and mayonnaise. The *W. diestrammenae* type strain was isolated from the gut of a camel cricket. The remaining 5 type strains sequenced

TABLE 3 | Consumption of carbon sources by *Weissella* strains evaluated after anaerobic incubation of Biolog AN MicroPlate (Biolog, USA).

	W. beninensis **LMG 25373^T**	***W. diestrammenae*** **DSM 27840^T**	***W. fabalis*** **LMG 26217^T**	***W. ghanensis*** **DSM 19935^T**	***W. uvarum*** **B18NM42^T**	***W. fabaria*** **LMG 24289^T**
N-Acetyl-D-Glucosamine	+[a]	+	+	+	+	+
N-Acetyl-β-DMannosamine	–[b]	+	+	+	+	+
Amygdalin	–	–	+	+	–	–
D-Arabitol	–	–	–	–	+	–
D-Cellobiose	+	–	+	+	+	+
α-Cyclodextrin	–	–	+	–	–	–
Dextrin	+	–	+	+	+	+
i-Erythritol	–	–	+	–	–	+
D-Fructose	+	+	+	+	+	+
L-Fucose	+	–	–	+	–	–
D-Galactose	+	–	–	–	–	–
Gentiobiose	–	–	+	+	–	–
D-Gluconic Acid	+	+	–	+	+	–
α-D-Glucose	+	+	+	+	+	+
α-D-Glucose-1-Phosphate	–	+	–	–	+	–
D-Glucose-6-Phosphate	–	+	–	–	–	–
Glycerol	–	–	–	+	+	+
α-D-Lactose	+	–	–	–	–	–
Lactulose	+	–	–	–	–	–
Maltose	–	+	–	+	–	–
Maltotriose	+	+	+	+	+	+
D-Mannitol	+	–	+	–	+	+
D-Mannose	+	+	+	+	+	+
D-Melibiose	+	–	–	–	–	–
3-Methyl-D-Glucose	+	–	+	–	+	+
β-Methyl-D-Galactoside	+	–	–	–	–	–
β-Methyl-D-Glucoside	–	–	+	+	–	–
Palatinose	+	+	+	+	+	+
D-Raffinose	+	–	–	–	–	–
L-Rhamnose	+	–	+	+	–	–
Salicin	–	–	+	+	–	–
D-Sorbitol	–	–	–	–	+	–
Sucrose	+	–	–	–	–	–
D-Trehalose	–	–	+	–	–	+
Turanose	+	+	+	+	+	+
Acetic Acid	–	–	–	–	–	+
Formic Acid	–	–	–	–	–	+
Fumaric Acid	–	–	+	–	–	–
Glyoxylic Acid	–	–	+	+	–	+
α-Hydroxybutyric Acid	+	+	–	–	+	+
α-Ketobutyric Acid	+	+	+	+	–	+
α-Ketovaleric Acid	–	+	+	+	–	+
D,L-Lactic Acid	+	+	+	+	+	–
L-Lactic Acid	+	+	+	+	–	–
D-Lactic Acid Methyl Ester	+	+	+	–	+	+
D-Malic Acid	–	–	+	–	–	+
L-Malic Acid	+	+	+	+	–	–
Propionic Acid	–	–	–	+	–	+
Pyruvic Acid	+	–	+	+	+	+

(Continued)

TABLE 3 | Continued

	W. beninensis LMG 25373^T	*W. diestrammenae* DSM 27840^T	*W. fabalis* LMG 26217^T	*W. ghanensis* DSM 19935^T	*W. uvarum* B18NM42^T	*W. fabaria* LMG 24289^T
Pyruvic Acid Methyl Ester	+	−	−	−	−	−
Glycyl-L-Glutamine	−	−	+	−	−	−
Glycyl-L-Methionine	+	−	+	+	−	+
L-Methionine	−	−	+	+	−	−
Inosine	+	+	−	−	+	−
Uridine	+	+	−	−	+	−
Uridine-5′-Mono-phosphate	+	−	−	−	−	−

The Weisella strains resulted all negative for N-acetyl-D-galactosamine, adonitol, arbutin, β-cyclodextrin, dulcitol, D-galacturonic acid, D-glucosaminic acid, D,L-α-glycerol phosphate, m-inositol, D-melezitose, α-methyl-D-galactoside, α-methyl-D-glucoside, stachyose, β-hydroxybutyric acid, itaconic acid, D-saccharic acid, succinamic acid, succinic acid, succinic acid mono-methyl ester, m-tartaric acid, urocanic acid, alaninamide, L-alanine, L-alanyl-L-glutamine, L-alanyl-L-histidine, L-alanyl-L-threonine, L-asparagine, L-glutamic acid, L-glutamine, glycyl-L-aspartic acid, glycyl-L-proline, L-phenylalanine, L-serine, L-threonine, L-valine, L-valine plus L-aspartic acid, 2′-deoxy adenosine, thymidine, thymidine-5′-mono-phosphate.
[a] +, utilization/oxidation of the carbon source; [b] -, not utilization/oxidation of the carbon source.

in this study originated from vegetables such as cassava, cocoa, and grapes, all sources containing mannose and/or its (oligo)polymers. Due to the described enzymatic activities, these *Weissellas* could be employed as starters for the fermentation of such mannose- and oligo mannose-containing vegetables. Their ability to metabolize mannose is corroborated by the fact that in every sequenced genome we retrieved (i) the *man*Xa transporter, which can transfer phosphorus-containing groups to D-mannose, and (ii) the mannose-6-phosphate isomerase [EC:5.3.1.8] for the conversion of D-mannose-6-phosphate to β-D-fructose 6-phosphate. The mannose transportation operon *man*XYZ (Jeckelmann and Erni, 2020) was identified in all the sequenced strains.

Except for *W. diestrammenae*, all the other strains possess an *scr*K gene, coding for a fructokinase. It catalyzes the transfer of phosphate to D-fructose, which is then converted to α-D-glucose 6-phosphate by the glucose-6-phosphate isomerase. In the *W. diestrammaneae* type strain, fructose utilization might proceed through its conversion to α-D-glucose, catalyzed by the xylose isomerase, only harbored by this strain (*xyl*A gene, KAR27_02655). In the *W. uvarum* and *W. fabaria* type strains, the *fru*AB gene, coding for the D-fructose phosphotransferase, was identified. Only in the *W. uvarum* type strain can D-fructose enter glycolysis through the subsequent action of the *fru*K fructokinase and the triosephosphate isomerase, which converts the β-D-fructose 1,6-bisphosphate to glyceraldehyde 3-phosphate.

According to the Biolog data, *W. beninensis* is the only species able to metabolize D-galactose and raffinose. In fact, the *gal*A gene was detected only in the genomic sequence of the type strain of this species, in addition to two copies of the β-galactosidase (KAK10_01750 and KAK10_05755). The *gal*A codes for the α-galactosidase [EC:3.2.1.22] (KAK10_01740), the exoglycosidase that hydrolyzes α-1,6 galactoside linkages found in sugars, such as raffinose, melibiose, and stachyose, and branched polysaccharides such as galactomannans and galacto-glucomannans. The ability to metabolize nondigestible sugars such as oligosaccharides of the raffinose family and galactomannans is common for several probiotic bacteria (Zartl et al., 2018). The α-galactosidase enzyme is well explored in the food industry, where it is used for removing raffinose family oligosaccharides (RFOs) in soymilk and sugar crystallization processes, and for the improvement of animal feed quality and biomass processing (Bhatia et al., 2020).

Only *W. beninensis* can consume D-sucrose. This is due to the presence of (i) a PTS transporter subunit EIIC (KAK10_07455), which transports sucrose into the cell, converting it to sucrose-6-phosphate; (ii) a sucrose-6-phosphate hydrolase (KAK10_07450), which catalyzes the formation of fructose and glucose-6-phosphate; and (iii) a sucrose phosphorylase (KAK10_07625), which reconverts the sucrose-6-phosphate into sucrose. Moreover, a preliminary evaluation of EPS production (data not shown) indicates that *W. beninensis* can produce ropiness when incubated in MRS added with 20 g/L of sucrose. *W. beninensis* was isolated from spontaneous fermentation of cassava, and it was previously characterized for the acid production from D-fructose, D-galactose, D-glucose, lactose, maltose, D-mannose, melibiose, D-raffinose, sucrose, *N*-acetylglucosamine, and D-mannitol (Padonou et al., 2010). Cassava usually contains a low amount of free sugar, which differs according to the variety and the tissue age in the storage root (Junior and Campos, 2004). Its spontaneous fermentation is accomplished by the presence of different species of yeasts and lactic acid bacteria (Padonou et al., 2009) and activates mutually stimulating interactions, as widely characterized in other products such as sourdough and kefir (De Vuyst and Neysens, 2005; Tofalo et al., 2020). The ability of *W. beninensis* to use numerous carbon sources may indicate its adaptation to this complex microbial and metabolic environment.

Trehalose, or α-D-glucopyranosyl-1,1-α-D-glucopyranoside, is a disaccharide made up of two α-D-glucose molecules joined by an α-1,1 glycosidic bond. It naturally occurs in bacteria, fungi, yeasts, algae, plants, and invertebrates, including insects, but it is absent in vertebrates. Its major dietary sources are mushrooms; in fact, it is contained in most edible fungi and is an important part of reconstituting dried shiitake mushrooms. For this reason, it is also referred to as a mushroom sugar, while it is called seaweed sugar in China since trehalose is contained in marine

plants such as "hijiki" seaweed. *Weissellas* that can metabolize this sugar may play a pivotal role in the fermentation and digestion of the above-mentioned (novel) foods. To metabolize trehalose, this disaccharide enters the cells throughout the action of the trehalose PTS permease (EC: 2.7.1.201). Consequently, the α,α-phosphotrehalase catalyzes the hydrolysis of α,α-trehalose 6-phosphate into D-glucose and D-glucose 6-phosphate. This enzyme was annotated in the genomic sequences of the *W. fabalis*, *W. fabaria*, and *W. ghanensis* type strains to occur in a genomic cluster, which shows differences in the organization among these species (**Figure 6**), while the corresponding genes in the *W. ghanensis* type strain were annotated in two separate contigs. All these 3 species share the same ecological niche of origin (*W. ghanensis* type strain was isolated from Ghanaian cocoa fermentation; *W. fabaria* type strain was isolated from fermented cocoa bean heaps; and *W. fabalis* was isolated from spontaneous cocoa bean fermentation). Common to all 3, within the cluster, there are (1) a gene coding for a hypothetical protein with a GH (family 13) catalytic domain, (2) the *tre*R gene (coding for the trehalose operon repressor), (3) the *tre*C gene (coding for the α,α-phosphotrehalase), (4) the genes coding for TreB (the trehalose transporter with the two domain phosphotransferases (PTS) subunit EIIC and IIA), (5) one flippase gene, (6) one gene coding for an EpsG family protein, (7) one for a sugar transferase, and (8) several genes coding for glycosyl transferases. In the *W. fabaria* type strain, this locus also comprises a sequence of 14 kbp downstream of *tre*R, where there are annotated rhamnose operon *rbf*ABCD and genes coding for (1) a dTMP kinase, (2) a LicD family protein, (3) a DUF1972 domain-containing protein, (4) a CpsD/CapB family tyrosine-protein kinase, (5) a capsular biosynthesis protein, (6) an LCP family protein, and (7) a tyrosine phosphatase. The trehalose operon has not yet been described in the *Weissella* species. In addition to the *tre*R and *tre*B genes, *Lactococcus lactis* includes the phosphomannomutase gene *fem*B, the trehalose/maltose hydrolase gene *tre*PP, and the β-phosphoglucomutase *pgm*B gene (Andersson et al., 2005) (**Figure 6**). The *pgm*B gene was instead annotated in *W. diestrammenae* (KAR27_07555), within a locus comprising genes coding for the transcriptional regulator LacI, an α-glycosidase, a glycoside hydrolase, a galactose maturotase, and sugar transporters.

The maltose operon was first described in *L. lactis* by Andersson and Rådström (2002). Maltose is transported by an ATP-dependent permease system, and then it is degraded by the concerted action of a Pi-dependent maltose phosphorylase (MP) and the β-phosphoglucomutase (β-PGM), which are located in two different operons. In the *Weissella* species analyzed in this study, *pgm*B occurred in the *W. diestrammenae* (KAR27_07555), *W. beninensis* (KAK10_00105), and *W. uvarum* (KAR63_02130) type strains. The *map*A gene, which codes for the maltose phosphorylase [EC:2.4.1.8] that catalyzes the reaction between maltose and phosphate to form D-glucose and β-D-glucose-1-phosphate, was found to be present in the *W. uvarum* (KAR63_02125), *W. beninensis* (KAK10_00115), and *W. diestrammenae* (KAR27_07545) type strains. In contrast to *L. lactis*, the *mal*P and *pgm*B genes are located closely on the genomic sequences of the species in which they were annotated. In the *W. uvarum* type strain, this operon also contains genes coding for the glucose transport system, formed by the PTS transporter subunit EIIC and PTS glucose transporter subunit IIA, and a trehalose operon repressor. However, the sequences surrounding this genomic area are different from the trehalose operon as above described in the other *Weissella* species, indicating a divergent evolutionary pattern. In the *pgm*B genomic locus of the *W. diestrammenae* type strain, there are also genes coding for a galactose mutarotase, a glycosyl

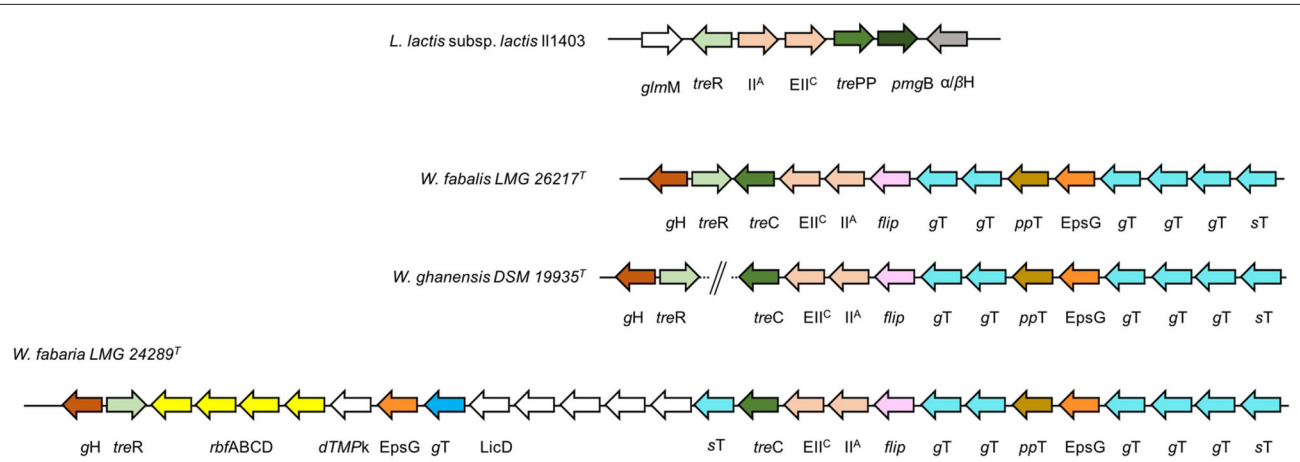


FIGURE 6 | Trehalose operon in *Weissella* species. Genomic organization of the putative trehalose operon in the *Weissella* species. Gene clustering is represented by the arrows superposed on the black horizontal line. Gene and intergenic spaces are not drawn in scale. *Lactococcus lactis* subsp. *lactis* II1403 (GenBank accession no. NC_002662): *glm*M, phosphoglucosamine mutase; *tre*R, trehalose operon repressor; IIA, PTS glucose transporter subunit IIA; EIIc, PTS transporter subunit EIIC; *pmg*B, beta-phosphoglucomutase; α/βH, alpha/beta hydrolase; gH, glycosyl-hydrolase; polysaccharide biosynthesis C-terminal domain-containing protein flip, flippase; *g*T, glycosyltransferase; *pp*T, polysaccharide pyruvyl transferase family protein; EpsG, EpsG family protein; *s*T, sugar transferase; *rbf*ABCD, rhamnose operon; dTMPk, dTMP kinase, LicD, LicD family protein.

hydrolase, the α-glycosidase (KAR27_07540) gene, a gene coding for an MFS transporter, one for a LacI family DNA-binding transcriptional regulator, one for an ECF transporter, one for an ABC transporter substrate-binding protein, and a permease gene. The *pgm*B gene codes for a β-phosphoglucomutase, which catalyzes the interconversion of D-glucose 1-phosphate (G1P) and D-glucose-6-phosphate (G6P), forming β-D-glucose 1,6-(bis)phosphate (β-G16P) as an intermediate. It is the catabolic activity in the maltose and trehalose pathways. Although *pgm*B has not yet been characterized in the *Weissella* species, this gene was reported as essential in the trehalose pathway of *L. lactis* (Andersson et al., 2001). The occurrence of this gene in the *W. diestrammenae* type strain, which, according to the carbon source consumption analysis, is not able to metabolize trehalose, suggests its involvement in the maltose metabolism. This sugar is indeed consumed by this strain, most likely through the subsequent reactions catalyzed by the maltose phosphorylase MapA and PgmB. In the *W. beninensis* type strain, the putative maltose operon was annotated in a very short contig; therefore, it was not possible to reconstruct the surrounding organization of the putative operon, although there are annotated genes coding for a galactose mutarotase, a glycosyl hydrolase, one MFS transporter, and two LacI family DNA-binding transcriptional regulators. In the W. *fabalis, W. uvarum*, and *W. ghanensis* type strains, the ability to hydrolyze maltose can also be associated with the presence of the maltase-glucoamylase (E. 3.2.10) that catalyzes maltose hydrolysis to produce two molecules of D-glucose.

Furthermore, α-glucosidase genes were retrieved in the *W. fabalis*, *W. ghanensis*, and *W. uvarum* type strains. The substrates for this hydrolytic enzyme are maltooligosaccharides, phenyl α-maltoside, nigerose, soluble starch, amylose, amylopectin, and β-limit dextrins (Tomasik and Horton, 2012). In a recent study by Wangpaiboon et al. (2021), this enzyme was characterized in *W. cibaria* as acting on short-chain maltooligosaccharides. Genes encoding 6-phospho-β-glucosidase enzymes were annotated in the *W. diestrammenae*, *W. fabalis*, *W. fabaria*, and *W. ghanensis* type strains. This enzyme is involved in the hydrolysis of phosphorylated disaccharides and usually does not have hydrolytic activity toward nonphosphorylated substrates. The resulting glucose and glucose 6-phosphate are further metabolized by the glycolytic pathway. β-D-glucosidase activity is widespread among lactic acid bacteria. It was suggested to play a role in the interaction with the human host. Furthermore, it is relevant for food fermentation processes and is involved in the release of β-glucosidases from their β-D-glucosylated precursors in several plant secondary metabolites (Michlmayr and Kneifel, 2014).

In *W. beninensis,* only one GH1 was annotated, in the putative ribose operon, comprising sugar transporters, and the *rbsC* and *rbs*D genes, coding for the ribose permease and the D-ribose pyranase, respectively.

Pectin lyases were identified in the genomic sequences of the *W. fabalis*, *W. fabaria*, and *W. ghanensis* type strains. These enzymes catalyze pectin degradation *via* eliminative cleavage of the α-(1,4) glycosidic linkages in homogalacturonan. The capability of these species to hydrolyze pectin suggested by the presence of these enzymes could be relevant for the exploitation in the food industry; pectin lyases are, in fact, usually employed in wine and juice prepress maceration, as well as in juice clarification methods (Mantovani et al., 2005; Kassara et al., 2019).

CONCLUSION

In this study, for the first time, we sequenced and analyzed the genomes of the type strains of 6 *Weissella* species discovered in the past decade, whose assemblies were not yet available at the time this study was initiated. We also performed a comprehensive phylogenomic analysis on all the *Weissella* species to date described and reassessed the phylogenetic structure of strains in the *Weissella* genus using 16S rRNA gene sequence and genome-based phylogeny. Updating the previous clustering proposed by other authors, we were able to identify six distinct species groups within the genus, namely, *W. beninensis, W. kandleri, W. confusa, W. halotolerans, W. oryzae*, and *W. paramesenteroides* species groups.

Moreover, we investigated the capability of the 6 type strains to metabolize 95 carbohydrates, demonstrating the strong carbohydrate utilization capabilities of the sequenced strains. This ability was also confirmed by the identification of the genetic determinants of the enzymes involved in carbohydrate metabolism. The genomic and phenotypic analyses provided further knowledge about the ability of the *W. beninensis*, *W. ghanensis*, *W. fabaria*, *W. fabalis*, *W. uvarum,* and *W. diestrammenae* type strains to metabolize certain carbohydrates and to detect their genetic determinants. In fact, the studies reporting the discovery of these type strains (De Bruyne et al., 2008, 2010; Padonou et al., 2010; Oh et al., 2013; Snauwaert et al., 2013; Nisiotou et al., 2014) provided data about the utilization of a maximum of 19 carbohydrates against the 95 carbohydrates tested herein by using the Biolog. Furthermore, the permutation analysis of the Biolog data confirmed the interspecific metabolic diversity of the analyzed type strains.

The increasing availability of the genomic sequences of the *Weissella* species will contribute to improving the knowledge about this genus and identifying the features defining its role in fermentative processes and its biotechnological potential.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

VF conceived the work and interpreted the data. G-SC performed the genomic sequencing. VF and FF organized and performed the bioinformatic work. DC checked the purity of the strains and prepared the working cultures for DNA extraction and phenotypic characterization. MM performed the phenotypic

characterization. VF, FF, and MM wrote the manuscript. All authors contributed to the revision of the manuscript, read, and approved the submitted version.

ACKNOWLEDGMENTS

The authors acknowledge Sebastien Santini (CNRS/AMU IGS UMR7256) and the PACA Bioinfo Platform (supported by IBISA) for the availability and management of the phylogeny.fr website used to analyze phylogenetic relationships between 16S rRNA gene sequences. They acknowledge Dr. Anastasia Nisiotou for kindly providing the type strain of *W. uvarum*.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.914036/full#supplementary-material

Supplementary Table S1 | List of *Weissella* strains used in this study.

Supplementary Table S2 | ANI values among *Weissella* strains.

Supplementary Table S3 | AAI values among *Weissella* strains.

Supplementary Table S4 | CAZymes present in sequenced *Weissella* species.

Supplementary Figure S1 | Cluster analysis of *Weissella* strains based on carbon source consumption by PermutMatrix.

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*. 46, W537–W544. doi: 10.1093/nar/gky379

Afolayan, A. O., Ayeni, F. A., and Ruppitsch, W. (2017). Antagonistic and quantitative assessment of indigenous lactic acid bacteria in different varieties of *ogi* against gastrointestinal pathogens. *Pan Afr. Med. J*. 27, 22. doi: 10.11604/pamj.2017.27.22.9707

Albesharat, R., Ehrmann, M. A., Korakli, M., Yazaji, S., and Vogel, R. F. (2011). Phenotypic and genotypic analyses of lactic acid bacteria in local fermented food, breast milk and faeces of mothers and their babies. *Syst. Appl. Microbiol*. 34, 148–155. doi: 10.1016/j.syapm.2010.12.001

Andersson, U., Levander, F., and Rådström, P. (2001). Trehalose-6-phosphate phosphorylase is part of a novel metabolic pathway for trehalose utilization in *Lactococcus lactis*. *J. Biol. Chem*. 276, 42707–42713. doi: 10.1074/jbc.M108279200

Andersson, U., Molenaar, D., Rådström, P., and de Vos, W. M. (2005). Unity in organisation and regulation of catabolic operons in *Lactobacillus plantarum*, L*actococcus lactis* and *Listeria monocytogenes*. *Syst. Appl. Microbiol*. 28, 187–195. doi: 10.1016/j.syapm.2004.11.004

Andersson, U., and Rådström, P. (2002). Physiological function of the maltose operon regulator, MalR, in *Lactococcus lactis*. *BMC Microbiol*. 2, 28. doi: 10.1186/1471-2180-2-28

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genom*. 9, 75. doi: 10.1186/1471-2164-9-75

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., et al. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res*. 44, W147–W153. doi: 10.1093/nar/gkw419

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol*. 19, 455–477. doi: 10.1089/cmb.2012.0021

Beasley, S. S., Manninen, T. J., and Saris, P. E. (2006). Lactic acid bacteria isolated from canine faeces. *J. Appl. Microbiol*. 101, 131–138. doi: 10.1111/j.1365-2672.2006.02884.x

Bhatia, S., Singh, A., Batra, N., and Singh, J. (2020). Microbial production and biotechnological applications of α-galactosidase. *Int. J. Biol. Macromol*. 150, 1294–1313. doi: 10.1016/j.ijbiomac.2019.10.140

Björkroth, K. J., Schillinger, U., Geisen, R., Weiss, N., Hoste, B., Holzapfel, W. H., et al. (2002). Taxonomic study of *Weissella confusa* and description of *Weissella cibaria* sp. nov., detected in food and clinical samples. *Int. J. Syst. Evol. Microbiol*. 52(Pt 1), 141–148. doi: 10.1099/00207713-52-1-141

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Cai, Y., Benno, Y., Nakase, T., and Oh, T. K. (1998). Specific probiotic characterization of *Weissella hellenica* DS-12 isolated from flounder intestine. *J. Gen. Appl. Microbiol*. 44, 311–316. doi: 10.2323/jgam.44.311

Chen, Y. S., Yanagida, F., and Shinohara, T. (2005). Isolation and identification of lactic acid bacteria from soil using an enrichment procedure. *Lett. Appl. Microbiol*. 40, 195–200. doi: 10.1111/j.1472-765X.2005.01653.x

Choi, A. R., Patra, J. K., Kim, W. J., and Kang, S. S. (2018). Antagonistic activities and probiotic potential of lactic acid bacteria derived from a plant-based fermented food. *Front. Microbiol*. 9, 1963. doi: 10.3389/fmicb.2018.01963

Choi, H. J., Cheigh, C. I., Kim, S. B., Lee, J. C., Lee, D. W., Choi, S. W., et al. (2002). *Weissella kimchii* sp. nov., a novel lactic acid bacterium from kimchi. *Int. J. Syst. Evol. Microbiol*. 52(Pt 2), 507–511. doi: 10.1099/00207713-52-2-507

Collins, M. D., Samelis, J., Metaxopoulos, J., and Wallbanks, S. (1993). Taxonomic studies on some Leuconostoc-like organisms from fermented sausages: description of a new genus *Weissella* for the *Leuconostoc paramesenteroides* group of species. *J. Appl. Bacteriol*. 75, 595–603. doi: 10.1111/j.1365-2672.1993.tb01600.x

De Bruyne, K., Camu, N., De Vuyst, L., and Vandamme, P. (2010). *Weissella fabaria* sp. nov., from a *Ghanaian cocoa* fermentation. *Int. J. Syst. Evol. Microbiol*. 60, 1999-−2005. doi: 10.1099/ijs.0.019323-0

De Bruyne, K., Camu, N., Lefebvre, K., De Vuyst, L., and Vandamme, P. (2008). *Weissella ghanensis* sp. nov., isolated from a Ghanaian cocoa fermentation. *Int. J. Syst. Evol. Microbiol*. 58(Pt 12), 2721–2725. doi: 10.1099/ijs.0.65853-0

De Vuyst, L., and Neysens, P. (2005). The sourdough microflora: biodiversity and metabolic interactions. *Trends Food Sci. Technol*. 16, 43–56. doi: 10.1016/j.tifs.2004.02.012

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res*. 36, W465–W469. doi: 10.1093/nar/gkn180

Emerenini, E. C., Afolabi, O. R., Okolie, P. I., and Akintokun, A. K. (2014). *In vitro* studies on antimicrobial activities of lactic acid bacteria isolated from fresh vegetables for biocontrol of tomato pathogens. *Brit. Microbiol. Res. J*. 4, 351–359. doi: 10.9734/BMRJ/2014/5423

Ennahar, S., and Cai, Y. (2004). Genetic evidence that *Weissella kimchii* Choi et al., 2002 is a later heterotypic synonym of *Weissella cibaria* Björkroth et al., 2002. *Int. J. Syst. Evol. Microbiol*. 54(Pt 2), 463–465. doi: 10.1099/ijs.0.02783-0

Espinoza-Monje, M., Campos, J., Alvarez Villamil, E., Jerez, A., Dentice Maidana, S., Elean, M., et al. (2021). Characterization of *Weissella viridescens* UCO-SMC3 as a potential probiotic for the skin: its beneficial role in the pathogenesis of acne vulgaris. *Microorganisms* 9, 1486. doi: 10.3390/microorganisms9071486

Fanelli, F., Chieffi, D., Di Pinto, A., Mottola, A., Baruzzi, F., and Fusco, V. (2020). Phenotype and genomic background of *Arcobacter butzleri* strains and taxogenomic assessment of the species. *Food Microbiol*. 89, 103416. doi: 10.1016/j.fm.2020.103416

Fusco, V., Quero, G. M., Cho, G. S., Kabisch, J., Meske, D., Neve, H., et al. (2015). The genus *Weissella*: taxonomy, ecology and biotechnological potential. *Front. Microbiol*. 6, 155. doi: 10.3389/fmicb.2015.00155

Fusco, V., Quero, G. M., Stea, G., Morea, M., and Visconti, A. (2011). Novel PCR-based identification of *Weissella confusa* using an AFLP-derived marker. *Int. J. Food Microbiol*. 145, 437–443. doi: 10.1016/j.ijfoodmicro.2011.01.015

Galle, S., Schwab, C., Arendt, E., and Gänzle, M. (2010). Exopolysaccharide-forming *Weissella* strains as starter cultures for sorghum and wheat sourdoughs. *J. Agric. Food Chem*. 58, 5834–5841. doi: 10.1021/jf1002683

Galli, V., Venturi, M., Coda, R., Maina, N. H., and Granchi, L. (2020). Isolation and characterization of indigenous *Weissella confusa* for in situ bacterial exopolysaccharides (EPS) production in chickpea sourdough. *Food Res. Int.* 138(Pt B), 109785. doi: 10.1016/j.foodres.2020.109785

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol*. 57, 81–91. doi: 10.1099/ijs.0.64483-0

Heo, J., Hamada, M., Cho, H., Weon, H. Y., Kim, J. S., Hong, S. B., et al. (2019). *Weissella cryptocerci* sp. nov., isolated from gut of the insect *Cryptocercus kyebangensis*. *Int. J. Syst. Evol. Microbiol*. 69, 2801–2806. doi: 10.1099/ijsem.0.003564

Hu, X., Shi, Y., Zhang, P., Miao, M., Zhang, T., and Jiang, B. (2016). D-Mannose: properties, production, and applications: an overview. *Comp. Rev. Food Sci. Food Saf*. 15, 773–785. doi: 10.1111/1541-4337.12211

Hyun, D. W., Lee, J. Y., Sung, H., Kim, P. S., Jeong, Y. S., Lee, J. Y., et al. (2021). *Brevilactibacter coleopterorum* sp. nov., isolated from the intestine of the dark diving beetle *Hydrophilus acuminatus,* and *Weissella coleopterorum* sp. nov., isolated from the intestine of the diving beetle *Cybister lewisianus*. *Int. J. Syst. Evol. Microbiol*. 71, 4779doi: 10.1099/ijsem.0.004779

Jeckelmann, J. M., and Erni, B. (2020). The mannose phosphotransferase system (Man-PTS) - Mannose transporter and receptor for bacteriocins and bacteriophages. *Biochim. Biophys. Acta Biomembr*. 1862, 183412. doi: 10.1016/j.bbamem.2020.183412

Junior, C. B., and Campos, L. (2004). Identification and characterization of a novel cassava (Manihot esculenta Crantz) clone with high free sugar content and novel starch. *Plant Mol. Biol*. 56, 643–659. doi: 10.1007/s11103-004-4873-9

Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci*. 29, 28–35. doi: 10.1002/pro.3711

Kassara, S., Li, S., Smith, P., Blando, F., and Bindon, K. (2019). Pectolytic enzyme reduces the concentration of colloidal particles in wine due to changes in polysaccharide structure and aggregation properties. *Int. J. Biol. Macromol*. 140, 546–555. doi: 10.1016/j.ijbiomac.2019.08.043

Kwak, M. J., Choi, S. B., Kim, B. Y., and Chun, J. (2019). Genome-based reclassification of *Weissella jogaejeotgali* as a later heterotypic synonym of *Weissella thailandensis*. *Int. J. Syst. Evol. Microbiol*. 69, 3672–3675. doi: 10.1099/ijsem.0.003315

Lee, K. W., Park, J. Y., Jeong, H. R., Heo, H. J., Han, N. S., and Kim, J. H. (2012). Probiotic properties of *Weissella* strains isolated from human faeces. *Anaerobe* 18, 96–102. doi: 10.1016/j.anaerobe.2011.12.015

Lee, S. H., Ku, H. J., Ahn, M. J., Hong, J. S., Lee, S. H., Shin, H., et al. (2015). *Weissella jogaejeotgali* sp. *nov.,* isolated from jogae jeotgal, a traditional Korean fermented seafood. *Int. J. Syst. Evol. Microbiol*. 65, 4674–4681. doi: 10.1099/ijsem.0.000631

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 47, W256–W259. doi: 10.1093/nar/gkz239

Li, J., Ai, L., Xu, F., Hu, X., Yao, Y., and Wang, L. (2021). Structural characterization of exopolysaccharides from *Weissella cibaria* NC516.11 in distiller grains and its improvement in gluten-free dough. *Int. J. Biol. Macromol*. 199, 17–23. doi: 10.1016/j.ijbiomac.2021.12.089

Li, Y. Q., Tian, W. L., and Gu, C. T. (2020). *Weissella sagaensis* sp. nov., isolated from traditional Chinese yogurt. *Int. J. Syst. Evol. Microbiol*. 70, 2485–2492. doi: 10.1099/ijsem.0.004062

Lin, S. T., Wang, L. T., Wu, Y. C., Guu, J. J., Tamura, T., Mori, K., et al. (2020). *Weissella muntiaci* sp. nov., isolated from faeces of Formosan barking deer (*Muntiacus reevesi*). *Int. J. Syst. Evol. Microbiol*. 70, 1578–1584. doi: 10.1099/ijsem.0.003937

Mantovani, C. F., Geimba, M. P., and Brandelli, A. (2005). Enzymatic clarification of fruit juices by fungal pectin lyase. *Food Biotechnol*. 19, 173–181. doi: 10.1080/08905430500316284

Martín, R., Heilig, H. G., Zoetendal, E. G., Jiménez, E., Fernández, L., Smidt, H., et al. (2007). Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women. *Res. Microbiol*. 158, 31–37. doi: 10.1016/j.resmic.2006.11.004

Masuda, Y., Zendo, T., Sawa, N., Perez, R. H., Nakayama, J., and Sonomoto, K. (2012). Characterization and identification of weissellicin Y and weissellicin M, novel bacteriocins produced by *Weissella hellenica* QU 13. *J. Appl. Microbiol*. 112, 99–108. doi: 10.1111/j.1365-2672.2011.05180.x

Michlmayr, H., and Kneifel, W. (2014). β-Glucosidase activities of lactic acid bacteria: mechanisms, impact on fermented food and human health. *FEMS Microbiol. Lett*. 352, 1–10. doi: 10.1111/1574-6968.12348

Montemurro, M., Pontonio, E., and Rizzello, C. G. (2021). Design of a "Clean-Label" gluten-free bread to meet consumers demand. *Foods* 10, 462. doi: 10.3390/foods10020462

Mortezaei, F., Royan, M., Allaf Noveirian, H., Babakhani, A., Alaie Kordghashlaghi, H., and Balcázar, J. L. (2020). *In vitro* assessment of potential probiotic characteristics of indigenous *Lactococcus lactis* and *Weissella oryzae* isolates from rainbow trout (*Oncorhynchus mykiss* Walbaum). *J. Appl. Microbiol*. 129, 1004–1019. doi: 10.1111/jam.14652

Mun, S. Y., and Chang, H. C. (2020). Characterization of *Weissella koreensis* SK isolated from kimchi fermented at low temperature (around 0 °c) based on complete genome sequence and corresponding phenotype. *Microorganisms* 8, 1147. doi: 10.3390/microorganisms8081147

Muñoz-Atienza, E., Gómez-Sala, B., Araújo, C., Campanero, C., del Campo, R., Hernández, P. E., et al. (2013). Antimicrobial activity, antibiotic susceptibility and virulence factors of lactic acid bacteria of aquatic origin intended for use as probiotics in aquaculture. *BMC Microbiol*. 13, 15. doi: 10.1186/1471-2180-13-15

Nisiotou, A., Dourou, D., Filippousi, M. E., Banilas, G., and Tassou, C. (2014). *Weissella uvarum* sp. nov., isolated from wine grapes. *Int. J. Syst. Evol. Microbiol*. 64(Pt 11), 3885–3890. doi: 10.1099/ijs.0.066209-0

O'Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M. M., Gormley, N. A., and Cox, A. J. (2015). NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics* 31, 2035–2037. doi: 10.1093/bioinformatics/btv057

Oh, S. J., Shin, N. R., Hyun, D. W., Kim, P. S., Kim, J. Y., Kim, M. S., et al. (2013). *Weissella diestrammenae* sp. nov., isolated from the gut of a camel cricket (*Diestrammena coreana*). *Int. J. Syst. Evol. Microbiol*. 63(Pt 8), 2951–2956. doi: 10.1099/ijs.0.047548-0

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 42, D206–D214. doi: 10.1093/nar/gkt1226

Padonou, S. W., Nielsen, D. S., Hounhouigan, J. D., Thorsen, L., Nago, M. C., and Jakobsen, M. (2009). The microbiota of Lafun, an African traditional cassava food product. *Int. J. Food Microbiol*. 133, 22–30. doi: 10.1016/j.ijfoodmicro.2009.04.019

Padonou, S. W., Schillinger, U., Nielsen, D. S., Franz, C. M. A. P., Hansen, M., Hounhouigan, J. D., et al. (2010). *Weissella beninensis* sp. nov., a motile lactic acid bacterium from submerged cassava fermentations, and emended description of the genus *Weissella*. *Int. J. Syst. Evol. Microbiol*. 60(Pt 9), 2193–2198. doi: 10.1099/ijs.0.014332-0

Prado, G. K. S., Torrinha, K. C., Cruz, R. E., Gonçalves, A. B. B., Silva, C. A. V., Oliveira, F. M. S., et al. (2020). *Weissella paramesenteroides* WpK4 ameliorate the experimental amoebic colitis by increasing the expression of MUC-2 and the intestinal epithelial regeneration. *J. Appl. Microbiol*. 129, 1706–1719. doi: 10.1111/jam.14671

Praet, J., Meeus, I., Cnockaert, M., Houf, K., Smagghe, G., and Vandamme, P. (2015). Novel lactic acid bacteria isolated from the bumble bee gut: *Convivina intestini* gen. *nov., sp. nov., Lactobacillus bombicola sp. nov.,* and *Weissella bombi sp.* nov. *Antonie Van Leeuwenhoek*. 107, 1337–1349. doi: 10.1007/s10482-015-0429-z

Ricciardi, A., Parente, E., and Zotta, T. (2009). Modelling the growth of *Weissella cibaria* as a function of fermentation conditions. *J. Appl. Microbiol*. 107, 1528–1535. doi: 10.1111/j.1365-2672.2009.04335.x

Rodriguez-R, L. M., Gunturu, S., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., et al. (2018). The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of archaea and bacteria at the whole genome level. *Nucleic Acids Res*. 46, W282–W288. doi: 10.1093/nar/gky467

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *Peer J. Prepr*. 4, e1900ve1901. doi: 10.7287/peerj.preprints.1900v1

Sandes, S., Figueiredo, N., Pedroso, S., Sant'Anna, F., Acurcio, L., Abatemarco Junior, M., et al. (2020). *Weissella paramesenteroides* WpK4 plays an immunobiotic role in gut-brain axis, reducing gut permeability, anxiety-like and depressive-like behaviors in murine models of colitis and chronic stress. *Food Res. Int*. 137, 109741. doi: 10.1016/j.foodres.2020.109741

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sica, M. G., Olivera, N. L., Brugnoni, L. I., Marucci, P. L., López-Cazorla, A. C., and Cubitto, M. A. (2010). Isolation, identification and antimicrobial activity of lactic acid bacteria from the Bahía Blanca estuary. *Rev. Biol. Mar. Oceanogr*. 45, 389–397. doi: 10.4067/S0718-19572010000300003

Snauwaert, I., Papalexandratou, Z., De Vuyst, L., and Vandamme, P. (2013). Characterization of strains of *Weissella fabalis* sp. nov. and *Fructobacillus tropaeoli* from spontaneous cocoa bean fermentations. *Int. J. Syst. Evol. Microbiol*. 63(Pt 5), 1709–1716. doi: 10.1099/ijs.0.040311-0

Srionnual, S., Yanagida, F., Lin, L. H., Hsiao, K. N., and Chen, Y. S. (2007). Weissellicin 110, a newly discovered bacteriocin from *Weissella cibaria* 110, isolated from plaa-som, a fermented fish product from thailand. *Appl. Environ. Microbiol*. 73, 2247–2250. doi: 10.1128/AEM.02484-06

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res*. 44, 6614–6624. doi: 10.1093/nar/gkw569

Teixeira, C. G., Fusieger, A., Milião, G. L., Martins, E., Drider, D., Nero, L. A., et al. (2021). *Weissella*: an emerging bacterium with promising health benefits. *Probiotics Antimicrob. Proteins* 13, 915–925. doi: 10.1007/s12602-021-09751-1

Tofalo, R., Fusco, V., Böhnlein, C., Kabisch, J., Logrieco, A. F., Habermann, D., et al. (2020). The life and times of yeasts in traditional food fermentations. *Crit. Rev. Food Sci. Nutr*. 60, 3103–3132. doi: 10.1080/10408398.2019.1677553

Tomasik, P., and Horton, D. (2012). Enzymatic conversions of starch. *Adv. Carbohydr. Chem. Biochem*. 68, 59–436. doi: 10.1016/B978-0-12-396523-3.00001-4

van der Maarel, M. J., van der Veen, B., Uitdehaag, J. C., Leemhuis, H., and Dijkhuizen, L. (2002). Properties and applications of starch-converting enzymes of the alpha-amylase family. *J. Biotechnol*. 94, 137–155. doi: 10.1016/S0168-1656(01)00407-2

Wang, C., Zhang, C. W., Chen, H. C., Yu, Q., Pei, X. F., and Liu, H. C. (2008). [Phylogeny analysis and identification of two bacterial strains sourcing from human intestine and having resistance to acid and bile]. *Sichuan Da Xue Xue Bao Yi Xue Ban*. 39, 263–366. Available online at: https://europepmc.org/article/med/18630699

Wang, J. P., Yoo, J. S., Jang, H. D., Lee, J. H., Cho, J. H., and Kim, I. H. (2011). Effect of dietary fermented garlic by *Weissella koreensis* powder on growth performance, blood characteristics, and immune response of growing pigs challenged with *Escherichia coli* lipopolysaccharide. *J. Anim. Sci*. 889, 2123–2131. doi: 10.2527/jas.2010-3186

Wang, W., Liu, W., and Chu, W. (2020). Isolation and preliminary screening of potentially probiotic *Weissella confusa* strains from healthy human feces by culturomics. *Microb. Pathog*. 147, 104356. doi: 10.1016/j.micpath.2020.104356

Wang, Y., Wu, J., Lv, M., Shao, Z., Hungwe, M., Wang, J., et al. (2021). Metabolism characteristics of lactic acid bacteria and the expanding applications in food industry. *Front. Bioeng. Biotechnol*. 9, 612285. doi: 10.3389/fbioe.2021.612285

Wangpaiboon, K., Laohawuttichai, P., Kim, S. Y., Mori, T., Nakapong, S., Pichyangkura, R., et al. (2021). A GH13 α-glucosidase from *Weissella cibaria* uncommonly acts on short-chain maltooligosaccharides. *Acta Crystallogr. D. Struct. Biol*. 77(Pt 8), 1064–1076. doi: 10.1107/S205979832100677X

Wolter, A., Hager, A. S., Zannini, E., Czerny, M., and Arendt, E. K. (2014). Influence of dextran-producing Weissella cibaria on baking properties and sensory profile of gluten-free and wheat breads. *Int. J. Food Microbiol*. 172, 83–91. doi: 10.1016/j.ijfoodmicro.2013.11.015

Xiong, L., Ni, X., Niu, L., Zhou, Y., Wang, Q., Khalique, A., et al. (2019). Isolation and preliminary screening of a *Weissella confusa* strain from Giant Panda (*Ailuropoda melanoleuca*). *Probiotics Antimicrob. Proteins*. 11, 535–544. doi: 10.1007/s12602-018-9402-2

Yeu, J. E., Lee, H. G., Park, G. Y., Lee, J., and Kang, M. S. (2021). Antimicrobial and antibiofilmactivities of *Weissella cibaria* against pathogens of upper respiratory tract infections. *Microorganisms* 9, 1181. doi: 10.3390/microorganisms9061181

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 40, W445–W451. doi: 10.1093/nar/gks479

Yu, H. S., Lee, N. K., Choi, A. J., Choe, J. S., Bae, C. H., and Paik, H. D. (2019). Anti-inflammatory potential of probiotic strain *Weissella cibaria* JW15 Isolated from Kimchi through regulation of NF-κB and MAPKs pathways in LPS-induced RAW 264.7 cells. *J. Microbiol. Biotechnol*. 29, 1022–1032. doi: 10.4014/jmb.1903.03014

Zannini, E., Jeske, S., Lynch, K. M., and Arendt, E. K. (2018). Development of novel quinoa-based yoghurt fermented with dextran producer *Weissella cibaria* MG1. *Int. J. Food Microbiol*. 268, 19–26. doi: 10.1016/j.ijfoodmicro.2018.01.001

Zannini, E., Mauch, A., Galle, S.,Gänzle, M., Coffey, A., Arendt, E. K., et al. (2013). Barley malt wortfermentation by exopolysaccharide-forming *Weissella cibaria* MG1for the production of a novel beverage. *J. Appl. Microbiol*. 115, 1379–1387. doi: 10.1111/jam.12329

Zartl, B., Silberbauer, K., Loeppert, R., Viernstein, H., Praznik, W., and Mueller, M. (2018). Fermentation of non-digestible raffinose family oligosaccharides and galactomannans by probiotics. *Food Funct*. 9, 1638–1646. doi: 10.1039/C7FO01887H

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

REVIEW
published: 30 June 2022
doi: 10.3389/fmicb.2022.889788

OPEN ACCESS

Edited by:
Caroline Barretto, Nestlé, Switzerland

Reviewed by:
Anna Gałązka, Institute of Soil Science and Plant Cultivation, Poland
Salma Mukhtar, Connecticut Agricultural Experiment Station, United States
Sandi Orlic, Rudjer Boskovic Institute, Croatia

***Correspondence:**
Sébastien Terrat
sebastien.terrat@inrae.fr
Lionel Ranjard
lionel.ranjard@inrae.fr

†ORCID:
Christophe Djemiel
orcid.org/0000-0002-5659-7876
Battle Karimi
orcid.org/0000-0002-4523-974X
Aurélien Cottin
orcid.org/0000-0002-0862-784X
Walid Horrigue
orcid.org/0000-0001-9983-1039
Sophie Sadet-Bourgeteau
orcid.org/0000-0002-7533-4289
Pierre-Alain Maron
orcid.org/0000-0003-2315-0741
Nicolas Chemidlin Prévost-Bouré
orcid.org/0000-0002-8575-4295
Lionel Ranjard
orcid.org/0000-0002-7720-5843
Sébastien Terrat
orcid.org/0000-0001-5209-6196

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *04 March 2022*
Accepted: *06 June 2022*
Published: *30 June 2022*

Citation:
Djemiel C, Dequiedt S, Karimi B, Cottin A, Horrigue W, Bailly A, Boutaleb A, Sadet-Bourgeteau S, Maron P-A, Chemidlin Prévost-Bouré N, Ranjard L and Terrat S (2022) Potential of Meta-Omics to Provide Modern Microbial Indicators for Monitoring Soil Quality and Securing Food Production. Front. Microbiol. 13:889788. doi: 10.3389/fmicb.2022.889788

Potential of Meta-Omics to Provide Modern Microbial Indicators for Monitoring Soil Quality and Securing Food Production

Christophe Djemiel[1†], Samuel Dequiedt[1], Battle Karimi[1,2†], Aurélien Cottin[1†], Walid Horrigue[1†], Arthur Bailly[1], Ali Boutaleb[1], Sophie Sadet-Bourgeteau[1†], Pierre-Alain Maron[1†], Nicolas Chemidlin Prévost-Bouré[1†], Lionel Ranjard[1†]* and Sébastien Terrat[1†]

[1]Agroécologie, INRAE, Institut Agro, Université Bourgogne, Université Bourgogne Franche-Comté, Dijon, France, [2]Novasol Experts, Dijon, France

Soils are fundamental resources for agricultural production and play an essential role in food security. They represent the keystone of the food value chain because they harbor a large fraction of biodiversity—the backbone of the regulation of ecosystem services and "soil health" maintenance. In the face of the numerous causes of soil degradation such as unsustainable soil management practices, pollution, waste disposal, or the increasing number of extreme weather events, it has become clear that (i) preserving the soil biodiversity is key to food security, and (ii) biodiversity-based solutions for environmental monitoring have to be developed. Within the soil biodiversity reservoir, microbial diversity including Archaea, Bacteria, Fungi and protists is essential for ecosystem functioning and resilience. Microbial communities are also sensitive to various environmental drivers and to management practices; as a result, they are ideal candidates for monitoring soil quality assessment. The emergence of meta-omics approaches based on recent advances in high-throughput sequencing and bioinformatics has remarkably improved our ability to characterize microbial diversity and its potential functions. This revolution has substantially filled the knowledge gap about soil microbial diversity regulation and ecology, but also provided new and robust indicators of agricultural soil quality. We reviewed how meta-omics approaches replaced traditional methods and allowed developing modern microbial indicators of the soil biological quality. Each meta-omics approach is described in its general principles, methodologies, specificities, strengths and drawbacks, and illustrated with concrete applications for soil monitoring. The development of metabarcoding approaches in the last 20 years has led to a collection of microbial indicators that are now operational and available for the farming sector. Our review shows that despite the recent huge advances, some meta-omics approaches (e.g., metatranscriptomics or meta-proteomics) still need developments to be operational for environmental bio-monitoring. As regards prospects, we outline the importance of building up repositories of soil quality indicators. These are essential for objective and robust diagnosis, to help actors and stakeholders improve soil management,

with a view to or to contribute to combining the food and environmental quality of next-generation farming systems in the context of the agroecological transition.

Keywords: meta-omics, microbial indicators, food value chain, biomonitoring, soil quality

LINK BETWEEN SOIL AS A SUPPORT FOR FOOD PRODUCTION AND A HABITAT FOR LIVING ORGANISMS

Soil is the most complex natural physicochemical matrix and the most structured one on our planet. This heterogeneous layer is the most superficial one of the Earth's crusts, and can be from a few centimeters to several tens of meters deep. The variability of this matrix can be structured into three dimensions: horizontal, vertical, and temporal. Horizontal variability results from the spatial distribution of the different types of soils, influenced by the type of parent material (limestone, granite, metamorphic rock) on a large scale (Chesworth, 2007). Vertical variability refers to the different layers – called "horizons" – that have differentiated over time (Hartemink et al., 2020). Temporal variability results from the pedogenesis process, which is at the origin of soil creation from the weathering of the parent material by climate, vegetation and living organisms or human interventions. Soil formation is a slow process: 0.05 mm of soil are formed each year on average (Evans et al., 2019). Altogether, this makes soils a mosaic of extremely rich and diversified habitats providing a large number of macro- and micro-environments for micro-, meso- and macroscopic living organisms (Nielsen et al., 2015). It is home to many organisms that spend part or all of their life cycle underground and are part of the soil biota (e.g., nematodes, tardigrades, earthworms, microorganisms; **Figure 1**; Bardgett and Van Der Putten, 2014). Thanks to this mosaic of micro- and macro-habitats, soils host almost one third of our planet's biodiversity (FAOSTAT, 2021).

The food value chain summarizes the expression "from farm to fork," including all steps from food production to food distribution to consumers, *via* food manufacturing. Soils are fundamental resources for agricultural production and play an essential role for food security (Kopittke et al., 2022). As providers of more than 95% of food and feed, they represent the cornerstone of the food value chain (Friedrichsen et al., 2019; FAOSTAT, 2021). Soil nutrient availability is unevenly distributed on regional and global scales and affects agricultural productivity (Silver et al., 2021). Soils harbor a large fraction of biodiversity – the backbone of the regulation of ecosystem services and soil multi-functionality. However, these are increasingly threatened by anthropogenic activities such as unsustainable soil management practices, pollution, waste disposal, or the increasing number of extreme weather events. Thus, monitoring the soil biodiversity has become an imperious and urgent issue to secure food production, and requires the development of microbial indicators (Kopittke et al., 2019). New initiatives have recently been launched at a global scale to try and build indicators of food system sustainability (Béné et al., 2019), but microbial diversity was missing from the 27 selected indicators.

Soil Microorganisms: Modern Bioindicators of Soil Quality in a Context of Agroecological Transition

Definition of Bioindicators

A bioindicator (or biological indicator) is a measurement of macro- or micro-organisms whose response in terms of presence/absence, abundance, activities/functions, morphology, physiology or behavior provides information on the state of a habitat or ecosystem (Martinez-Salgado et al., 2010; Astudillo-García et al., 2019). It is used to assess the quality of an environment and its evolution over time. Bio-monitoring of soil quality is carried out at different spatial (local, landscape, regional or global) and temporal (years, decades, millennia) scales, but also at different levels of biological organization (individual, population, community, ecosystem; Pulleman et al., 2012).

Soil bioindicators are used to monitor soil responses to disturbances such as land use changes, agricultural management, or soil contamination, and to assess environmental risks (Geisen et al., 2019). It is essential for a bioindicator to have a good sensitivity to soil management and environmental pressures, and to be relevant to soil functions (e.g., organic matter decomposition, mineralization of soil nitrogen, formation of the soil structure) or related to soil quality (Bloem et al., 2005). From an operational perspective, an indicator should (i) be user-friendly in the field (ii) be precise and reproducible, and (iii) have an affordable financial cost to be used by the greatest number of actors and on a large scale (Griffiths et al., 2018). It should be as universal as possible for the various microbiomes (Schloter et al., 2018). Macro-fauna organisms (e.g., nematodes and earthworms) long remained the only indicators of soil biological quality because they were easily observed and well established in ISO standards (Martinez-Salgado et al., 2010; Schloter et al., 2018; Phillips et al., 2021). Microorganisms were used as bioindicators of soil quality later, but only when the molecular ecology era began (Maron et al., 2011). There are currently several bioindicators of soil quality available, based on microbial taxa and their associated functions (Schloter et al., 2018; Thiele-Bruhn et al., 2020).

Microorganisms Are Good Bioindicators of Soil Quality

Microorganisms (mainly bacteria, fungi and protists) and their functional guilds (pathotrophs, saprotrophs, symbiotrophs) are involved in many ecosystem services (**Figure 1**), in particular through their role in the biogeochemical cycles of major elements (Hermans et al., 2017; Frac et al., 2018; Geisen et al., 2018). They are involved in N cycling by fixing atmospheric N,

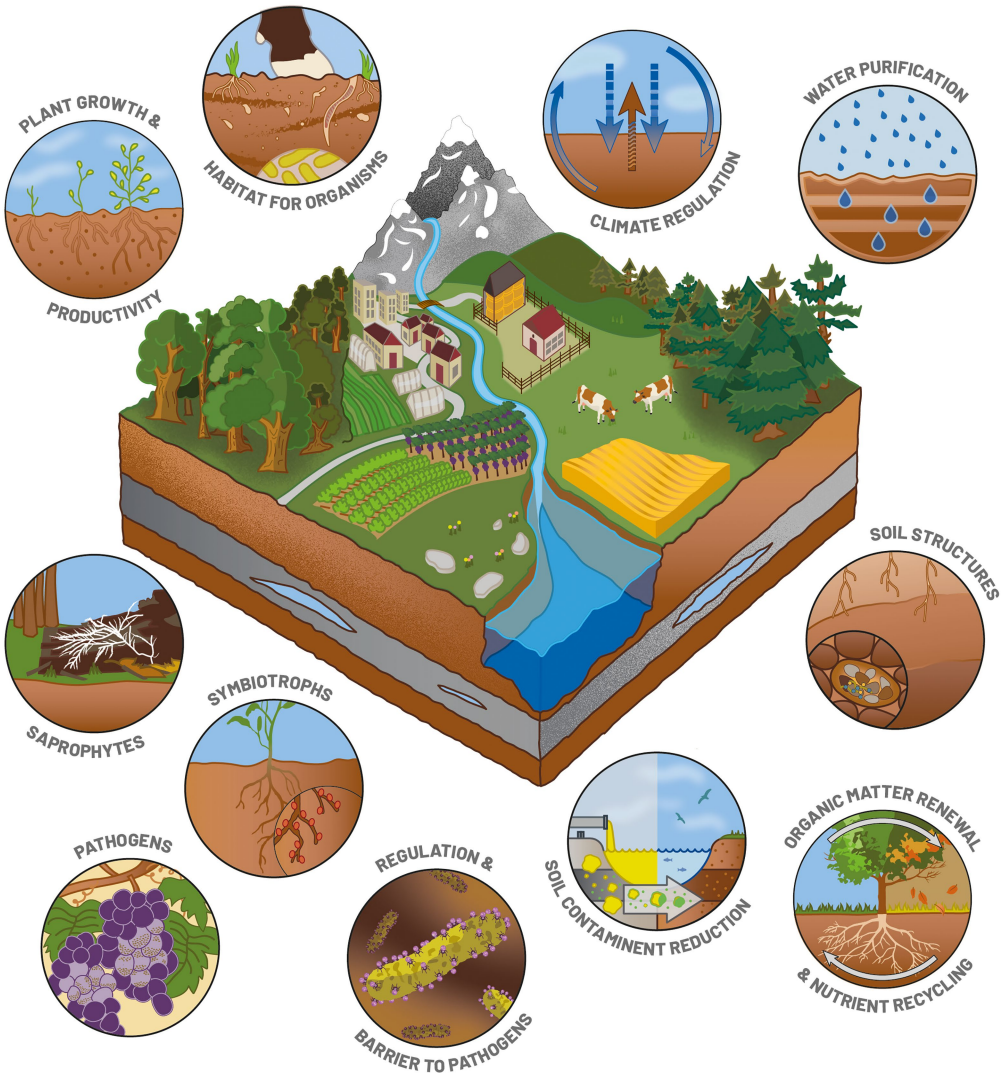


FIGURE 1 | Illustration of the main roles of soil functions and microorganisms as essential players in various ecosystem services.

denitrification and ammonification (Hayatsu et al., 2008). Soil organic matter mineralization is mainly driven by microbial communities that transform the complex carbohydrates into mineral elements available for plant growth (Nielsen et al., 2015; Maron et al., 2018; Yang et al., 2018). Various microorganisms – especially bacteria and fungi – can achieve symbiosis with plants and benefit to plant growth and productivity (Prudent et al., 2020; **Figure 1**). They can also control plant pathogens thanks to metabolites and provide a barrier against pathogens (**Figure 1**). Soil microbial communities deliver other kinds of services that promote the soil structure or reduce soil contamination (FAO, ITPS, GSBI, and SCBD, 2020; **Figure 1**). Their deep involvement in ecological services make microorganisms ideal candidate bioindicators for monitoring the soil quality and food safety (**Figure 2**). In addition, their high sensitivity to environmental disturbances combined with their short generation time makes them early indicators of disturbances. As the cost of some meta-omics technologies is getting lower, the trends are now to define news bioindicators, and an inventory of operational bioindicators is necessary to provide a benchmark for soil quality assessment. This evaluation needs to include (i) the currently available meta-omics approaches so as to evaluate their benefits and technical or conceptual drawbacks (ii) the establishment of an adapted sampling design, and (iii) the development of repositories for an objective interpretation of the results.

CONTRIBUTION OF -OMICS TECHNOLOGIES TO THE MONITORING OF THE SOIL MICROBIOLOGICAL QUALITY

Over the past two decades, many molecular tools have emerged that have accelerated the exploration of the soil microbial diversity, and later contributed to soil quality assessment

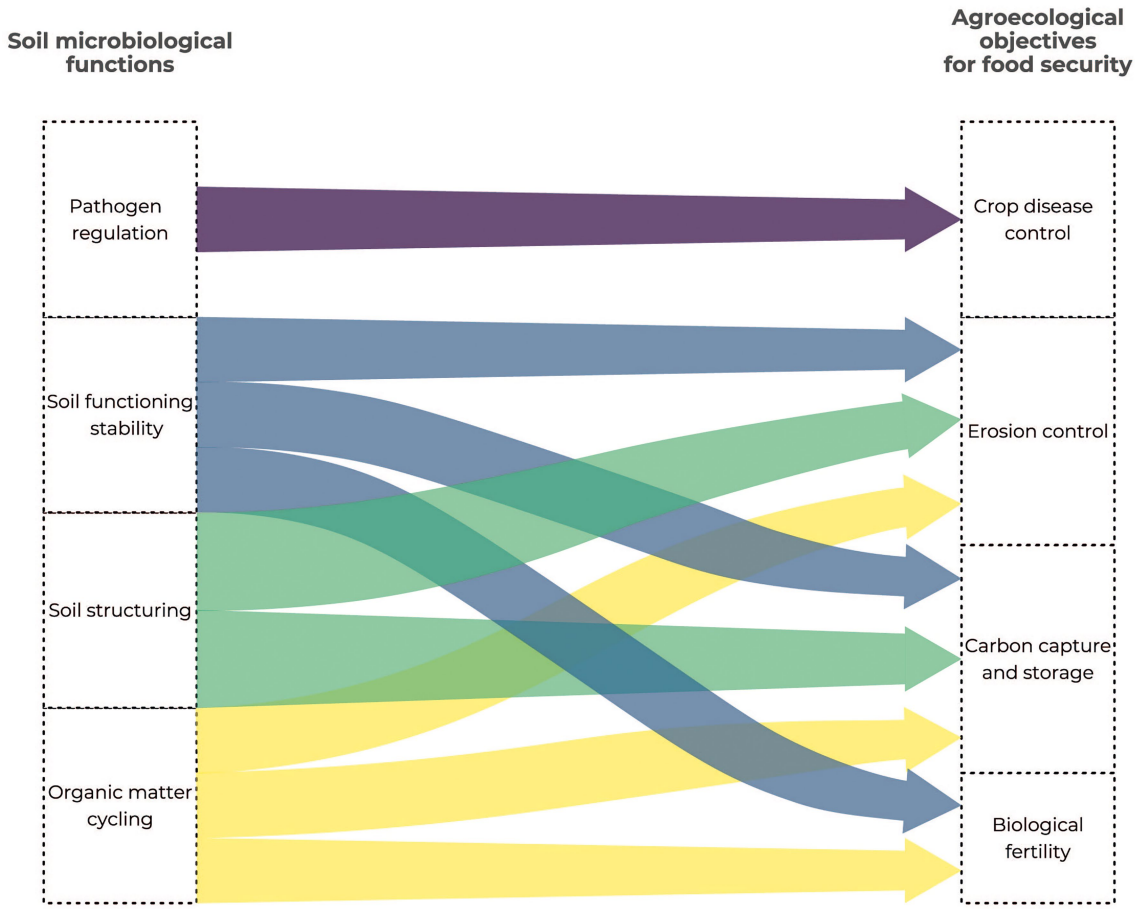


FIGURE 2 | Link between the soil microbiological functions and the agroecological objectives for food security.

(Schloter et al., 2018; Nkongolo and Narendrula-Kotha, 2020). These methods can be divided into three major categories: (i) quantitative methods (ii) hybridization methods (microarrays), and (iii) sequencing methods.

Quantitative Methods

Different quantitative methods have been developed to investigate the relative or absolute abundance of taxa but also their functional profiles by targeting specific genes or gene families of interest (e.g., antibiotic resistome, functional genes in the N cycle; Oshiki et al., 2018). Quantitative polymerase chain reaction (qPCR), high-throughput quantitative PCR (HT-qPCR) and droplet digital PCR (ddPCR) provide access to the numbers of copies of one or more genes. Technological advances have made it possible to simultaneously analyze several thousands of samples and targets (Porter and Hajibabaei, 2018; Mehle and Dreo, 2019). More sensitive and more accurate, ddPCR appears to be promising for an absolute quantification of genes from complex samples such as soil samples because it can overcome the limitations associated with the presence of inhibitors (Gutiérrez-Aguirre et al., 2015).

Microarray Methods

DNA microarray is based on hybridization between (i) specific oligonucleotides used as probes and fixed on a solid surface, and (ii) the target corresponding to the DNA/RNA soil sample (Srivastava et al., 2019). Combining hybridization and labeling (e.g., by fluorescence) of DNA, this semi-quantitative technology is used to detect the presence or relative abundance of target markers or to monitor gene expression profiles at the individual or community levels (Sessitsch et al., 2006; Porter and Hajibabaei, 2018). Various microarrays such as PhyloChip, GeoChip (He et al., 2007), PathoChip (Lee et al., 2013), StressChip (Zhou et al., 2013), CAZyChip (Abot et al., 2016) have been developed to capture a snapshot of specific microbial populations and/or functional genes. For example, GeoChip was designed to detect genes from various metabolic pathways including the carbon, nitrogen, sulfur and phosphorus cycles. The latest version includes about 160,000 distinct probes covering about 1,500 functional gene families. PathoChip is less widespread and contains over 3,000 probes characterizing virulence factors (e.g., adhesion, colonization, motility, toxins) covering 1,400 species. CAZyChip was developed in 2016 to monitor the expression profiles of genes encoding bacterial enzymes specifically involved in polysaccharide degradation. These DNA

chips contain 55,220 probes targeting glycoside hydrolases covering more than 70% of the families present in the CAZy database. Some of these microarrays have become tools for diagnosis, in particular for targeting viruses, bacterial or fungal pathogens or harmful organisms.

From DNA Fingerprinting to High-Throughput Sequencing: Characterizing Microbial Diversity and Functions

DNA fingerprinting was the first high-throughput analytical technique used to comprehensively characterize the genetic structure of soil microbial communities in terms of presence, absence or relative abundance. These approaches are frequently based on PCR amplification of a molecular marker, and classified in two groups depending on the polymorphism of the studied sequences. The first group is usually based on the sequence length polymorphism and includes T-RFLP (restriction fragment length polymorphism), ARDRA [amplified rDNA (ribosomal DNA) restriction analysis], or ARISA (automated ribosomal intergenic spacer analysis). The second group relies on the polymorphism in the composition of the nucleotide sequence and involves techniques such as DGGE (denaturing gradient gel electrophoresis) or TGGE (temperature gradient gel electrophoresis; Rincon-Florez et al., 2013). These tools have been used for various soil microbial diagnostics because they are inexpensive and relatively fast (Sessitsch et al., 2006; Rincon-Florez et al., 2013; van Dorst et al., 2014). However, their sensitivity only identifies the dominant groups, and the comparison of samples (inter-gel) remains difficult (Schloter et al., 2018). Moreover, they do not provide any identification of the organisms or measurement of microbial diversity. These fingerprinting tools have been totally replaced by high-throughput sequencing.

Second- and third-generation high-throughput sequencing offers new ways of answering those various technical and scientific issues. Sequencing coverage, raw data quality, read length, technical biases vary depending on the method (Goodwin et al., 2016). Nevertheless, the performance of high-throughput sequencing platforms is constantly evolving to better support scientific goals at increasingly affordable costs. This has enabled the democratization of several meta-omics approaches for scientific teams but also the transfer of bio-indicators for the agri-food industries. In the next section, we explore and describe the different promising meta-omics approaches for the development of soil quality bioindicators (**Figure 3**).

META-OMICS APPROACHES: GUIDELINES FOR AND APPLICATIONS TO THE FOOD VALUE CHAIN

A paradigm shift emerged in environmental microbiology studies at the end of the 1990's, as the studies based on culture-dependent methods disappeared and were replaced by culture-independent methods. The transition was facilitated by our capacity to extract and access the molecular resources (DNA, RNA, proteins, metabolites) of microbial communities directly from environmental samples. This determined the beginning of the present meta-omics area. These high-throughput approaches for screening hundreds to thousands of samples simultaneously and retrieving a huge amount of data have been developed and have evolved considerably over time. Studies related to soil and/or food have known rapid and important progress thanks to meta-omics since 2010 (**Figure 4**). However, very few studies related to soil and food were based on meta-omics compared to the number of studies focused on soil or food independently of each other. The metagenomic and metabarcoding surveys were the most frequently used approaches to study the soil and food microbiota and the microbiome, while metatranscriptomics and meta-proteomics were rarely investigated (**Figure 4**). The countries that produced the greatest numbers of publications about soil and/or food microbial research based on meta-omics were the USA and China, followed by Germany and France for European countries (**Figure 5**).

Assessing Analytical Specificity in Soil Analysis

Consequences of the Sampling Effort and Sample Conservation on Soil Analysis

Any sampling survey first involves a series of decisions that determine how the data should be analyzed (Dickie et al., 2018), e.g., what is sampled, what is not sampled, where it is sampled, how many samples to be taken, how many replicates, location of subsamples within plots, etc. Even storage conditions of soil samples can influence the quality of the molecular resource and should be well adapted. For example, the use of sieved soils and −40°C storage is efficient to access and keep DNA, but not to preserve and extract a more sensitive molecule like RNA (Ranjard et al., 2009). Fresh soils are undoubtedly an ideal material for characterizing microbial taxonomic and functional diversity, but using them is sometimes impossible in the case of large sampling surveys or comparisons among chronological samples requiring storage conditions (Cui et al., 2014; **Table 1**).

Extraction of Soil Genetic Resources: DNA and RNA

Meta-omics approaches started with isolating DNA directly from environmental samples. Success essentially depended upon the quality and the quantity of isolated DNA (Terrat et al., 2015). Various homemade DNA extraction protocols and commercial kits were developed to improve the efficiency of soil DNA extraction (Martin-Laurent et al., 2001; Delmont et al., 2011; Terrat et al., 2012). However, each method is characterized by its own advantages and biases, leading to variations in DNA representativeness. These variations have consequences on the observed effects on soil microbial assessments, and this makes it difficult to compare studies.

The choice of a specific DNA extraction method determines the inherent bias involved throughout the whole project

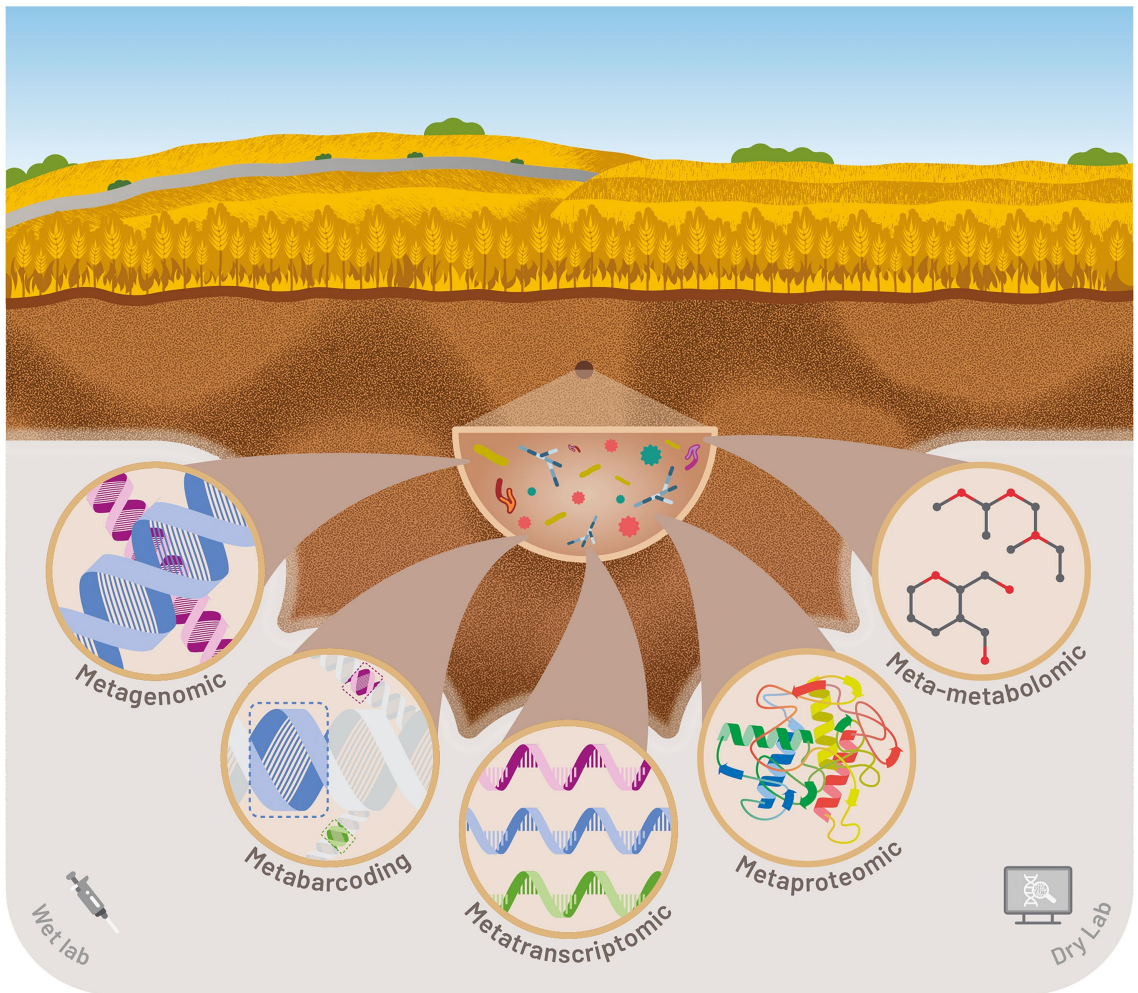


FIGURE 3 | Illustration of various meta-omics approaches for monitoring soil quality and securing food production.

(Vestergaard et al., 2017). This is why standardized methods (like the ISO 11063: 2012 standard) are advised in order to provide consistent results across studies, more particularly when measurements are interpreted relatively to references used as bioindicators of soil quality (**Table 1**).

DNA provides information only about the presence of microorganisms in soils but does not inform on microbial activities. On the contrary, RNA analyses potentially describe what those microorganisms are actually doing and how fast they are doing it (Prosser, 2015). But RNA is a very sensitive molecule, with a low stability and a short half-life time (Hambraeus et al., 2003), and needs to be protected immediately after soil sampling using dedicated methods or solutions (e.g., LifeGuard Soil Preservation Solution or RNA*later*). Moreover, to protect RNA from degradation by RNases, a clean laboratory must be dedicated. Inactivation reagents for RNases such as guanidine thiocyanate must be added to extraction buffers to inactivate RNase molecules immediately after their release from lyzed cells. Finally, no method for RNA extraction from all types of soils is available; this hinders the study of microbial gene expression in soils (Wang et al., 2012).

Extraction of Soil Molecular Resources: Proteins and Metabolites

Due to its huge potential for linking functional and phylogenetic information on soil microbial communities, research in soil ecology has showed a rising interest in meta-proteomics to investigate microbe-driven ecosystem functions (Prosser, 2015; Keiblinger et al., 2016). However, the application of meta-proteomics to soils still faces several challenges such as the heterogeneity of soil matrices or the ecosystem-specific dominance of a few microbial species, which limit the development of a generic protocol. The extraction of soil proteins is sensitive to the presence of several organic compounds such as complex carbohydrates, lipids, phenolics (e.g., lignin), humic substances, but also inorganic compounds such as silt and clay minerals (Keiblinger et al., 2016). Proteins may be adsorbed or linked anchored or embedded onto solid particles and soil

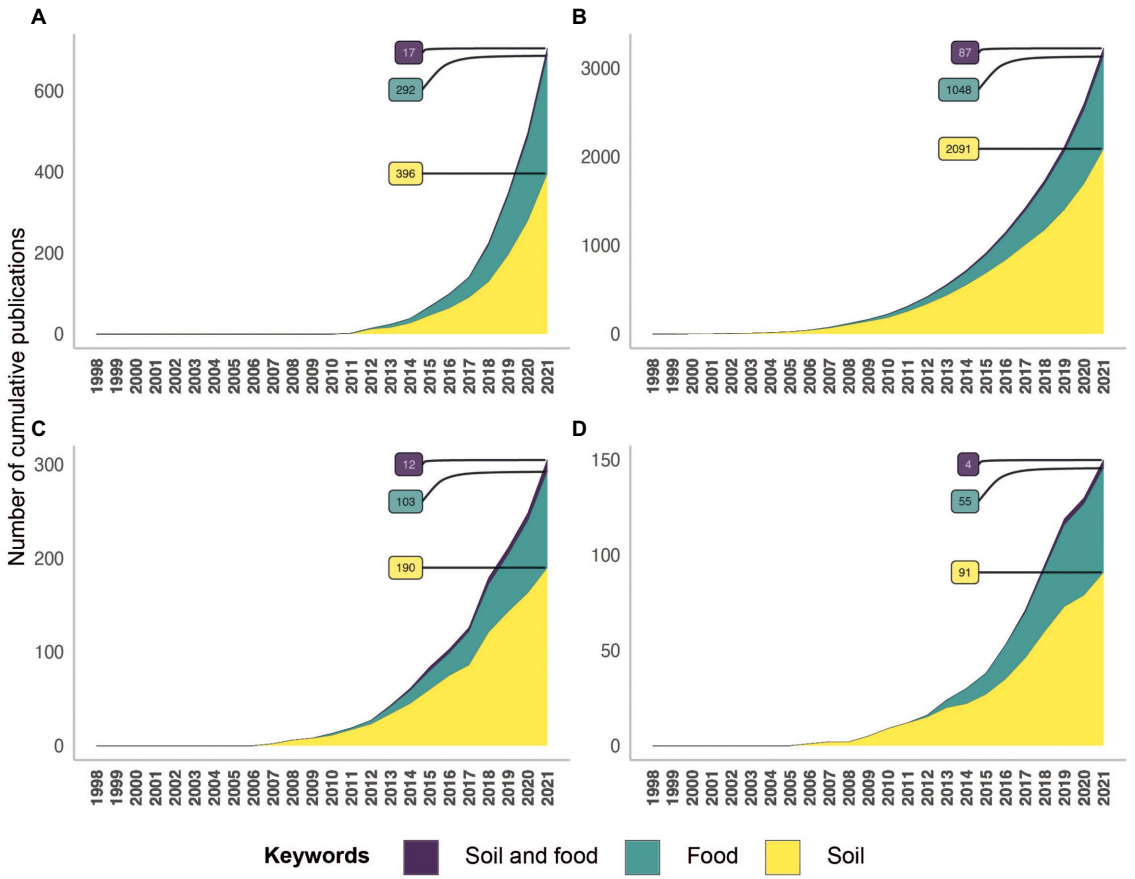


FIGURE 4 | Temporal evolution of the number of references indexed in the Pubmed database concerning meta-omics approaches in relation to soil (yellow), food (green) and soil and food (purple). This analysis relies on keywords present in the title or abstract, related to meta-omics (metabarcoding **(A)**, metagenomics **(B)**, metatranscriptomics **(C)** and meta-proteomics **(D)**) across the different research topics (soil; food; soil and food). Concerning the analysis of metagenomics keywords, these can be biased because a lot of publications used "metagenomics" instead of metabarcoding.

organic matter, hence reduced extraction efficiency. In addition, measurements of soil enzyme activities can be affected in different ways depending on the storage method, the soil type and even on the assay method. Freezing of small sampled aliquots retains the activity of most enzymes, but ultra-freezing is preferred for storage of samples with a high organic matter content (Wallenius et al., 2010). Altogether, processing fresh samples whenever possible or storage at −80°C is recommended to minimize the activity of naturally occurring proteases in environmental samples (Keiblinger et al., 2016).

Soil organic matter – more precisely the dissolved fraction – is composed of many small molecules of plant and microbial origin (Swenson et al., 2015). Classical methods for extracting soil metabolites involve long extraction times followed by compound-specific analyses (Swenson et al., 2015). However, the soil physical properties such as mineral composition, surface area, shape and porosity represent different challenges when it comes to extracting soil metabolites and metabolomics approaches. For example, amines adsorb to clay material in soils, and so do other metabolites, especially carboxylic acids that potentially form covalent bonds with organic matter or are chelated by metals or other ions (from the soil or from the extractant). Moreover, each metabolite has specific half-life times depending on the soil specificities, but also on the activities of the organisms (i.e., enzymatic degradation). Finally, the quantity, stability and specificity of metabolites after extraction can be influenced by soil storage conditions (**Table 1**).

Metabarcoding to Investigate Microbial Diversity

Metabarcoding, also known as targeted metagenomics, uses the polymerase chain reaction (PCR) method to target a genomic region common to all studied microorganisms (Semenov, 2021). This molecular marker can be a gene allowing to study the relationships between microorganisms (considered as a taxonomic marker; Woese et al., 1990) or a functional gene/family (considered as a functional marker; Barbi et al., 2014). DNA fragments (amplicons) produced by the PCR and sequenced by high-throughput sequencing provide an overview of microbial diversity (Nilsson et al., 2019a). Several molecular markers exist to characterize

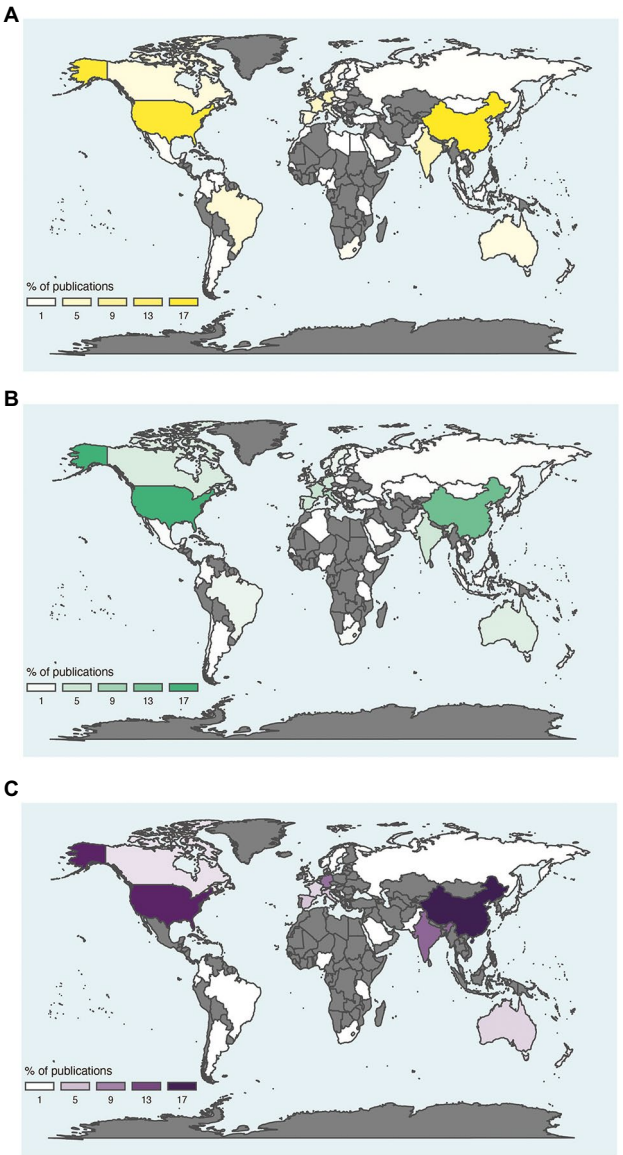


FIGURE 5 | Distribution of publications on meta-omics approaches in relation to soil **(A)**, food **(B)** and soil and food **(C)**. World distribution of publications on soil **(A)**, food **(B)** and soil and food **(C)** using meta-omics approaches. This analysis relies on keywords present in the title or abstract, related to meta-omics (merging studies on metabarcoding, metagenomics, metatranscriptomics and meta-proteomics) and based on the country of the first author. Gray areas, absence of publication.

TABLE 1 | Influence of the conservation strategy of the soil sample on the quality of the extracted biological material.

Preservation strategy	DNA	RNA	Proteins	Metabolites
Immediate sample processing	+++	+++	+++	+++
Drying (room temperature)	+/–	–	–	–
Cold storage (4°C)	+/–	–	–	–
Freezing (–20°C)	++	+/–	+/–	+
Low freezing (–40°C to –80°C)	+++	+++	+++	+++
Chemical preservation (use of stabilizing solution)	+/–	++	++	unknown

+++, Image close to that of the sample; ++, Good protection of the sample; +, Average protection of the sample; +/–, Moderate impact of the storage method on the image; –, Strong impact of the storage method on the results.

different microbial communities. The most popular marker is the gene encoding the small subunit ribosomal RNA (SSU rRNA), essential to ribosome formation. The 16S rRNA gene is used to determine the diversity of Archaea and Bacteria (Woese and Fox, 1977). Similar to the 16S rRNA gene, the 18S rRNA gene is used to explore eukaryotic organisms, and in particular the diversity of the fungal kingdom (Hadziavdic et al., 2014). Considered as the universal DNA barcode marker to investigate fungal diversity, the nuclear ribosomal internal transcribed spacer (ITS) is also applied to identity fungi. Among further taxonomic markers of other important functions are the *gyrB* (DNA gyrase subunit B), *recA* (recombinase A), or *rpoB* (RNA polymerase subunit beta) genes (Holmes et al., 2004; Poirier et al., 2018; Ogier et al., 2019). The choice of the marker depends on the targeted scientific question/objective and microbial communities. Using the metabarcoding approach is advantageous in several ways. It has become very popular in the last decade, and is undoubtedly the most affordable meta-omics approach. Thus, many projects have emerged to evaluate soil quality at large and local scales. Data analysis is now fully operational thanks to a significant range of open-source bioinformatics pipelines such as mothur (Schloss et al., 2009), qiime2 (Bolyen et al., 2019), FROGS (Escudié et al., 2018), or BIOCOM-PIPE (Djemiel et al., 2020a) and the associated standard operating procedures (SOPs) or tutorials. Nevertheless, this wide choice requires grasping well the subtleties of the algorithms implemented in the pipelines and their parameters because the accuracy of results is highly dependent on them (Pauvert et al., 2019). Another major effort in this area is the availability of reference databases for taxonomic assignment whatever the molecular marker (Quast et al., 2013; Nilsson et al., 2019b; Djemiel et al., 2020b). It should also be emphasized that some sequences could contain misidentifications or errors (Hofstetter et al., 2019). Therefore, it is important to be careful when interpreting taxonomic results, especially at the genus or species ranks. The initiation of a metabarcoding project requires choosing a molecular marker and then one or more regions to be amplified. This choice is crucial because the regions do not all have the same resolution and depend on the targeted samples and communities (Banos et al., 2018; Bukin et al., 2019). As a result, amplicon length is also crucial and impacts on the sequencers to be chosen and library preparation (Nilsson et al., 2019a; Castaño et al., 2020). The sequence length generally ranges from a few dozen to one hundred base pairs; this size does not allow fine taxonomic resolution to determine the species names.

The use of metabarcoding approaches to monitor soil quality and also other compartments of the food chain value is scarce. However, some studies have already applied such approaches and brought significant knowledge on microbiota interactions

(Stefanini and Cavalieri, 2018). In a recent study, Zarraonaindia and colleagues investigated 725 samples from soils and grapevine organs using a metabarcoding approach to determine the influence of the vine cultivar., edaphic parameters, the vine developmental stage (dormancy, flowering, pre-harvest), and vineyard characteristics on the spatial and temporal dynamics of bacterial communities (Zarraonaindia et al., 2015). They concluded that most organ-associated taxa originated from the soil, and their distribution reflected the influence of highly localized biogeographic factors and vineyard management. Moreover, the vineyard-associated localization of bacterial taxa has implications for wine growers who rely on the assumption that the soil imparts a unique quality to the wine specific to each vineyard, called the "terroir effect." Complementary to the evaluation of a microbial terroir, such studies investigate the environmental and human-related factors that influence the composition of microbial populations and potentially affect their performances, like fermentation (Stefanini and Cavalieri, 2018).

Metagenomics to Explore the Functional Potential of Microbial Communities

The metagenomics approach – or shotgun sequencing (without any *a priori* compared to metabarcoding) – rests upon the sequencing of total DNA from environmental samples. The extracted DNA is directly sequenced (or after fragmentation or nebulization). This avoids PCR amplification, and non-culturable and/or unknown organisms can be characterized, such as viruses (Prosser, 2015). Its main purpose is to characterize whole microbial genomes across microbiome samples. With the increasing efficiency of sequencing technologies, this approach can now be applied with a sufficient amount of acquired sequences to reconstruct "population genomes" of major microorganisms (Prosser, 2015). Metagenomics leads to an overview of the potential genomic structure of the microbial community, its potential functional properties, in a more sensitive manner than metabarcoding does (Bünemann et al., 2018). Moreover, with the democratization of sequencing technologies, a variety of tools and analysis pipelines have been developed to analyze metagenomics data (Kieser et al., 2020). Configuring various tools, linking them with advanced binning and annotation tools, and managing the updates and parameters modifications of each tool is now easier and more accessible for researchers with the efficient and automated deployment of workflows like ATLAS, MG-RAST or MOCAT2 (Meyer et al., 2008; Kultima et al., 2016; Kieser et al., 2020).

The use of metagenomics for characterizing microbial communities presents several advantages. As already stated, metagenomics characterizes unknown microorganisms in the absence of prior knowledge, and provides information on potential functions based on the genomic content. This allows a deeper understanding of the response of microbial communities to environmental changes (Aguiar-Pulido et al., 2016a). Some drawbacks have also been identified. The presence of a functional gene is not proof of its expression. The host organism may be dormant, inactive or only active when alternatively expressed metabolic pathways do not require the function of the detected gene (Prosser, 2015). Moreover, for economic reasons the resulting datasets often have low coverage of the genomes in the microbiome. Finally, even with available bioinformatics pipelines, the analysis of such datasets needs expertise and can be time-consuming (Aguiar-Pulido et al., 2016a), with a high percentage of reads considered as "Unknown" due to the absence of similar reference genomes.

Metagenomics outcompetes metabarcoding in that it provides insights into the metabolic, virulence or resistance potential of microbial communities. A good example is the study of Wind and colleagues (Wind et al., 2021) focused on antibiotic resistance genes (ARG) across a vegetable production system. The study was interested on three pre-harvest "control-points," i.e., manure and compost amendments, soils, and lettuce samples, to track antibiotic resistance and the associated microbiomes in order to identify the potential impacts of key agricultural management strategies. The soils were found highly stable in the composition of the resistome, which has the potential to act as a natural ecological buffer to ARG proliferation, at least over one harvest cycle. But lettuce grown in soils amended with compost produced from the manure of cows had higher total relative ARG abundance and risk scores compared to the soil amended with stockpiled manure. Similar results were described with poultry manure application: ARG abundance increased on lettuce surfaces at the greenhouse scale (Zhang et al., 2019). The authors of these studies advise reducing the spreading of ARG from amendment onto the lettuce leaf surface as this may influence human health negatively. Many other scientific questions can be tackled using metagenomics, such as Zn availability in grains for biofortification (Wang et al., 2021), or the detection of foodborne pathogens in production chains (Yang et al., 2016).

Metatranscriptomics Dedicated to Studying Gene Expression in Microbial Communities

Metatranscriptomics is quite similar to metagenomics, but different in that it focuses on microbial gene expression from complex environments. Metatranscriptomics provides the diversity of the active genes within a community, and measures the changes in the level of expression due to modified environmental conditions (Mukherjee and Reddy, 2020). Extracted RNAs are reverse-transcribed into cDNA and sequenced, as in metagenomics studies. Pioneering studies aiming to identify expressed genes in environmental samples date back to 2005 and represent the dawn of metatranscriptomics (Prosser, 2015). As in metagenomics studies, reads are either mapped to a reference genome or *de novo* assembled into contigs and scaffolds (Prosser, 2015). Several advantages of metatranscriptomics can be highlighted. This approach characterizes unknown microorganisms, as metagenomics does. It is complementary to metagenomics in that it provides a snapshot of expressed genes (as only coding sequences are obtained). Thus, the active microbiome is explored to understand how patterns of gene expression change following biotic and

abiotic stresses and various environmental disturbances. Yet, although it is easy to recover large quantities of total RNA from a pure bacterial culture (tens of micrograms *per* extraction), this is rather difficult in soil samples (Wang et al., 2012). According to reports, the extraction yields of soil RNA range from tens of nanograms to several micrograms *per* gram of soil. Such a range could be explained by the amount of living microorganisms in the soil samples, by contamination by humic substances or by RNA loss during purification (Wang et al., 2012). The technical issues associated with metatranscriptomics are similar to those associated with metagenomics, but with further potential biases introduced during rRNA depletion (as rRNAs can represent more than 80% of total RNAs), the construction of cDNA libraries, and the required rapid inactivation of samples to prevent mRNA turnover (Prosser, 2015).

As metatranscriptomics approaches can be complicated to apply across the whole food value chain, only specific compartments have been studied. For example, the microbiota of various Swiss-type cheeses were analyzed during ripening (Duru et al., 2018). Samples were collected from three cheeses at two time points – during warm room ripening and cold room ripening. The resulting data supported a better understanding of the flavor-forming mechanism, i.e., the succession of up-regulated and down-regulated pathways involving microbial species during warm and cold ripening of the cheese. Such studies are a first step toward improving the use of food microbiomes in terms of flavor, quality and security (Afshari et al., 2020).

Characterization of Microbial Community Functionality and Activity: Meta-Proteomics and Meta-Metabolomics

Meta-proteomics and meta-metabolomics are devoted to the description of microbial functionality and activity. Therefore, they are the functional complement of studies focused on DNA and/or RNA. Extracellular enzymes synthesized by microbial communities contribute to various ecosystem services such as organic matter decomposition or nutrient cycling (Van Emon, 2016). For this purpose, meta-proteomics characterizes the entire protein content of an environmental microbiome at a given location and time. Meta-metabolomics is dedicated to the comprehensive analysis of all metabolites (i.e., small molecules released by microorganisms into the environment) contained in a sample. Due to its large potential for providing a link between functional and phylogenetic information on soil microbial communities, there has been growing interest in the application of meta-proteomics in soil ecology to study microbe-driven ecosystem functions (Keiblinger et al., 2016). Bacteria, fungi and protists indeed excrete a wide range of volatile organic compounds to interact and communicate with each other, promote crop growth or modulate plant defense (Schulz-Bohm et al., 2017; Poveda, 2021).

Both approaches use common methodologies to extract proteins or metabolites from soil matrices. Direct extraction can be applied, using physical, chemical or mechanical lysis of cells, leading to comprehensive protein (or metabolite) recovery from soil microorganisms. For the choice of the cell lysis method, the target proteins and soil texture should be considered (Keiblinger et al., 2016). This direct approach is often complex due to the presence of other organic compounds co-extracted with proteins or metabolites. These co-extracted products can also complexify the analysis steps. Moreover, the choice of the extraction method can clearly impact the results of the meta-metabolomics approach. Another solution consists in separating microbial cells from the soil matrix prior to extraction, e.g., by density gradient centrifugation (Keiblinger et al., 2016). This indirect extraction reduces the problems caused by interfering and coextracted substances. However, all proteins, enzymes and metabolites excreted by microorganisms are ignored.

Generating meta-proteomics and meta-metabolomics data differs significantly from generating metagenomics and metatranscriptomics data, which rely heavily on sequencing (Speda et al., 2017). Concerning meta-proteomics, the complexity of environmental samples still outstrips the capabilities of mass spectrometry approaches (Keiblinger et al., 2016). Thus, protein separation is mandatory to reduce sample complexity before mass spectrometry analysis. 2D gel-based protein separation methods prior to enzymatic digestion can be successfully employed, but these methods have major drawbacks, particularly regarding the analysis of proteins with extreme molecular weights, isoelectric points or hydrophobicity values. Gel-free approaches include different protein extraction procedures, followed by digestion into peptides. Peptides are further separated by reversed-phase liquid chromatography or 2D chromatographic separation using strong cation exchange chromatography in combination with reversed-phase liquid chromatography. Then, peptides are analyzed by mass spectrometry, the measured spectra are compared with theoretical spectra of peptides from a database. Due to its high efficiency and automation, this method is the main strategy for protein identification, quantification and detection. However, only protein sequences represented in the database can be identified.

Identifying and quantifying metabolites is typically carried out using a combination of chromatography techniques (i.e., liquid chromatography, gas chromatography) and detection methods, such as mass spectrometry, and nuclear magnetic resonance (Aguiar-Pulido et al., 2016b). These technologies produce spectra consisting of patterns of peaks that allow metabolite identification and quantification, using comparisons with spectral databases like BioMagResBank, the MassBank, or the Golm Metabolome Database (Aguiar-Pulido et al., 2016b). Prior to this analysis, denoising and peak-picking processes are essential to improve the quality of the treated data.

Data storage and management are a major issue for both approaches because the generated data take valuable space and are barely standardized (Keiblinger et al., 2016; Aguiar-Pulido et al., 2016b). Moreover, meta-proteomics and meta-metabolomics are relatively costly in terms of equipment, and need a high degree of scientific knowledge to compute, analyze and treat

the data. But, contrary to other meta-omics approaches, meta-proteomics and meta-metabolomics are the most direct methods for identifying the functionalities of a natural environment because they work as semi-quantitative methods. Moreover, as these approaches are very sensitive, they can be considered as a good bioindicator of soil quality. Due to their large potential for providing precise information on soil microbial communities, meta-proteomics and meta-metabolomics are currently most advanced among meta-omics approaches.

In parallel to metatranscriptomics, meta-proteomics and meta-metabolomics are generally applied only for specific compartments of food value chains (soils, plants, or transformed products). For example, the composition of soybean leaf metabolites was investigated using metabolomics (Yun et al., 2018) and linked with the soil physicochemical properties. The authors characterized leaf metabolites of 40 samples harvested at two growth stages and from two contrasting sites, with ten replicates *per* condition. They concluded that the soybean leaf metabolome is clearly dependent on the geographical area and influenced by environmental factors such as the soil properties. However, recent studies are a starting point for further, more detailed studies of more compartments of food chains (Mattarozzi et al., 2020). These authors applied meta-proteomics on rhizosphere soils to have a better understanding of the impact of biostimulants on maize seeds. They demonstrated that biostimulants increased the activity of the bacterial community to different extents, and created a permissive and nurturing rhizosphere by stimulating microbial metabolism. As a result, plant growth and root elongation were promoted, and the activity of the bacterial community – especially of species beneficial to plant growth – was increased, without the composition of the microbiota being modified.

CONTRIBUTION OF META-OMICS TO SOIL QUALITY INDICATORS

Issues and Obstacles to Making Meta-Omics Usable for Operational Bioindicators

As previously detailed, meta-omics can be used to characterize the soil microbial communities and evaluate the soil biological quality, but environmental heterogeneity can complexify the analyses. For microbial communities, environmental heterogeneity provides a multitude of habitats at the microscale and macroscale (Fierer, 2017; Karimi et al., 2020). Moreover, soil physico-chemical differences across fields, landscapes, regions and/or continents create an extraordinary pedodiversity that complexifies the study and the survey of soil microbial communities (Hermans et al., 2019). To ensure robust study strategies and avoid the effects of storage time and conditions on the microbial community composition, several international standardized methods (like ISO standards) have been developed to recommend the most accurate sampling method (composite, stratified, systematic, random) and storage conditions of soil samples.

Along with the application of meta-omics methods to characterize microbial diversity in soil samples, several criteria must be verified to lead to an operational bioindicator (**Figure 6**). First, the repeatability, sensitivity and reproducibility levels of the method must be informed and controlled. Moreover, standard operating procedures must be created, evaluated, and tested by independent laboratories (Terrat et al., 2015). The chosen meta-omics procedure must be easily usable as routine analysis (**Supplementary Table 1**).

The understanding and interpretation of any bioindicator of soil quality needs a repository to diagnose the biological status of soils (Lemanceau et al., 2014; Hermans et al., 2017; Jansson and Hofmockel, 2018). Such repositories are used to evaluate whether the measured values are within the range of variations considered as "normal operating range," namely for a given soil type and land use. Data rely on prior descriptive biodiversity studies, the determination of the relationships between microbial abundance and diversity, taxonomic abundance, microbial activities and functioning, and environmental filters (soil, physicochemical, meteorological and spatial parameters). This knowledge of microbial biogeography needs studies at a large spatial scale. A significant number of soil microbial surveys are listed at the national or territory scale (**Figure 7**). Some of these studies aim to monitor soil quality thanks to repositories, and propose modern and operational microbial indicators (**Figure 7**; Rutgers et al., 2009; Griffiths et al., 2011; Powell et al., 2015; Bissett et al., 2016; Horrigue et al., 2016; Terrat et al., 2017; Tresch et al., 2018; George et al., 2019; Ji et al., 2019; Norris et al., 2020). Some are used to measure the impact of land use changes combined with pedoclimatic conditions (Horrigue et al., 2016; Terrat et al., 2017), or of farming practices (Norris et al., 2020) or to evaluate the quality of urban soils (Tresch et al., 2018). The main measured indicators are microbial biomass, microbial diversity (16S, 18S rDNA and ITS markers) using metabarcoding, and more rarely the microbial functional potential through metagenomics (Bissett et al., 2016; Norris et al., 2020). Other research works have sought to understand the spatial distribution and the main environmental drivers, but can be considered as microbial monitoring frameworks for practical conservation of soils (Toju et al., 2016; Hermans et al., 2017; Coller et al., 2019; Jiao and Lu, 2020; Tedersoo et al., 2020; **Figure 7**).

Examples Using Meta-Omics to Assess Soil Quality as a Part of the Food Value Chain

The extensive use of meta-omics approaches to better characterize soil microbial communities has clearly improved our knowledge on soil functioning. However, the integration of -omics data to understand and evaluate all the compartments of the food value chain remain scarce and only some recent studies used it to deal with the microbiological risk assessment or the

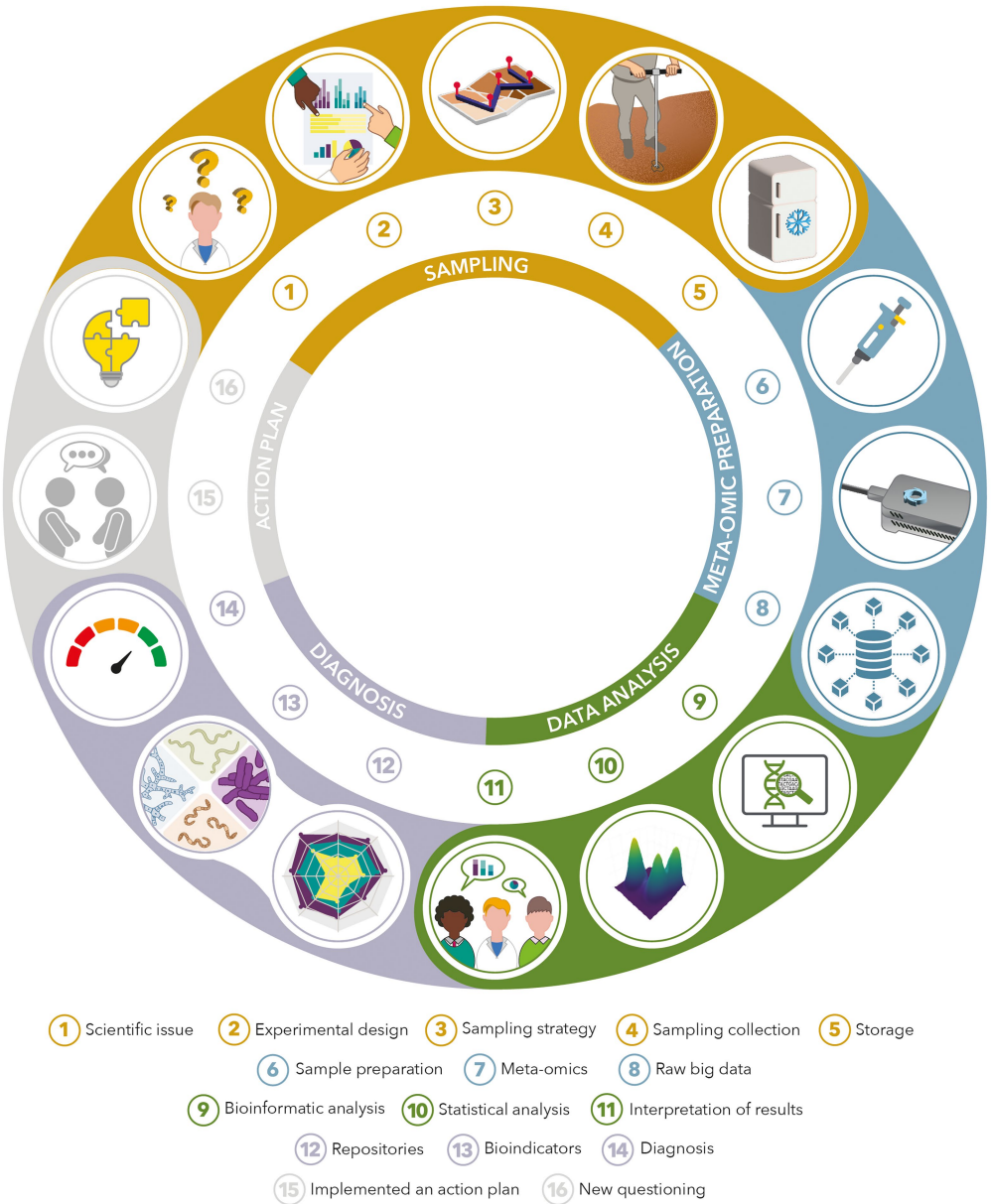


FIGURE 6 | Example of a strategy for studying soil samples with meta-omics approaches: from basic research issues to operational diagnosis.

agroecological transition, with the objective of a high food chain quality.

Impact of Agricultural Practices on Food Security

Soil fertilization with animal-based organic products, such as manure, can cause human pathogens to develop in the soil (Berge et al., 2009; Semenov et al., 2009). One of the greatest challenges for microbiologists is the exploitation of -omics data in microbiological risk assessment (Cocolin et al., 2018). To date, few studies using metagenomics have assessed the pathogenic status of amended soils (Fang et al., 2015; Meneghine et al., 2017). The most dominant human pathogens detected in soil after a long-term application of chicken manure would be *Mycobacterium tuberculosis* and *Mycobacterium ulcerans* (Fang et al., 2015). The study of (Meneghine et al., 2017) used whole metagenomics sequencing approaches to investigate the taxonomic and functional profiles of microbial communities in soil amended with organic compost made from animal carcasses. Bacterial pathogens generally affiliated with fecal contamination were found in low abundance in the soil (less than 0.4% of the total identified bacterial genera), and no toxigeny-related gene or fecal contaminant was detected. More recently, virulence genes and ARGs were searched

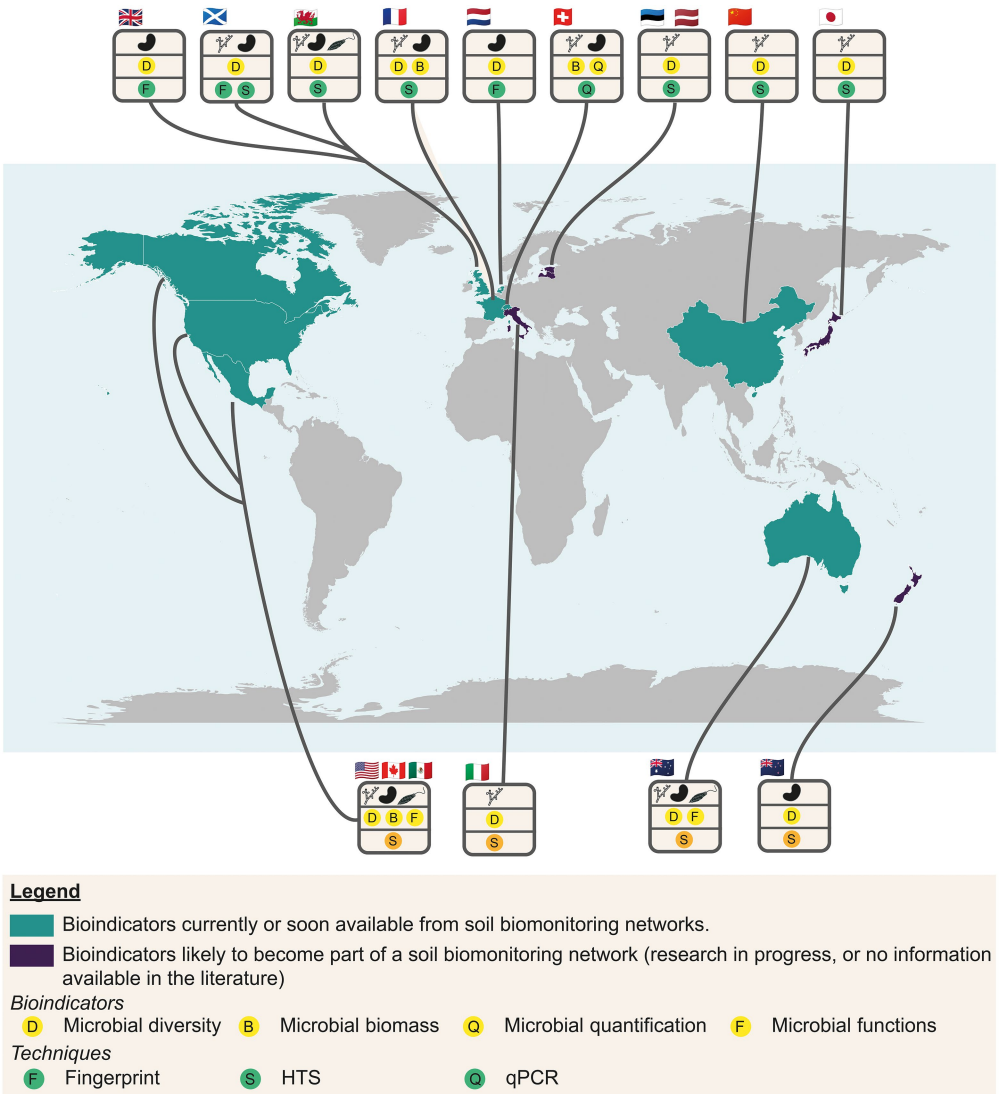


FIGURE 7 | World map showing the main surveys used to monitor the soil quality thanks to repositories based on modern and operational microbial indicators or in progress (turquoise) and studies in the research phase (purple). Microbial indicators are symbolized by circles with a yellow background, and methods by circles with a green background.

for in long-term manure-amended greenhouse soils using metagenomics (Zhang et al., 2020). Manure application led to the accumulation of such virulence and ARGs in soils, and accumulation increased with the frequency of manure application. Such results highlight a risk of widespread ARGs and virulence genes in amended agricultural soils, that can be regulated by mitigating manure application frequency or by pre-treating manure.

From Soil Microbiology to Food Quality at the Territorial Scale

The agroecological transition praised at a global scale, but it operates at local scales where particular environmental, social and economic constraints weigh on agricultural production (Schmitt et al., 2017). Based on this observation, projects are now developing that target focused areas and take into account the whole food value chain, from soils to food quality and consumption, including economic aspects (Magrini et al., 2019).

In France, products with labels of quality of origin (e.g., Protected Designation of Origin, aka. PDO; Protected Geographic Indication; "Label Rouge") are good examples of food value chains in which the quality or the specificity of the products are related to the specificities of the "terroirs" of production, including the local soil properties. Production specifications have long been established independently of environmental issues, hence potential soil degradation. The

question of the relationship between the soil ecological quality and the quality of agricultural products has been regularly suggested in recent years, but it still needs objective scientific demonstrations. The IFEP project was set up to bring new scientific knowledge in this field by focusing on the relationship between soil and product specificity for Comté PDO cheese. It aims at investigating the microbial communities along the continuum between the soil, the grass phyllosphere, cow teats and milk, using a metabarcoding approach and across a network of 44 dairy farms (Chemidlin Prévost-Bouré et al., 2021). The results highlighted strong and significant links between the four compartments of this continuum. These links were modulated by environmental factors: soil pH, plant diversity and elevation, but also farming practices, in particular organic fertilization, cattle density and cow-teat care. Thanks to metabarcoding, the IFEP project provided a proof-of-concept of the importance of the soil microbial quality in the definition of "terroirs," and more largely in the food value chain and its sustainability through microbial transfers from agroecosystems to final products. The next step will be to investigate to what extent integrating soil microbiological quality may improve the sustainability of the food production models, with projects extending this approach to other products and going further up the food value chain. To answer this question, an approach similar to the Living-lab appeared to be most promising and relevant, due to its numerous advantages for innovation in agriculture (Lacoste et al., 2022).

The "Dijon, sustainable food supply 2030" project is another example of the integration of soil quality in the food value chain at a territorial scale. This project is an ambitious and innovative ten-year plan aimed at reaching a fully sustainable agri-food model in the Dijon Métropole territory (Burgundy, France). Even though not obvious, metabarcoding of the soil DNA represents an important asset for this project. It will be implemented on a sampling of 600 points covering the entire 3,500 km^2 of Dijon Metropole's urban and agricultural territories, and will lead to several outputs. From a scientific point of view, the implementation of metabarcoding on a territorial scale will provide more insights into the drivers of the spatial distribution and regulation of the soil microbial diversity. From a more operational point of view, it will lead to building up the first territorial repository of the soil microbial quality in terms of microbial biomass and biodiversity. Such a repository will serve to implement a robust and sensitive diagnosis of the soil microbial quality by the actors and stakeholders involved in the use of soils (stakeholders of biomonitoring programs, farmers, land managers, agri-food industries, and consumers). This is a pre-requisite for combining food and the environmental quality of future farming systems developed in the context of the agroecological transition. Indicators and associated repositories will also be used to implement innovative labels for agricultural products, based on performance obligations rather than just best endeavors obligations. Such labels will inform the consumers about the effective environmental fingerprint of the agricultural products. They will also help move toward the development of local food systems. In addition, based on diagnosis at the territorial scale, the most sustainable farming systems and practices will be identified and communicated to the stakeholders involved in the use of soils, whether urban or rural. Soil quality indicators obtained from metabarcoding will also provide bases for soil value estimates integrating the intrinsic ecological state, which has not been taken into account to date. Ultimately, the implications of this shift of paradigm will be huge in terms of soil conservation, legal terms and economic aspects. Altogether, this project illustrates to what extent metabarcoding can be a powerful tool to help in the transition of urban and rural spaces toward more sustainable systems. As a pilot project, it is aimed at being replicated in other territories in order to supply collectivities and soil users with tools for measuring the impact of their activities on the soil quality and promoting sustainable systems.

ONGOING AND FUTURE DEVELOPMENTS

Recent Advanced Technologies

New technological developments frequently emerge in microbiology. Adapted from the Lawrence Livermore Microbial Detection Array (LLMDA), the Axiom Microbiome Array (AMA) is a high-throughput platform that identifies species from environmental samples. This microarray contains 1.38 million DNA probes that can detect just over 12,000 species of viruses, archaea and bacteria, fungi, and protozoa, and probably represents the most thorough method (Thissen et al., 2019).

New platforms devoted to third-generation sequencing have emerged. They provide long reads and avoid PCR steps for some technologies (Tedersoo et al., 2021). Although these technologies are increasingly used in microbial ecology research, they are not yet available for the microbial diagnosis of soils. Nevertheless, they have proved useful to characterize microbial communities in a context of diagnosis in plant pathology (Tedersoo et al., 2019). The small and large subunits of the rRNA operon can be sequenced in full length (Tedersoo et al., 2018), improving the species-level identification of prokaryote and eukaryote organisms (Benítez-Páez et al., 2016; Singer et al., 2016) and *in fine* diagnosis. The feasibility of the method was demonstrated by the Oxford Nanopore MinION TM system and PacBio sequencing with environmental samples for metabarcoding (Schloss et al., 2016; Kerkhof, 2021) and metagenomics (Van Goethem et al., 2021) studies. Real-time DNA sequencing coupled with bioinformatic analysis also appears very promising to carry out a rapid assessment of biodiversity directly in the field and might represent a reliable bio-monitoring tool in the years to come (Pomerantz et al., 2018; Krehenwinkel et al., 2019; Maestri et al., 2019; Albrecht et al., 2020).

Recent meta-omics approaches are essential to better understand the soil microbiome, but it would be interesting

to also explore the microorganisms of relevance for improving the soil quality and food safety. Although significant improvements have been made in culturing so-far unculturable microorganisms, finding an appropriate culture medium remains difficult. High-throughput culturing platforms have emerged to isolate and screen a large number of microbial cells (Mafla-Endara et al., 2021; Trivedi et al., 2021). The ability to isolate single cells using microfluidics coupled with technologies amplifying genomic DNA from single cells will revolutionize the study of unculturable species and of microheterogeneity within species (Chen et al., 2021). As recently stated, microfluidics has a great potential for experimenting on the soil microbiome so as to determine the mechanisms underpinning specific microbial metabolic interactions and to understand and predict the influence of environmental gradients on specific microbial functions with precise spatiotemporal control (Jansson and Hofmockel, 2018). The latest data indicated that droplet-based microfluidic systems compartmentalize single microorganisms in droplets, detect and sort the droplets at a rate of up to 1,000 *per* second and screen ~1 million microorganisms *per* hour (Watterson et al., 2020).

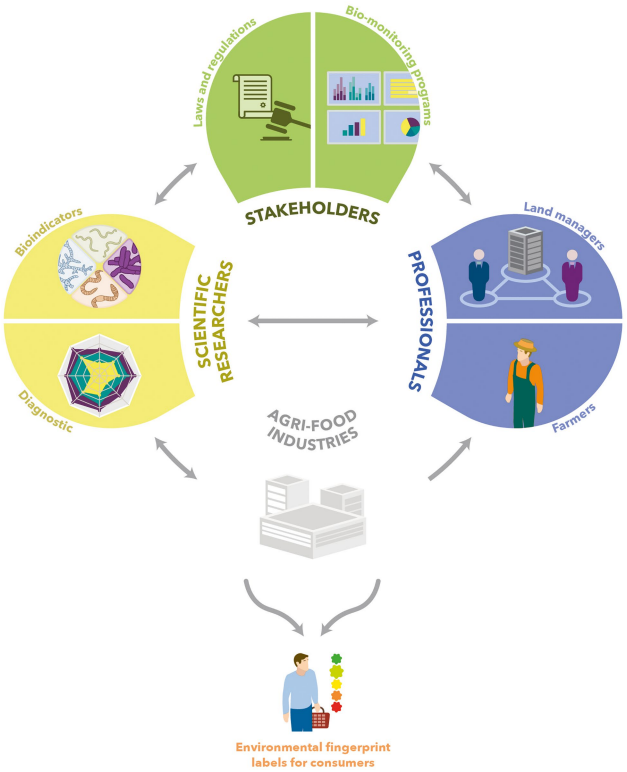


FIGURE 8 | Schematic overview of field soil quality measurement and its actors. Scientific researchers explore microbial diversity in order to create repositories and develop bioindicators. These bioindicators are forwarded to the agri-food industry to make them largely accessible to professionals such as farmers or land managers. In addition, exchanges with stakeholders of bio-monitoring programs take place.

Old Data for Future Prospects, New Perspectives

Over the last decade, multiple tools or databases have been developed to maximize the use of meta-omics results. Major advances have been made in metabarcoding research, ands have provided functional information (Djemiel et al., 2022). Functional inference and functional trait assignments can be carried out in addition to the characterization of microbial diversity. Functional inference based on phylogenetic models and genomic data predicts putative bacterial and fungal functions (Douglas et al., 2020). One of the advantages is to go beyond the metagenomics approach while saving time and money. However, the major limitation is that the robustness of the results is strongly linked to amplicon resolution (corresponding to metabarcoding) and the availability of microbial reference genomes (corresponding to inference). Further tools and databases can be used to supplement microbial diversity results with trait and character data (Nguyen et al., 2016; Põlme et al., 2020). For example, based on taxonomic classification, fungal genera/species can be associated with their lifestyles, body types or habitats. Another alternative way to evaluate the quality and functioning of soil microbial ecosystems is biotic interaction network analysis (Barberán et al., 2012; Layeghifard et al., 2017). Based on the analysis of taxonomic co-abundance patterns, the topology of each network highlights the relationship between taxa – positive in the case of co-occurrence, or negative in the case of co-exclusion (Karimi et al., 2017). Exploring the keystone taxa or guilds highly connected whatever their abundance will provide a complementary approach and detect taxa with a crucial role in the soil microbiota structure and functioning (Banerjee et al., 2018), which can be modified by soil disturbances linked to agricultural intensification (Banerjee et al., 2019; Karimi et al., 2019).

CONCLUSION

Global awareness is required for a greater consideration of microbial indicators for monitoring soil biodiversity and achieving the ambition of a sustainable food value chain. This involves a wide-scale collaboration of governments, local authorities, stakeholders of bio-monitoring programs, farmers, land managers, agri-food industries, and consumers (Lehmann et al., 2020; Fierer et al., 2021; **Figure 8**). Microbial indicators for soil quality monitoring by scientific researchers (Fierer et al., 2021) and routine use by field professionals (**Figure 8**) need to be developed and validated. Although studying microbial diversity used to be difficult owing to technical constraints, a broad range of approaches have emerged in the past few years or decades (Yang et al., 2020). Meta-omics techniques are already being used to study microbial communities along the food production chain, but insufficiently integrate the soil microbiome (Cocolin et al., 2018; Yap et al., 2022). This holistic vision is essential to offer high-value bioindicators. Moreover,

robust repositories are badly needed for presently existing soil monitoring networks (Yang et al., 2020). In light of these elements, meta-omics has provided and will keep on providing modern microbial indicators for monitoring the soil biodiversity and securing food production and quality, and the related environmental footprint with the creation of labels for the consumers. This ambition to achieve sustainable food systems also applies to urban environments (Vargas-Hernández, 2020).

AUTHOR CONTRIBUTIONS

CD and ST conceptualized the manuscript and drafted the manuscript with contributions from SD, BK, AC, WH, ArB, AlB, SS-B, P-AM, NCP-B, and LR. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank Alexandrine Thore for graphical design in **Figures 1**, **3**, **6**, and **8**. They also thank Annie Buchwalter for correcting and improving English language in the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.889788/full#supplementary-material

Supplementary Table 1 | Overview of the various meta-omics approaches and their operationality for soil quality biomonitoring. R: In research phase, P: Provided by agri-food industries, XXX: Adapted technique with a very good data / price ratio, XX: Adapted technique with a moderate data / price ratio, X: Adapted technique with a low data/price ratio.

REFERENCES

Abot, A., Arnal, G., Auer, L., Lazuka, A., Labourdette, D., Lamarre, S., et al. (2016). CAZyChip: dynamic assessment of exploration of glycoside hydrolases in microbial ecosystems. *BMC Genomics* 17:671. doi: 10.1186/s12864-016-2988-4

Afshari, R., Pillidge, C. J., Dias, D. A., Osborn, A. M., and Gill, H. (2020). Cheesomics: the future pathway to understanding cheese flavour and quality. *Crit. Rev. Food Sci. Nutr.* 60, 33–47. doi: 10.1080/10408398.2018.1512471

Aguiar-Pulido, V., Huang, W., Suarez-ulloa, V., Cickovski, T., Mathee, K., and Narasimhan, G. (2016a). Approaches for microbiome analysis. *Evol. Bioinforma.* 12, 5–16. doi: 10.4137/EBO.S36436.TYPE

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., and Narasimhan, G. (2016b). Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. *Evol. Bioinforma.* 12, 5–16. doi: 10.4137/EBO.S36436

Albrecht, B., Bağcl, C., and Huson, D. H. (2020). MAIRA- real-time taxonomic and functional analysis of long reads on a laptop. *BMC Bioinformatics* 21:390. doi: 10.1186/s12859-020-03684-2

Astudillo-García, C., Hermans, S. M., Stevenson, B., Buckley, H. L., and Lear, G. (2019). Microbial assemblages and bioindicators as proxies for ecosystem health status: potential and limitations. *Appl. Microbiol. Biotechnol.* 103, 6407–6421. doi: 10.1007/s00253-019-09963-0

Banerjee, S., Schlaeppi, K., and van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nat. Rev. Microbiol.* 16, 567–576. doi: 10.1038/s41579-018-0024-1

Banerjee, S., Walder, F., Büchi, L., Meyer, M., Held, A. Y., Gattinger, A., et al. (2019). Agricultural intensification reduces microbial network complexity and the abundance of keystone taxa in roots. *ISME J.* 13, 1722–1736. doi: 10.1038/s41396-019-0383-2

Banos, S., Lentendu, G., Kopf, A., Wubet, T., Glöckner, F. O., and Reich, M. (2018). A comprehensive fungi-specific 18S rRNA gene sequence primer toolkit suited for diverse research issues and sequencing platforms. *BMC Microbiol.* 18:190. doi: 10.1186/s12866-018-1331-4

Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi: 10.1038/ismej.2011.119

Barbi, F., Bragalini, C., Vallon, L., Prudent, E., Dubost, A., Fraissinet-Tachet, L., et al. (2014). PCR primers to study the diversity of expressed fungal genes encoding lignocellulolytic enzymes in soils using high-throughput sequencing. *PLoS One* 9:e116264. doi: 10.1371/journal.pone.0116264

Bardgett, R. D., and Van Der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature* 515, 505–511. doi: 10.1038/nature13855

Béné, C., Prager, S. D., Achicanoy, H. A. E., Toro, P. A., Lamotte, L., Bonilla, C., et al. (2019). Global map and indicators of food system sustainability. *Sci. Data* 6:279. doi: 10.1038/s41597-019-0301-5

Benítez-Páez, A., Portune, K. J., and Sanz, Y. (2016). Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. *Gigascience* 5, 1–9. doi: 10.1186/s13742-016-0111-z

Berge, A. C. B., Glanville, T. D., Millner, P. D., and Klingborg, D. J. (2009). Methods and microbial risks associated with composting of animal carcasses in the United States. *J. Am. Vet. Med. Assoc.* 234, 47–56. doi: 10.2460/javma.234.1.47

Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P. M., Reith, F., Dennis, P. G., et al. (2016). Introducing BASE: The biomes of Australian soil environments soil microbial diversity database. *Gigascience* 5:s13742-016. doi: 10.1186/s13742-016-0126-5

Bloem, J., Hopkins, D. W., and Benedetti, A. (2005). *Microbiological Methods for Assessing Soil Quality*. United Kingdom: CABI.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., and Zemskaya, T. I. (2019). The effect of 16s rRNA region choice on bacterial community metabarcoding results. *Sci. Data* 6, 1–14. doi: 10.1038/sdata.2019.7

Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., De Goede, R., et al. (2018). Soil quality – A critical review. *Soil Biol. Biochem.* 120, 105–125. doi: 10.1016/j.soilbio.2018.01.030

Castaño, C., Berlin, A., Brandström Durling, M., Ihrmark, K., Lindahl, B. D., Stenlid, J., et al. (2020). Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities. *New Phytol.* 228:nph.16731. doi: 10.1111/nph.16731

Chemidlin Prévost-Bouré, N., Karimi, B., Sadet-Bourgeteau, S., Djemiel, C., Brie, M., Dumont, J., et al. (2021). Microbial transfers from permanent grassland ecosystems to milk in dairy farms in the Comté cheese area. *Sci. Rep.* 11:18144. doi: 10.1038/s41598-021-97373-6

Chen, Q. L., Hu, H. W., He, Z. Y., Cui, L., Zhu, Y. G., and He, J. Z. (2021). Potential of indigenous crop microbiomes for sustainable agriculture. *Nat. Food* 2, 233–240. doi: 10.1038/s43016-021-00253-5

Chesworth, W. (2007). *Encyclopedia of soil Science*. Germany: Springer Science.

Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M. F., Jagadeesan, B., et al. (2018). Next generation microbiological risk assessment meta-omics: The next need for integration. *Int. J. Food Microbiol.* 287, 10–17. doi: 10.1016/j.ijfoodmicro.2017.11.008

Coller, E., Cestaro, A., Zanzotti, R., Bertoldi, D., Pindo, M., Larger, S., et al. (2019). Microbiome of vineyard soils is shaped by geography and management. *Microbiome* 7:140. doi: 10.1186/s40168-019-0758-7

Cui, H., Wang, C., Gu, Z., Zhu, H., Fu, S., and Yao, Q. (2014). Evaluation of soil storage methods for soil microbial community using genetic and

metabolic fingerprintings. *Eur. J. Soil Biol.* 63, 55–63. doi: 10.1016/j.ejsobi.2014.05.006

Delmont, T. O., Robe, P., Clark, I., Simonet, P., and Vogel, T. M. (2011). Metagenomic comparison of direct and indirect soil DNA extraction approaches. *J. Microbiol. Methods* 86, 397–400. doi: 10.1016/j.mimet.2011.06.013

Dickie, I. A., Boyer, S., Buckley, H., Duncan, R. P., Gardner, P., Hogg, I. D., et al. (2018). Towards robust and repeatable sampling methods in eDNA based studies. *Mol. Ecol. Resour.* 5, 940–952. doi: 10.1111/1755-0998.12907

Djemiel, C., Dequiedt, S., Karimi, B., Cottin, A., Girier, T., El Djoudi, Y., et al. (2020a). BIOCOM-PIPE: a new user-friendly metabarcoding pipeline for the characterization of microbial diversity from 16S, 18S and 23S rRNA gene amplicons. *BMC Bioinformatics* 21:492. doi: 10.1186/s12859-020-03829-3

Djemiel, C., Maron, P., Terrat, S., Dequiedt, S., Cottin, A., and Ranjard, L. (2022). Inferring microbiota functions from taxonomic genes: a review. *Gigascience* 11, 1–30. doi: 10.1093/gigascience/giab090

Djemiel, C., Plassard, D., Terrat, S., Crouzet, O., Sauze, J., Mondy, S., et al. (2020b). µ green-db: a reference database for the 23S rRNA gene of eukaryotic plastids and cyanobacteria. *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-62555-1

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., et al. (2020). PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* 38, 685–688. doi: 10.1038/s41587-020-0548-6

Duru, I. C., Laine, P., Andreevskaya, M., Paulin, L., Kananen, S., Tynkkynen, S., et al. (2018). Metagenomic and metatranscriptomic analysis of the microbial community in Swiss-type Maasdam cheese during ripening. *Int. J. Food Microbiol.* 281, 10–22. doi: 10.1016/j.ijfoodmicro.2018.05.017

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., et al. (2018). FROGS: find, rapidly, OTUs with galaxy solution. *Bioinformatics* 34, 1287–1294. doi: 10.1093/bioinformatics/btx791

Evans, D. L., Quinton, J. N., Tye, A. M., Rodés, Á., Davies, J. A. C., Mudd, S. M., et al. (2019). Arable soil formation and erosion: a hillslope-based cosmogenic nuclide study in the United Kingdom. *Soil* 5, 253–263. doi: 10.5194/soil-5-253-2019

Fang, H., Wang, H., Cai, L., and Yu, Y. (2015). Prevalence of antibiotic resistance genes and bacterial pathogens in long-term manured greenhouse soils As revealed by metagenomic survey. *Environ. Sci. Technol.* 49, 1095–1104. doi: 10.1021/es504157v

FAO, ITPS, GSBI, and SCBD (2020). *State of Knowledge of Soil Biodiversity - Status, Challenges and Potentialities*. Rome, Italy: FAO. 618.

FAOSTAT (2021). Food and Agriculture Organization of the United Nations (FAO). Available at: https://www.fao.org/faostat/en/#home

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* 15, 579–590. doi: 10.1038/nrmicro.2017.87

Fierer, N., Wood, S. A., and Bueno de Mesquita, C. P. (2021). How microbes can, and cannot, be used to assess soil health. *Soil Biol. Biochem.* 153:108111. doi: 10.1016/j.soilbio.2020.108111

Frac, M., Hannula, S. E., Belka, M., and Jędryczka, M. (2018). Fungal biodiversity and their role in soil health. *Front. Microbiol.* 9:707. doi: 10.3389/fmicb.2018.00707

Friedrichsen, C. N., Daroub, S. H., Monroe, M. C., Stepp, J. R., and Gerber, S. (2019). Stakeholders' mental models of soil food value chain in the Everglades. *Geoderma* 343, 166–175. doi: 10.1016/j.geoderma.2019.02.037

Geisen, S., Mitchell, E. A. D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., et al. (2018). Soil protists: a fertile frontier in soil biology research. *FEMS Microbiol. Rev.* 42, 293–323. doi: 10.1093/femsre/fuy006

Geisen, S., Wall, D. H., and van der Putten, W. H. (2019). Challenges and opportunities for soil biodiversity in the Anthropocene. *Curr. Biol.* 29, R1036–R1044. doi: 10.1016/j.cub.2019.08.007

George, P. B. L., Lallias, D., Creer, S., Seaton, F. M., Kenny, J. G., Eccles, R. M., et al. (2019). Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nat. Commun.* 10, 1–11. doi: 10.1038/s41467-019-09031-1

Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351. doi: 10.1038/nrg.2016.49

Griffiths, B. S., Faber, J., and Bloem, J. (2018). Applying soil health indicators to encourage sustainable soil use: The transition from scientific study to practical application. *Sustainability* 10:3021. doi: 10.3390/su10093021

Griffiths, R. I., Thomson, B. C., James, P., Bell, T., Bailey, M., and Whiteley, A. S. (2011). The bacterial biogeography of British soils. *Environ. Microbiol.* 13, 1642–1654. doi: 10.1111/j.1462-2920.2011.02480.x

Gutiérrez-Aguirre, I., Rački, N., Dreo, T., and Ravnikar, M. (2015). "Droplet digital PCR for absolute quantification of pathogens," in *Plant Pathology* (New York, NY: Humana Press), 331–347.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., and Troedsson, C. (2014). Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. *PLoS One* 9:e87624. doi: 10.1371/journal.pone.0087624

Hambraeus, G., Von Wachenfeldt, C., and Hederstedt, L. (2003). Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol. Gen. Genomics.* 269, 706–714. doi: 10.1007/s00438-003-0883-6

Hartemink, A. E., Zhang, Y., Bockheim, J. G., Curi, N., Silva, S. H. G., Grauer-Gray, J., et al. (2020). Soil horizon variation: A review. *Adv. Agron.* 160, 125–185. doi: 10.1016/bs.agron.2019.10.003

Hayatsu, M., Tago, K., and Saito, M. (2008). Various players in the nitrogen cycle: diversity and functions of the microorganisms involved in nitrification and denitrification. *Soil Sci. Plant Nutr.* 54, 33–45. doi: 10.1111/j.1747-0765.2007.00195.x

He, Z., Gentry, T. J., Schadt, C. W., Wu, L., Liebich, J., Chong, S. C., et al. (2007). GeoChip: A comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* 1, 67–77. doi: 10.1038/ismej.2007.2

Hermans, S. M., Buckley, H. L., Case, B. S., Curran-Cournane, F., Taylor, M., and Lear, G. (2017). Bacteria as emerging indicators of soil condition. *Appl. Environ. Microbiol.* 83:e02826. doi: 10.1128/AEM.02826-16

Hermans, S. M., Buckley, H. L., and Lear, G. (2019). Perspectives on the impact of sampling design and intensity on soil microbial diversity estimates. *Front. Microbiol.* 10:1820. doi: 10.3389/fmicb.2019.01820

Hofstetter, V., Buyck, B., Eyssartier, G., Schnee, S., and Gindro, K. (2019). The unbearable lightness of sequenced-based identification. *Fungal Divers.* 96, 243–284. doi: 10.1007/s13225-019-00428-3

Holmes, D. E., Nevin, K. P., and Lovley, D. R. (2004). Comparison of 16S rRNA, nifD, recA, gyrB, rpoB and fusA genes within the family Geobacteraceae fam. Nov. *Int. J. Syst. Evol. Microbiol.* 54, 1591–1599. doi: 10.1099/ijs.0.02958-0

Horrigue, W., Dequiedt, S., Chemidlin Prévost-Bouré, N., Jolivet, C., Saby, N. P. A., Arrouays, D., et al. (2016). Predictive model of soil molecular microbial biomass. *Ecol. Indic.* 64, 203–211. doi: 10.1016/j.ecolind.2015.12.004

Jansson, J. K., and Hofmockel, K. S. (2018). The soil microbiome — from metagenomics to metaphenomics. *Curr. Opin. Microbiol.* 43, 162–168. doi: 10.1016/j.mib.2018.01.013

Ji, N., Gao, C., Sandel, B., Zheng, Y., Chen, L., Wu, B., et al. (2019). Late Quaternary climate change explains soil fungal community composition rather than fungal richness in forest ecosystems. *Ecol. Evol.* 9, 6678–6692. doi: 10.1002/ece3.5247

Jiao, S., and Lu, Y. (2020). Abundant fungi adapt to broader environmental gradients than rare fungi in agricultural fields. *Glob. Chang. Biol.* 26, 4506–4520. doi: 10.1111/gcb.15130

Karimi, B., Dequiedt, S., Terrat, S., Jolivet, C., Arrouays, D., Wincker, P., et al. (2019). Biogeography of soil bacterial networks along a gradient of cropping intensity. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-40422-y

Karimi, B., Maron, P. A., Bouré, C. P., Bernard, N., Gilbert, D., and Ranjard, L. (2017). Microbial diversity and ecological networks as indicators of environmental quality. *Environ. Chem. Lett.* 15, 265–281. doi: 10.1007/s10311-017-0614-6

Karimi, B., Villerd, J., Dequiedt, S., Terrat, S., Chemidlin-Prévost Bouré, N., Djemiel, C., et al. (2020). Biogeography of soil microbial habitats across France. *Glob. Ecol. Biogeogr.* 29, 1399–1411. doi: 10.1111/geb.13118

Keiblinger, K. M., Fuchs, S., Zechmeister-Boltenstern, S., and Riedel, K. (2016). Soil and leaf litter metaproteomics-A brief guideline from sampling to understanding. *FEMS Microbiol. Ecol.* 92, 1–18. doi: 10.1093/femsec/iw180

Kerkhof, L. J. (2021). Is Oxford Nanopore sequencing ready for analyzing complex microbiomes? *FEMS Microbiol. Ecol.* 97:fiab001. doi: 10.1093/femsec/fiab001

Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M., and McCue, L. A. (2020). ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinform.* 21:257. doi: 10.1186/s12859-020-03585-4

Kopittke, P. M., Berhe, A. A., Carrillo, Y., Cavagnaro, T. R., Chen, D., Chen, Q.-L., et al. (2022). Ensuring planetary survival: the centrality of organic carbon

in balancing the multifunctional nature of soils. *Crit. Rev. Environ. Sci. Technol.* 1–17. doi: 10.1080/10643389.2021.2024484

Kopittke, P. M., Menzies, N. W., Wang, P., McKenna, B. A., and Lombi, E. (2019). Soil and the intensification of agriculture for global food security. *Environ. Int.* 132:105078. doi: 10.1016/j.envint.2019.105078

Krehenwinkel, H., Pomerantz, A., and Prost, S. (2019). Genetic biomonitoring and biodiversity assessment using portable sequencing technologies: current uses and future directions. *Genes* 10:858. doi: 10.3390/genes10110858

Kultima, J. R., Coelho, L. P., Forslund, K., Huerta-Cepas, J., Li, S. S., Driessen, M., et al. (2016). MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* 32, 2520–2523. doi: 10.1093/bioinformatics/btw183

Lacoste, M., Cook, S., McNee, M., Gale, D., Ingram, J., Bellon-Maurel, V., et al. (2022). On-farm experimentation to transform global agriculture. *Nat. Food* 3, 11–18. doi: 10.1038/s43016-021-00424-4

Layeghifard, M., Hwang, D. M., and Guttman, D. S. (2017). Disentangling interactions in the microbiome: A network perspective. *Trends Microbiol.* 25, 217–228. doi: 10.1016/j.tim.2016.11.008

Lee, Y. J., Van Nostrand, J. D., Tu, Q., Lu, Z., Cheng, L., Yuan, T., et al. (2013). The PathoChip, a functional gene array for assessing pathogenic properties of diverse microbial communities. *ISME J.* 7, 1974–1984. doi: 10.1038/ismej.2013.88

Lehmann, J., Bossio, D. A., Kögel-Knabner, I., and Rillig, M. C. (2020). The concept and future prospects of soil health. *Nat. Rev. Earth Environ.* 1, 544–553. doi: 10.1038/s43017-020-0080-8

Lemanceau, P., Maron, P. A., Mazurier, S., Mougel, C., Pivato, B., Plassart, P., et al. (2014). Understanding and managing soil biodiversity: a major challenge in agroecology. *Agron. Sustain. Dev.* 35, 67–81. doi: 10.1007/s13593-014-0247-0

Maestri, C., Paterno, F., and Garces, M. (2019). A rapid and accurate MinION-based workflow for tracking species biodiversity in the field. *Genes* 10:468. doi: 10.3390/genes10060468

Mafla-Endara, P. M., Arellano-Caicedo, C., Aleklett, K., Pucetaite, M., Ohlsson, P., and Hammer, E. C. (2021). Microfluidic chips provide visual access to in situ soil ecology. *Commun. Biol.* 4:889. doi: 10.1038/s42003-021-02379-5

Magrini, M.-B., Martin, G., Magne, M.-A., Duru, M., Couix, N., Hazard, L., et al. (2019). "Agroecological transition from farms to territorialised Agri-food systems: issues and drivers," in *Agroecological Transitions: From Theory to Practice in Local Participatory Design*. eds. J. E. Bergez, E. Audouin and O. Therond (Cham: Springer International Publishing), 69–98.

Maron, P.-A., Mougel, C., and Ranjard, L. (2011). Soil microbial diversity: methodological strategy, spatial overview and functional interest. *C. R. Biol.* 334, 403–411. doi: 10.1016/j.crvi.2010.12.003

Maron, P.-A., Sarr, A., Kaisermann, A., Lévêque, J., Mathieu, O., Guigue, J., et al. (2018). High microbial diversity promotes soil ecosystem functioning. *Appl. Environ. Microbiol.* 84:e02738-17. doi: 10.1128/AEM.02738-17

Martinez-Salgado, M. M., Gutierrez-Romero, V., Jannsens, M., and Ortega-Blu, R. (2010). Biological soil quality indicators: a review. *Curr. Res. Technol. Educ. Top. Appl. Microbiol. Microb. Biotechnol.* 319–328.

Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J. C., Soulas, G., et al. (2001). DNA extraction from soils: old Bias for new microbial diversity analysis methods. *Appl. Environ. Microbiol.* 67, 2354–2359. doi: 10.1128/AEM.67.5.2354-2359.2001

Mattarozzi, M., Di Zinno, J., Montanini, B., Manfredi, M., Marengo, E., Fornasier, F., et al. (2020). Biostimulants applied to maize seeds modulate the enzymatic activity and metaproteome of the rhizosphere. *Appl. Soil Ecol.* 148:103480. doi: 10.1016/j.apsoil.2019.103480

Mehle, N., and Dreo, T. (2019). "Quantitative analysis with droplet digital PCR," in *Notes on the Greek Text of Genesis*. ed. J. W. Wevers (United States: SBL Press).

Meneghine, A. K., Nielsen, S., Varani, A. M., Thomas, T., and Carareto Alves, L. M. (2017). Metagenomic analysis of soil and freshwater from zoo agricultural area with organic fertilization. *PLoS One* 12:e0190178. doi: 10.1371/journal.pone.0190178

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386

Mukherjee, A., and Reddy, M. S. (2020). Metatranscriptomics: an approach for retrieving novel eukaryotic genes from polluted and related environments. *Biotech* 10:71. doi: 10.1007/s13205-020-2057-1

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., et al. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* 20, 241–248. doi: 10.1016/j.funeco.2015.06.006

Nielsen, U. N., Wall, D. H., and Six, J. (2015). Soil biodiversity and the environment. *Annu. Rev. Environ. Resour.* 40, 63–90. doi: 10.1146/annurev-environ-102014-021257

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., and Tedersoo, L. (2019a). Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nat. Rev. Microbiol.* 17, 95–109. doi: 10.1038/s41579-018-0116-y

Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., et al. (2019b). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47, D259–D264. doi: 10.1093/nar/gky1022

Nkongolo, K. K., and Narendrula-Kotha, R. (2020). Advances in monitoring soil microbial community dynamic and function. *J. Appl. Genet.* 61, 249–263. doi: 10.1007/s13353-020-00549-5

Norris, C. E., Bean, G. M., Cappellazzi, S. B., Cope, M., Greub, K. L. H., Liptzin, D., et al. (2020). Introducing the north American project to evaluate soil health measurements. *Agron. J.* 112, 3195–3215. doi: 10.1002/agj2.20234

Ogier, J.-C., Pagès, S., Galan, M., Barret, M., and Gaudriault, S. (2019). rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiol.* 19:171. doi: 10.1186/s12866-019-1546-z

Oshiki, M., Segawa, T., and Ishii, S. (2018). Nitrogen cycle evaluation (NiCE) Chip for simultaneous analysis of multiple N cycle-associated genes. *Appl. Environ. Microbiol.* 84:e02615-17. doi: 10.1128/AEM.02615-17

Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., et al. (2019). Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecol.* 41, 23–33. doi: 10.1016/j.funeco.2019.03.005

Phillips, H. R. P., Bach, E. M., Bartz, M. L. C., Bennett, J. M., Beugnon, R., Briones, M. J. I., et al. (2021). Global data on earthworm abundance, biomass, diversity and corresponding environmental properties. *Sci. Data* 8:136. doi: 10.1038/s41597-021-00912-z

Poirier, S., Rué, O., Peguilhan, R., Coeuret, G., Zagorec, M., Champomier-Vergès, M. C., et al. (2018). Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using gyrB amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing. *PLoS One* 13:e0204629. doi: 10.1371/journal.pone.0204629

Põlme, S., Abarenkov, K., Henrik Nilsson, R., Lindahl, B. D., Clemmensen, K. E., Kauserud, H., et al. (2020). FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Divers.* 105, 1–16. doi: 10.1007/s13225-020-00466-2

Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., et al. (2018). Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *Gigascience* 7, 1–14. doi: 10.1093/gigascience/giy033

Porter, T. M., and Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Mol. Ecol.* 27, 313–338. doi: 10.1111/mec.14478

Poveda, J. (2021). Beneficial effects of microbial volatile organic compounds (MVOCs) in plants. *Appl. Soil Ecol.* 168:104118. doi: 10.1016/j.apsoil.2021.104118

Powell, J. R., Karunaratne, S., Campbell, C. D., Yao, H., Robinson, L., and Singh, B. K. (2015). Deterministic processes vary during community assembly for ecologically dissimilar taxa. *Nat. Commun.* 6:8444. doi: 10.1038/ncomms9444

Prosser, J. I. (2015). Dispersing misconceptions and identifying opportunities for the use of "omics" in soil microbial ecology. *Nat. Rev. Microbiol.* 13, 439–446. doi: 10.1038/nrmicro3468

Prudent, M., Dequiedt, S., Sorin, C., Girodet, S., Nowak, V., Duc, G., et al. (2020). The diversity of soil microbial communities matters when legumes face drought. *Plant Cell Environ.* 43, 1023–1035. doi: 10.1111/pce.13712

Pulleman, M., Creamer, R., Hamer, U., Helder, J., Pelosi, C., Pérès, G., et al. (2012). Soil biodiversity, biological indicators and soil ecosystem services-an overview of European approaches. *Curr. Opin. Environ. Sustain.* 4, 529–538. doi: 10.1016/j.cosust.2012.10.009

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data

processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Ranjard, L., Dequiedt, S., Lelievre, M., Maron, P. A., Mougel, C., Morin, F., et al. (2009). Platform GenoSol: a new tool for conserving and exploring soil microbial diversity. *Environ. Microbiol. Rep.* 1, 97–99. doi: 10.1111/j.1758-2229.2009.00023.x

Rincon-Florez, V., Carvalhais, L., and Schenk, P. (2013). Culture-independent molecular tools for soil and rhizosphere microbiology. *Diversity* 5, 581–612. doi: 10.3390/d5030581

Rutgers, M., Schouten, A. J., Bloem, J., Van Eekeren, N., De Goede, R. G. M., Jagersop Akkerhuis, G. A. J. M., et al. (2009). Biological measurements in a nationwide soil monitoring network. *Eur. J. Soil Sci.* 60, 820–832. doi: 10.1111/j.1365-2389.2009.01163.x

Schloss, P. D., Jenior, M. L., Koumpouras, C. C., Westcott, S. L., and Highlander, S. K. (2016). Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* 4:e1869. doi: 10.7717/peerj.1869

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Schloter, M., Nannipieri, P., Sørensen, S. J., and Van Elsas, J. D. (2018). Microbial indicators for soil quality. *Biol. Fertil. Soils* 54, 1–10. doi: 10.1007/s00374-017-1248-3

Schmitt, E., Galli, F., Menozzi, D., Maye, D., Touzard, J.-M., Marescotti, A., et al. (2017). Comparing the sustainability of local and global food products in Europe. *J. Clean. Prod.* 165, 346–359. doi: 10.1016/j.jclepro.2017.07.039

Schulz-Bohm, K., Geisen, S., Wubs, E. R. J., Song, C., de Boer, W., and Garbeva, P. (2017). The prey's scent – volatile organic compound mediated interactions between soil bacteria and their protist predators. *ISME J.* 11, 817–820. doi: 10.1038/ismej.2016.144

Semenov, M. V. (2021). Metabarcoding and metagenomics in soil ecology research: achievements, challenges, and prospects. *Biol. Bull. Rev.* 11, 40–53. doi: 10.1134/S2079086421010084

Semenov, A. V., Van Overbeek, L., and Van Bruggen, A. H. C. (2009). Percolation and survival of *Escherichia coli* O157:H7 and *Salmonella enterica* serovar typhimurium in soil amended with contaminated dairy manure or slurry. *Appl. Environ. Microbiol.* 75, 3206–3215. doi: 10.1128/AEM.01791-08

Sessitsch, A., Hackl, E., Wenzl, P., Kilian, A., Kostic, T., Stralis-Pavese, N., et al. (2006). Diagnostic microbial microarrays in soil ecology. *New Phytol.* 171, 719–736. doi: 10.1111/j.1469-8137.2006.01824.x

Silver, W. L., Perez, T., Mayer, A., and Jones, A. R. (2021). The role of soil in the contribution of food and feed. *Philos. Trans. R. Soc. B Biol. Sci.* 376:20200181. doi: 10.1098/rstb.2020.0181

Singer, E., Bushnell, B., Coleman-Derr, D., Bowman, B., Bowers, R. M., Levy, A., et al. (2016). High-resolution phylogenetic microbial community profiling. *ISME J.* 10, 1–13. doi: 10.1038/ismej.2015.249

Speda, J., Jonsson, B.-H., Carlsson, U., and Karlsson, M. (2017). Metaproteomics-guided selection of targeted enzymes for bioprospecting of mixed microbial communities. *Biotechnol. Biofuels* 10:128. doi: 10.1186/s13068-017-0815-z

Srivastava, N., Gupta, B., Gupta, S., Danquah, M. K., and Sarethy, I. P. (2019). "Analyzing functional microbial diversity," in *Microbial diversity in the genomic era*. eds. S. Das and H.R. Dash (United States: Academic Press).

Stefanini, I., and Cavalieri, D. (2018). Metagenomic approaches to investigate the contribution of the vineyard environment to the quality of wine fermentation: potentials and difficulties. *Front. Microbiol.* 9:991. doi: 10.3389/fmicb.2018.00991

Swenson, T. L., Jenkins, S., Bowen, B. P., and Northen, T. R. (2015). Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biol. Biochem.* 80, 189–198. doi: 10.1016/j.soilbio.2014.10.007

Tedersoo, L., Albertsen, M., Anslan, S., and Callahan, B. (2021). Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Appl. Environ. Microbiol.* 87, 1–19. doi: 10.1128/AEM.00626-21

Tedersoo, L., Anslan, S., Bahram, M., Drenkhan, R., Pritsch, K., Buegger, F., et al. (2020). Regional-scale in-depth analysis of soil fungal diversity reveals strong pH and plant species effects in northern Europe. *Front. Microbiol.* 11:1953. doi: 10.3389/fmicb.2020.01953

Tedersoo, L., Drenkhan, R., Anslan, S., Morales-Rodriguez, C., and Cleary, M. (2019). High-throughput identification and diagnostics of pathogens and pests: overview and practical recommendations. *Mol. Ecol. Resour.* 19, 47–76. doi: 10.1111/1755-0998.12959

Tedersoo, L., Tooming-Klunderud, A., and Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytol.* 217, 1370–1385. doi: 10.1111/nph.14776

Terrat, S., Christen, R., Dequiedt, S., Lelievre, M., Nowak, V., Regnier, T., et al. (2012). Molecular biomass and MetaTaxogenomic assessment of soil microbial communities as influenced by soil DNA extraction procedure. *Microb. Biotechnol.* 5, 135–141. doi: 10.1111/j.1751-7915.2011.00307.x

Terrat, S., Horrigue, W., Dequiedt, S., Saby, A., Lelièvre, M., Nowak, V., et al. (2017). Correction: mapping and predictive variations of soil bacterial richness across France. *PLoS One* 12:e0186766. doi: 10.1371/journal.pone.0186766

Terrat, S., Plassart, P., Bourgeois, E., Ferreira, S., Dequiedt, S., Adele-Dit-De-Renseville, N., et al. (2015). Meta-barcoded evaluation of the ISO standard 11063 DNA extraction procedure to characterize soil bacterial and fungal community diversity and composition. *Microb. Biotechnol.* 8, 131–142. doi: 10.1111/1751-7915.12162

Thiele-Bruhn, S., Schloter, M., Wilke, B. M., Beaudette, L. A., Martin-Laurent, F., Cheviron, N., et al. (2020). Identification of new microbial functional standards for soil quality assessment. *Soil* 6, 17–34. doi: 10.5194/soil-6-17-2020

Thissen, J. B., Be, N. A., McLoughlin, K., Gardner, S., Rack, P. G., Shapero, M. H., et al. (2019). Axiom microbiome Array, the next generation microarray for high-throughput pathogen and microbiome analysis. *PLoS One* 14, 1–17. doi: 10.1371/journal.pone.0212045

Toju, H., Kishida, O., Katayama, N., and Takagi, K. (2016). Networks depicting the fine-scale co-occurrences of fungi in soil horizons. *PLoS One* 11:e0165987. doi: 10.1371/journal.pone.0165987

Tresch, S., Moretti, M., Le Bayon, R.-C., Mäder, P., Zanetta, A., Frey, D., et al. (2018). Urban soil quality assessment—A comprehensive Case study dataset of urban garden soils. *Front. Environ. Sci.* 6, 1–5. doi: 10.3389/fenvs.2018.00136

Trivedi, P., Mattupalli, C., Eversole, K., and Leach, J. E. (2021). Enabling sustainable agriculture through understanding and enhancement of microbiomes. *New Phytol.* 230, 2129–2147. doi: 10.1111/nph.17319

van Dorst, J., Bissett, A., Palmer, A. S., Brown, M., Snape, I., Stark, J. S., et al. (2014). Community fingerprinting in a sequencing world. *FEMS Microbiol. Ecol.* 89, 316–330. doi: 10.1111/1574-6941.12308

Van Emon, J. M. (2016). The omics revolution in agricultural research. *J. Agric. Food Chem. Food Chem.* 64, 36–44. doi: 10.1021/acs.jafc.5b04515

Van Goethem, M. W., Osborn, A. R., Bowen, B. P., Andeer, P. F., Swenson, T. L., Clum, A., et al. (2021). Long-read metagenomics of soil communities reveals phylum-specific secondary metabolite dynamics. *Commun. Biol.* 4:1302. doi: 10.1038/s42003-021-02809-4

Vargas-Hernández, J. G. (2020). *Transition of Ecosystem Services Based on Urban Agro Ecology*. Hershey, Pennsylvania: IGO Global.

Vestergaard, G., Schulz, S., Schöler, A., and Schloter, M. (2017). Making big data smart—how to use metagenomics to understand soil quality. *Biol. Fertil. Soils* 53, 479–484. doi: 10.1007/s00374-017-1191-3

Wallenius, K., Rita, H., Simpanen, S., Mikkonen, A., and Niemi, R. M. (2010). Sample storage for soil enzyme activity and bacterial community profiles. *J. Microbiol. Methods* 81, 48–55. doi: 10.1016/j.mimet.2010.01.021

Wang, S., Guo, Z., Wang, L., Zhang, Y., Jiang, F., Wang, X., et al. (2021). Wheat rhizosphere metagenome reveals newfound potential soil Zn-mobilizing Bacteria contributing to cultivars' variation in grain Zn concentration. *Front. Microbiol.* 12:689855. doi: 10.3389/fmicb.2021.689855

Wang, Y., Hayatsu, M., and Fujii, T. (2012). Extraction of bacterial RNA from soil: challenges and solutions. *Microbes Environ.* 27, 111–121. doi: 10.1264/jsme2.me11304

Watterson, W. J., Tanyeri, M., Watson, A. R., Cham, C. M., Shan, Y., Chang, E. B., et al. (2020). Droplet-based high-throughput cultivation for accurate screening of antibiotic resistant gut microbes. *elife* 9, 1–22. doi: 10.7554/eLife.56998

Wind, L., Keenum, I., Gupta, S., Ray, P., Knowlton, K., Ponder, M., et al. (2021). Integrated metagenomic assessment of multiple pre-harvest control points on lettuce Resistomes at field-scale. *Front. Microbiol.* 12:683410. doi: 10.3389/fmicb.2021.683410

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5088–5090.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576

Yang, X., Noyes, N. R., Doster, E., Martin, J. N., Linke, L. M., Magnuson, R. J., et al. (2016). Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain. *Appl. Environ. Microbiol.* 82, 2433–2443. doi: 10.1128/AEM.00078-16

Yang, T., Siddique, K. H. M., and Liu, K. (2020). Cropping systems in agriculture and their impact on soil health-A review. *Glob. Ecol. Conserv.* 23:e01118. doi: 10.1016/j.gecco.2020.e01118

Yang, G., Wagg, C., Veresoglou, S. D., Hempel, S., and Rillig, M. C. (2018). How soil biota drive ecosystem stability. *Trends Plant Sci.* 23, 1057–1067. doi: 10.1016/j.tplants.2018.09.007

Yap, M., Ercolini, D., Álvarez-Ordóñez, A., O'Toole, P. W., O'Sullivan, O., and Cotter, P. D. (2022). 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. doi: 10.1146/annurev-food-052720-010751

Yun, D. Y., Kang, Y. G., Kim, E. H., Kim, M., Park, N. H., Choi, H. T., et al. (2018). Metabolomics approach for understanding geographical dependence of soybean leaf metabolome. *Food Res. Int.* 106, 842–852. doi: 10.1016/j.foodres.2018.01.061

Zarraonaindia, I., Owens, S. M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., et al. (2015). The soil microbiome influences grapevine-associated microbiota. *MBio* 6:e02527-14. doi: 10.1128/mBio.02527-14

Zhang, M., He, L.-Y., Liu, Y.-S., Zhao, J.-L., Zhang, J.-N., Chen, J., et al. (2020). Variation of antibiotic resistome during commercial livestock manure composting. *Environ. Int.* 136:105458. doi: 10.1016/j.envint.2020.105458

Zhang, Y. J., Hu, H. W., Chen, Q. L., Singh, B. K., Yan, H., Chen, D., et al. (2019). Transfer of antibiotic resistance from manure-amended soils to vegetable microbiomes. *Environ. Int.* 130:104912. doi: 10.1016/j.envint.2019.104912

Zhou, A., He, Z., Qin, Y., Lu, Z., Deng, Y., Tu, Q., et al. (2013). Stress Chip as a high-throughput tool for assessing microbial community responses to environmental stresses. *Environ. Sci. Technol.* 47, 9841–9849. doi: 10.1021/es4018656

Conflict of Interest: BK was employed by the company Novasol Experts.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 02 August 2022
DOI 10.3389/fmicb.2022.951182

OPEN ACCESS

EDITED BY
Jerome Combrisson,
Mars (United States), United States

REVIEWED BY
Minakshi Prasad,
Lala Lajpat Rai University of Veterinary and Animal Sciences, India
Christian U. Riedel,
University of Ulm, Germany
Silvina Graciela Fadda,
CONICET Centro de Referencia para Lactobacilos (CERELA), Argentina

*CORRESPONDENCE
Frédéric Borges
frederic.borges@univ-lorraine.fr

SPECIALTY SECTION
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 23 May 2022
ACCEPTED 14 July 2022
PUBLISHED 02 August 2022

CITATION
Borges F, Briandet R, Callon C, Champomier-Vergès M-C, Christieans S, Chuzeville S, Denis C, Desmasures N, Desmonts M-H, Feurer C, Leroi F, Leroy S, Mounier J, Passerini D, Pilet M-F, Schlusselhuber M, Stahl V, Strub C, Talon R and Zagorec M (2022) Contribution of omics to biopreservation: Toward food microbiome engineering. *Front. Microbiol.* 13:951182. doi: 10.3389/fmicb.2022.951182

Contribution of omics to biopreservation: Toward food microbiome engineering

Frédéric Borges[1]*, Romain Briandet[2], Cécile Callon[3], Marie-Christine Champomier-Vergès[2], Souad Christieans[4], Sarah Chuzeville[5], Catherine Denis[6], Nathalie Desmasures[7], Marie-Hélène Desmonts[8], Carole Feurer[9], Françoise Leroi[10], Sabine Leroy[11], Jérôme Mounier[12], Delphine Passerini[10], Marie-France Pilet[13], Margot Schlusselhuber[7], Valérie Stahl[8], Caroline Strub[14], Régine Talon[11] and Monique Zagorec[13]

[1]Université de Lorraine, LIBio, Nancy, France, [2]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, Jouy-en-Josas, France, [3]Université Clermont Auvergne, INRAE, VetAgro Sup, UMR 545 Fromage, Aurillac, France, [4]ADIV, Clermont-Ferrand, France, [5]ACTALIA, Pôle d'Expertise Analytique, Unité Microbiologie Laitière, La Roche sur Foron, France, [6]ACTALIA, Sécurité des Aliments, Saint Lô, France, [7]Normandie Univ, UNICAEN, UNIROUEN, ABTE, Caen, France, [8]Aerial, Illkirch, France, [9]IFIP, Institut de la Filière Porcine, Le Rheu, France, [10]Ifremer, MASAE, Laboratoire EM3B, Nantes, France, [11]Université Clermont Auvergne, INRAE, MEDIS, Clermont-Ferrand, France, [12]Univ Brest, Laboratoire Universitaire de Biodiversité et Ecologie Microbienne, Plouzané, France, [13]Oniris, INRAE, SECALIM, Nantes, France, [14]Qualisud, Univ Montpellier, Avignon Université, CIRAD, Institut Agro, IRD, Université de La Réunion, Montpellier, France

Biopreservation is a sustainable approach to improve food safety and maintain or extend food shelf life by using beneficial microorganisms or their metabolites. Over the past 20 years, omics techniques have revolutionised food microbiology including biopreservation. A range of methods including genomics, transcriptomics, proteomics, metabolomics and meta-omics derivatives have highlighted the potential of biopreservation to improve the microbial safety of various foods. This review shows how these approaches have contributed to the selection of biopreservation agents, to a better understanding of the mechanisms of action and of their efficiency and impact within the food ecosystem. It also presents the potential of combining omics with complementary approaches to take into account better the complexity of food microbiomes at multiple scales, from the cell to the community levels, and their spatial, physicochemical and microbiological heterogeneity. The latest advances in biopreservation through omics have emphasised the importance of considering food as a complex and dynamic microbiome that requires integrated engineering strategies to increase the rate of innovation production in order to meet the safety, environmental and economic challenges of the agri-food sector.

KEYWORDS

shelf life, food, safety, pathogen, spoilage, microbiome, fermentation, biopreservation

Introduction

Foods of animal or plant origins are complex ecosystems, rich in nutrients, with physicochemical characteristics enabling microbial growth during processing and storage. These ecosystems are colonised by microbial communities that can include pathogenic or spoilage microorganisms but also beneficial ones. The consumption of food contaminated with pathogens is an important cause of morbidity and mortality worldwide. Every year, approximately 600 million people – 1 in 10 people – get sick from foodborne pathogens, 420,000 of whom die. Human damage caused by foodborne pathogens results in colossal economic losses amounting to USD 110 billion due to lost productivity and health expenses (World Health Organization, 2015). Spoilage organisms are responsible for colour, odour, texture, taste or packaging defects leading to inedible products. Food microbial contaminants (spoilage as well as pathogenic microorganisms) contribute to global food loss and waste. According to the Food and Agriculture Organization (FAO), food waste occurs during the retail and consumption stages while food loss occurs after harvest or slaughter until retail (FAO, 2019). The causes of food waste and loss are numerous but a significant part of food destruction linked to microbial contamination is due to non-compliance with pathogen-related regulations or spoilage. At the European Union level, 20% of total available food is lost or wasted, fruits and vegetables being the most impacted category (43.5% of the food group), ahead of meat and fish products (26.3%; Caldeira et al., 2019). Worldwide, one-third of food produced for human consumption, about 1.3 billion tonnes per year, is estimated to be lost or wasted along the food supply chain (HPLE, 2014), while about 12% of the world population suffers from hunger (Agriculture and Economic Development Analysis Division, 2013). Reducing food wastage is thus crucial not only for ethical reasons but also for economic reasons. Food loss and waste are responsible for direct costs of about USD one trillion every year, but hidden costs extend much further. Indeed, global costs (including environmental, social, and economic costs) are evaluated by the FAO to amount to USD 2.6 trillion per year (FAO, 2014).

In addition, as food waste is correlated with high greenhouse gas emissions, reducing the undesirable impact of microorganisms has become a major objective in a situation of climate emergency. In this context, microorganisms also offer potential levers of action and can represent real opportunities for resolving food safety issues (Cavicchioli et al., 2019). In the global food system, approaches based on the barrier properties of biological systems appear attractive because of their efficiency and sustainability. These approaches are grouped under various names including biosanitation, biocontrol, bioprotection, and biopreservation. In this review, we will refer to food biopreservation, which is based on the hurdle technology that consists in using microorganisms (often lactic acid bacteria) as protective cultures and/or their metabolites to optimise the microbiological quality and shelf life of food by ensuring safety or reducing food waste, as defined by Stiles (Stiles, 1996). The use of protective cultures is often considered as an alternative to chemical additives or as a replacement for certain ingredients. Therefore, biopreservation should also help to meet the strong expectations of consumers who want "healthier" and more "natural" foods, and contribute to nutritional recommendations aimed at reducing salts, sugars, and additives in foods.

The concept of biopreservation was inspired by food fermentation ancestrally used to preserve food, except that fermentation involves substantial transformation of the food matrix, which is usually not the intention when engineering biopreservation systems. Consistently, the intentional addition of microorganisms or their metabolites for specific preservation purposes was largely investigated on fermented food, i.e., dairy products (cheeses, yoghurt), bakery products, or fermented sausages. Nevertheless, biopreservation was successfully extended to non-fermented food such as seafood, raw meat and non-fermented plant products. For example, biopreservation of seafood as fresh fish fillets, smoked fish or cooked shrimps aimed at controlling spoilers or pathogens such as *Listeria monocytogenes, Staphylococcus aureus, Vibrio* and histamine-producing bacteria (for a recent review, see Rathod et al., 2022). Biopreservation of raw meat (lamb, pork, beef, or poultry) and processed meat (sausage, cooked ham) has also focused on the control of pathogenic and spoilage bacteria as well as extension of shelf life (Jones et al., 2010; Zagorec and Champomier-Vergès, 2017; Wang et al., 2019). Biopreservation of non-fermented plant products is essentially dedicated to fighting against spoilage microorganisms including yeasts, moulds, spore-forming bacteria (*Bacillus subtilis, Bacillus licheniformis*), and pathogenic bacteria such as *L. monocytogenes* (Leyva Salas et al., 2017).

Most species used for food biopreservation by the food industry are lactic acid bacteria belonging essentially to the genera *Lactococcus*, *Lactobacillus lato sensu* and *Carnobacterium*. One strain of the Gram-negative species *Hafnia alvei* is also available on the market for anti-*Escherichia coli* purposes (Callon et al., 2016; Frétin et al., 2020). These bacteria are derived from food microbiota and are therefore particularly well adapted to food matrices. Moreover, as these species have been studied for several decades, their use as protective cultures is considered safe. *Latilactobacillus sakei, Latilactobacillus curvatus, Carnobacterium divergens, Carnobacterium maltaromaticum* and *Lactococcus lactis* are the main species providing protective cultures for meat/seafood products (Leroi et al., 2015; Comi et al., 2016; Ramaroson et al., 2018; Iacumin et al., 2020), while lactobacilli or even yeast strains can be used for vegetable food biopreservation (Siedler et al., 2019; Truchado et al., 2020; Windholtz et al., 2021). Although several biopreservation technologies are already available on the market, the rate and speed of innovation

in this area needs to increase considerably in order to meet global climate-related challenges. Indeed, the scientific literature dealing with biopreservation is considerably larger than the actual application of biopreservation in the food sector. Foods are complex systems because of their diversity, their various physical, chemical, and biological structures, and the numerous processes used to produce them. Moreover, food microbial community dynamics during shelf life depends on abiotic parameters, which are mostly linked to production processes or storage conditions, and biotic parameters where microbial interactions play a major role. The complexity and diversity of food microbial communities has been a major barrier to the widespread use of biopreservation. Therefore, the conception of efficient biopreservation technology requires the implementation of methodological approaches adapted to the high complexity of food ecosystems. Food microbiology has long been studied by classical culture-dependent methods. During the last decade, omics techniques including genomics, transcriptomics, proteomics, metabolomics, culturomics, and phenomics have revolutionised all areas of the life sciences, including food quality and safety assessment, because of their ability to decipher food systems as a whole (Cook and Nightingale, 2018). The application of omics provides a more realistic portrait of the complex interactions occurring in the food ecosystem (Gálvez et al., 2007), and it has significantly increased our understanding of the potential of microbiomes to increase the productivity and sustainability of food systems. These omics approaches have caused a paradigm shift from unsocial undesirable microorganisms colonising food to strongly interacting microorganisms establishing stable networks (Berg et al., 2020). Omics approaches have been applied to explore various aspects of biopreservation, such as selection and characterisation of protective cultures, investigation of the mechanisms involved in the protective effect, or the impact on food microbial communities. After a brief overview of food biopreservation, the different omics approaches used in studies dealing with biopreservation are reviewed below and analysed by illustrating to what extent they can answer questions related to the impact of biopreservation on the food microbial ecosystem.

Selection of biopreservation agents

Classical approaches to identify protective culture candidates are mainly based on the detection of inhibition zones in laboratory conditions. Recently, the use of phenomics and genomics has proved highly effective. Phenomics can be defined as the high-throughput study of phenotypes. In the case of biopreservation, the phenotype of interest is the inhibition of spoilage microorganisms or food-borne pathogens. Phenomics has been successfully used to identify strains exhibiting remarkable anti-*L. monocytogenes* properties by using a high-throughput liquid handling system and a genetically engineered luminescent strain of *L. monocytogenes* to set up mixed culture competition assays in food matrices (Riedel et al., 2007; El Kheir et al., 2018). This method was first used to select anti-*Listeria* candidates from a collection of strains isolated from raw milk and a collection of *Lactococcus piscium*. The majority of the candidates obtained did not produce an inhibitory halo following a classical agar diffusion-based method, suggesting promising inhibitory mechanisms (El Kheir et al., 2018). Lately, this high-throughput competition assay was implemented to study the inhibition phenotype of a collection of *C. maltaromaticum* strains under multiple varying conditions. This method resulted in the selection of robust antagonistic *C. maltaromaticum* strains whose anti-*Listeria* properties are insensitive to fluctuations, i.e., inoculation level and time lag of *L. monocytogenes* and candidate inoculation (Borges and Revol-Junelles, 2019). It is expected that phenomics approaches extended to other target microorganisms (Besnard et al., 2021), as well as other phenotypes related to the sensory profile or use of nutrients (Wiernasz et al., 2017; Acin-Albiac et al., 2020), may in the future enable the identification of microorganisms with high biopreservation performances.

Genomics is an interesting approach for selecting protective cultures as genome mining may point out important features that can be involved in the preservation effect (Baltz, 2019), and also prove the absence of some unwanted functions such as antibiotic resistance or biogenic amine synthesis. Genome mining involves the analysis of functional gene annotation, resulting in particular from antiSMASH (Blin et al., 2017) and BAGEL (van Heel et al., 2018), which are designed to identify clusters of genes involved in the biosynthesis of antimicrobial compounds, combined with genome comparison to find correlation between the presence/absence of genes and protective properties. As an example, the genome sequence of *L. sakei* 23K, a meat adapted bacterium used as a starter for sausage fermentation, but also proposed as a biopreservative agent for raw meat products (Zagorec and Champomier-Vergès, 2017), revealed its strong ability to be competitive in meat products (Chaillou et al., 2005; Eijsink and Axelsson, 2005). Indeed, genome analysis highlighted the presence in the genome of elements putatively enabling the use of alternative carbon sources, such as ribose, inosine, and adenosine. Their efficacy was subsequently proven and helps to explain the fitness in meat of *L. sakei*, which thereby escapes competition for energy sources (Rimaux et al., 2011). Also, the requirement for haem and iron, two components present in meat, could be assessed by genomics through the gene repertoire and further evidenced by functional genomics (Duhutrel et al., 2010; Verplaetse et al., 2020).

Another example of genomics input is the production of antagonistic molecules by protective strains. This has long been studied through the production of bacteriocins. The

genome sequence analysis of *L. curvatus* CRL705, a strain known to produce two bacteriocins (lactocin 705 and AL705), revealed the presence of additional genes putatively involved in bacteriocin production (sakacin P, sakacin Q, sakacin X, and sakacin T; Hebert et al., 2012). Divercin V41 is a bacteriocin involved in the protective function of *C. divergens* V41 a lactic acid bacterium strain whose operon sequence was reported more than two decades ago (Metivier et al., 1998). The genome analysis of this strain revealed that an additional gene was present in the divercin V41 operon (Remenant et al., 2016), the function of which was shown to be important for bacteriocin production (Back et al., 2015). Comparative genomics of *Carnobacterium* highlighted potential new candidate strains for biopreservation, efficient against *L. monocytogenes* and harbouring original bacteriocin gene equipment. For example, five different bacteriocins and a 16 kDa new one were predicted in the *C. maltaromaticum* SF668 and *C. maltaromaticum* EBP3019 genomes, respectively (Begrem et al., 2020). Combining genome analysis and peptidomics of a *Companilactobacillus crustorum* strain enabled, from the peptides produced during the growth of this strain, the discovery of eight novel bacteriocins and two other antimicrobials with a broad spectrum of action against Gram-positive and Gram-negative pathogens (Yi et al., 2018).

L. piscium CNCM I-4031 (EU2241) is a protective strain for seafood products that improves the sensory quality of shrimp and cold-smoked salmon (Fall et al., 2012; Leroi et al., 2015). This strain is also particularly efficient against the pathogen *L. monocytogenes* by reducing growth and virulence (Saraoui et al., 2018). Combined phenotyping and genome analyses evidenced that the inhibitory effect is dependent on cell-to-cell contact instead of extracellular molecules such as bacteriocins, organic acids, or hydrogen peroxide (Saraoui et al., 2016; Marché et al., 2017). This unusual mechanism still remains to be elucidated; nevertheless medium- and high-throughput screening revealed that other strains of the same species could be selected as new protective cultures, notably related to their large antimicrobial capacities (Wiernasz et al., 2017; El Kheir et al., 2018).

Genomics can reveal other unexpected features, as exemplified by the genome sequence of *C. divergens* V41 which contains an intriguing long genomic island (~40 kb) encoding polyketide synthases/non-ribosomal peptide synthases (PKS/NRPS) or PKS/NRPS-like enzymes, putatively involved in the production of a secondary metabolite of unknown function (Remenant et al., 2016). Such molecules may have antimicrobial functions or be associated with oxidative stress resistance, and immunomodulatory or cytotoxicity activities. Comparative genomic analysis of *Carnobacterium* strains showed that this PKS/NRPS gene cluster is unique in *C. divergens* V41 (Begrem et al., 2020). Other PKs/NRPs antimicrobial compounds, such as milkisin produced by a *Pseudomonas* sp. strain, with potentially interesting properties for the biopreservation of milk products have been described (Schlusselhuber et al., 2018, 2020). Reuterin-producing *Limosilactobacillus reuteri* are known as bioprotective agents for dairy products (Ortiz-Rivera et al., 2017). Some rare strains harbour a PKS/NRPS genomic island encoding *rtc* genes involved in the synthesis, regulation, immunity, and secretion of an antimicrobial tetramic acid named reutericyclin (Gänzle et al., 2000; Lin et al., 2015).

While many published papers describe the screening and evaluation of bioprotective antifungal lactic acid bacteria strains *in vitro* and *in situ* (Delavenne et al., 2012; Leyva Salas et al., 2018), as well as the identification of metabolites involved in their antifungal activities, the application of genomics and functional genomics to the selection of antifungal microorganisms is still in its infancy. Studies providing insights into the metabolic pathways of antifungal metabolites such as phenyllactic acid (Wu et al., 2020), clearly point out that in-depth studies of antifungal lactic acid bacteria using comparative and functional genomics are needed. Such studies should help to elucidate yet uncharacterised biosynthetic pathways of known antifungal molecules and potentially reveal new ones and to establish whether other competition exclusion phenomena exist between lactic acid bacteria and spoilage fungi, with the objective to further understand their action mechanism and possibly provide helpful tools for strain selection.

Some species with bioprotective properties are also described as potential spoilage organisms, depending on the type of food or process (Brillet et al., 2004; Leisner et al., 2007; Andreevskaya et al., 2015; Leroi et al., 2015; Saraoui et al., 2016; Poirier et al., 2018). Thus comparative genomics between strains known to be responsible for spoilage or on the contrary known as protective cultures should be an interesting approach for the selection of strains of interest. In addition, as a complementary approach to phenotypic tests, genome analysis of potent protective strains can be a complementary tool for their safety assessment i.e., screening for the presence of genes related to biogenic amine production, antibiotic resistance genes, as well as their location with respect to mobile genetic elements, i.e., plasmids and bacteriophages. Such an approach was recently applied to a *Lactiplantibacillus plantarum* starter culture used in the manufacturing of *nahm* fermented pork (Chokesajjawatee et al., 2020), as well as to a potent *L. plantarum* protective culture (Barbosa et al., 2021).

Impact of biopreservation on targeted microorganisms

The role of protective cultures is mainly related to their fitness (nutritional competition, ability to resist harsh conditions encountered in food), enabling them to establish and dominate at the expense of other undesirable species, as well as to their ability to produce antimicrobial molecules,

TABLE 1 Main characteristics of biopreservation in food and mechanisms involved.

Protective mode of action	Resulting effect	Effect at cellular and/or molecular levels	Mode of use	References
Nutritional competition	- Jameson effect - Growth impairment because of lack of nutrients	- Early entry into stationary phase of targeted microorganisms, and protective cultures - Growth cessation of targeted microorganisms	Live microorganisms added to the food	Jameson, 1962; Guillier et al., 2008; Hibbing et al., 2010
Production of organic acids	- Extracellular pH drop - Diffusion across the microbial cytoplasmic membrane	- Cytoplasmic pH decrease - Collapse of the proton gradient across the membrane - Disruption of cellular processes		Warnecke and Gill, 2005
Production of hydrogen peroxide	- Oxidation of cellular components	- Peroxidation and disruption of membrane layers - Oxidation of oxygen scavengers - Enzyme inhibition - Oxidation of nucleosides - Disruption of protein synthesis - Growth decrease at low concentrations - Cell death at high concentrations		Finnegan et al., 2010
Respiration	Change in atmosphere composition (O_2 decrease) leading to microaerophilic conditions	- Inhibition of strict aerobic bacteria	Live microorganisms added to the food	Ben Said et al., 2019
Production of bacteriocins	Bactericidal or bacteriostatic activity against species taxonomically related to the producing strain	- Pore formation in the cytoplasmic membrane - Loss of structure and subsequent cell death	Live microorganisms added to the food	Elsser-Gravesen and Elsser-Gravesen, 2013
			Bacteriocins purified from cultivated producer strains	Martinez et al., 2016

such as organic acids, hydrogen peroxide or specialised metabolites as bacteriocins, biosurfactants, or lipopeptides (Ben Said et al., 2019; Bourdichon et al., 2021; **Table 1**). The expected impact is induction of decay or at least growth inhibition of the undesirable microorganism. Classical methods such as qRT-PCR are *a priori* methods that involve a prior selection of the target genes to be studied. On the contrary, omics approaches such as RNAseq can be used without *a priori*, opening the possibility to identify original unsuspected mechanisms. These approaches have revealed that the mechanisms involved are more complex than previously thought, and involve microbial sensing capacity, gene regulation, and the potential for combining antimicrobial compounds for microbial inactivation (**Figure 1**).

Listeria monocytogenes

Two recent studies have explored the mechanism of inhibition of *L. monocytogenes* by antimicrobial lipopeptide (P34) or the peptide nisin (encapsulated or not) by proteomic analysis after incubation in a laboratory medium (Pinilla et al., 2021; Stincone et al., 2021). The lipopeptide P34 caused the downregulation of proteins involved in manganese transport and upregulation of proteins related to iron transport in *L. monocytogenes*. In addition, reduction of stress tolerance proteins related to the sigma B and VirR regulons and modulation of phosphoenolpyruvate phosphotransferase systems for sugar transport were observed (Stincone et al., 2021). Exposure of *L. monocytogenes* to nisin induced the synthesis of proteins related to ATP-binding cassette transporter systems, transmembrane proteins, RNA-binding proteins, and diverse stress response proteins (Pinilla et al., 2021). Some of the proteins detected in the presence of free nisin were related to translocation of *L. monocytogenes* virulence factors, activation of the LiaR-mediated stress defence, and glycosylation of cell wall teichoic acid. The comparison of treatment by free and encapsulated nisin revealed that *L. monocytogenes* did not express some stress proteins when nisin was encapsulated, suggesting the production of nisin-resistance factors by exposure to encapsulated nisin. The authors suggested that liposomes allow controlled release of nisin in the medium, resulting in fewer interactions between nisin and bacteria compared with free nisin, which may impact the mechanism of action of nisin (Pinilla et al., 2021). The induction of resistance factors was also observed by exposing *L. monocytogenes* to sublethal doses of pediocin. Transcriptomic analysis revealed the expression modulation of the two-component system LisRK and the alternative sigma factors SigB and SigL, resulting in an increased resistance of *L. monocytogenes* to this bacteriocin (Laursen et al., 2015).

Recently, the ability of *C. maltaromaticum*, *Leuconostoc gelidum*, or *L. piscium* strains to inhibit *L. monocytogenes* was demonstrated in seafood (Saraoui et al., 2016, 2018;

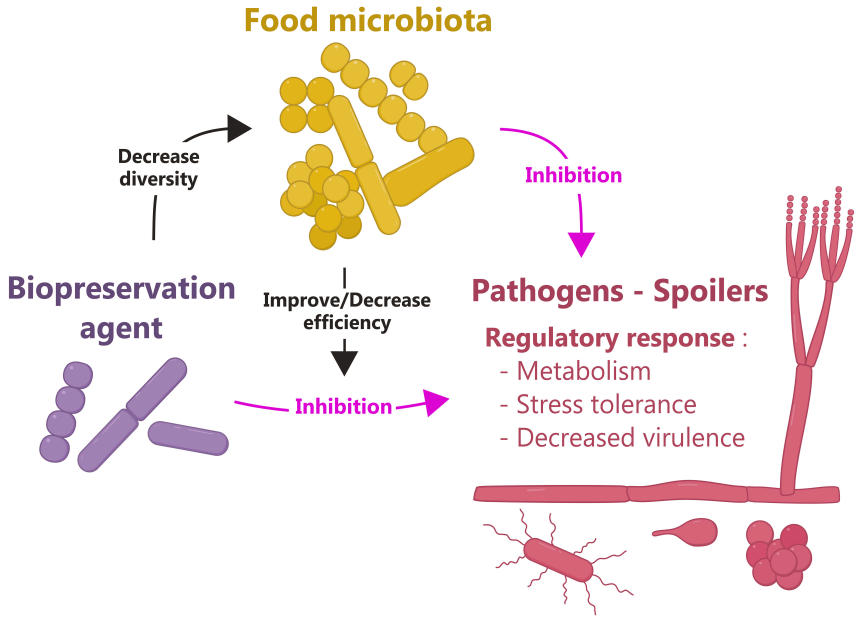


FIGURE 1
Major advances in the field of biopreservation thanks to omics approaches. The objective of biopreservation is to induce growth cessation or decay of pathogens and spoilers. Omics approaches have shown that in response to biopreservation, target microorganisms modulate the expression of genes involved in cell metabolism, stress tolerance and virulence. In addition, omics revealed that biopreservation can decrease the diversity of food microbiota. Conversely, depending on its structure, the food microbiota can either improve or reduce the effectiveness of biopreservation agents. Food microbiota can also have intrinsic protective properties and thus inhibit unwanted microorganisms.

Wiernasz et al., 2017). *L. piscium* inhibition required cell-to-cell contact with *L. monocytogenes*, affecting its cell surface, and decreasing its virulence (Saraoui et al., 2018). The metabolomic fingerprints suggested that this inhibition might not involve nutritional competition and remains to be explored (Saraoui et al., 2016).

Staphylococcus aureus

Lactococcus garvieae was shown to inhibit *S. aureus* growth in milk, in cheese, and *in vitro*, potentially through hydrogen peroxide (H_2O_2) production (Delbes-Paus et al., 2010). To better characterise this mechanism of inhibition *in vitro*, the transcriptomes of *L. garvieae* and *S. aureus* co-culture have been explored by RNA-seq and RT-qPCR (Delpech et al., 2015, 2017). *L. garvieae* repressed the expression of *S. aureus* genes involved in stress response, including oxidative stress generated by H_2O_2, and in cell division. It also modulated the expression of virulence-related genes (particularly *agrA*, *hld*, and enterotoxin-encoding genes; Delpech et al., 2015). For *L. garvieae*, a high concentration of H_2O_2 was not associated with higher expression of the H_2O_2 synthesis genes *pox*, *sodA*, and *spxA1*, but rather with repression of H_2O_2-degradation genes (*trxB1*, *ahpC*, *ahpF*, and *gpx*; Delpech et al., 2017). The interaction between *L. lactis* and *S. aureus* has also been widely studied and transcriptomic analyses were performed with microarrays and RT-qPCR. In a chemically defined medium held at a constant pH value of 6.6, the growth of the two bacteria in co-culture was not modified, but their transcriptome was modulated (Even et al., 2009; Nouaille et al., 2009). The expression of *S. aureus* virulence-related genes was impaired by *L. lactis*: the expression of genes encoding global regulators, including *agr* and consequently the *agr*-controlled enterotoxin genes and *sar*, was strongly reduced (Even et al., 2009). This downregulation of *agr* in *S. aureus* was associated with the reducing properties of *L. lactis* (Nouaille et al., 2014). *L. lactis* genes associated with amino acid metabolism, ion transport, oxygen response, menaquinone metabolism, cell surface, and phage expression were differentially expressed in co-culture compared to monoculture (Nouaille et al., 2009). In a complex medium such as the cheese matrix, the acidifying, proteolytic, and reducing activities of *L. lactis* were shown to affect carbohydrate and nitrogen metabolisms and the stress response of *S. aureus* (Cretenet et al., 2011). Enterotoxin gene expression was positively or negatively modulated by both *L. lactis* and the cheese matrix itself, depending on the enterotoxin type. Again, the *agr* operon was downregulated by the presence of *L. lactis*, in part because of a drop in pH (Cretenet et al., 2011). All these data highlight the intimate link between environment, metabolism, and virulence expression. A third binary interaction between *Enterococcus faecalis* and *S. aureus* was studied in milk and in cheese (Viçosa et al., 2018; Nogueira Viçosa et al., 2019). When

co-cultured, the growth of *S. aureus* was decreased and the classical enterotoxins were not produced. The expression of several enterotoxins and global regulator genes (including *agr*) was downregulated, while the expression of genes involved in metabolism was upregulated. Finally, the interaction of *S. aureus* with a mixed culture of *Enterococcus durans, E. faecalis* and *L. lactis* in milk confirmed that the production of enterotoxins was reduced in mixed culture and the expression of several genes involved in virulence was inhibited (Zdenkova et al., 2016). All these studies on the interaction of *S. aureus* with different lactic acid bacteria converge on a very important idea that the mechanism of action of biopreservation can result in the inhibition of virulence by inhibiting the production of enterotoxins through the decreased expression of genes involved in their synthesis.

Pathogenic *Escherichia coli*

L. curvatus, *L. plantarum*, and *Enterococcus mundtii* strains can inhibit the growth of *E. coli* O157:H7 when co-cultured in a meat model medium (Orihuel et al., 2018). The antagonistic effect of the most efficient *E. mundtii* strains against *E. coli* O157:H7 were characterised by a proteomic analysis (Orihuel et al., 2018). The expression of *E. mundtii* proteins involved in carbohydrate/amino acid metabolisms, energy production, transcription/translation, and cell division was modified in the presence of *E. coli*. Reciprocally, the presence of *E. mundtii* resulted in repression of *E. coli* synthesis of proteins related to metabolism and transport of amino acids and nucleotides, as well as overexpression of proteins involved in stress, energy production, and transcription (Orihuel et al., 2019). In addition, proteins associated with adhesion to extracellular matrix proteins of meat were modulated in *E. coli* in accordance with its decreased adhesion capacity when co-cultured with *E. mundtii*. *E. mundtii* did not influence the lytic cycle of the *E. coli* O157:H7 strain, indicating its potentially safe use as a bioprotective agent, since engagement in the lytic cycle results in the production of shiga toxin (Orihuel et al., 2019). The interaction of *E. coli* O157:H7 with ground beef microbiota was also studied (Galia et al., 2017). A beef piece was divided in two parts with the inner part considered sterile and the outer part as encompassing a natural meat microbiota. The microbial community structure was assessed by 16S rDNA amplicon sequencing, and the transcriptome of two inoculated strains (*E. coli* O157:H7 and *E. coli* O26:H11) was studied by RNAseq comparing samples of sterile and naturally contaminated meat. This study revealed that the two *E. coli* strains behave differently. On the one hand, an upregulation of genes involved in detoxification and stress response and a downregulation of *peR*, a gene negatively associated with virulence phenotype, were observed in *E. coli* O157:H7. On the other hand, the interaction of *E. coli* O26:H11 with ground beef microbiota revealed that genes involved in division, peptidoglycan synthesis, DNA repair, metal acquisition, and carbohydrate and amino acid metabolism were downregulated (Galia et al., 2017).

Fungal food spoilers

In a recent review of biopreservation against moulds in dairy products (Shi and Maktabdar, 2022), the authors point out newly described antifungal mechanisms. Among these is a perfect example of how omics technologies shed new light on the understanding of the protective mechanisms of antifungal lactobacilli toward dairy product spoilage fungi (Siedler et al., 2019). This work revealed that manganese scavenging by *Lacticaseibacillus rhamnosus* and *Lacticaseibacillus paracasei* antifungal strains, previously known as a defence mechanism against oxidative stress, was a main inhibitory mechanism (i.e., competitive exclusion) against many yeast and mould species involved in dairy product spoilage. Indeed, following milk fermentation and supplementation with manganese, their bioactivity was completely lost. A transcriptomic approach based on RNA-seq further showed that one of the most highly expressed gene products in these strains encoded a manganese transporter (MntH1). The role of MntH1 in manganese scavenging was confirmed in a Δ*mntH1 L. paracasei* strain in which no significant antifungal activity was detected, while bioactivity was restored in the Δ*mntH1* mutant complemented with a plasmid containing the *mntH1* gene under its own promoter.

Besides the above-mentioned discovery that competitive exclusion for manganese was an important antifungal mechanism, production of antifungal metabolites and pH decrease were believed to be the main mechanisms involved in the bioactivity of antifungal protective lactic acid bacteria. Through the use of metabolomic targeted and untargeted approaches with or without prior medium fractionation and bioactivity testing, more than 60 molecules have been thought to play a role in antifungal activity (see recent reviews by Leyva Salas et al., 2017 and Siedler et al., 2019). These metabolites include molecules produced through carbohydrate metabolism (e.g., organic acids such as lactic, acetic, formic, and succinic acids, volatile compounds such as diacetyl), proteolysis (e.g., bioactive peptides resulting from casein cleavage), amino acid metabolism (e.g., phenyllactic acid), lipolysis and free fatty acid metabolism (e.g., 3-hydroxydecanoic, caproic – i.e., hexanoic- and caprylic – i.e., octanoic – acids), but also complex compounds derived from bioconversions (e.g., benzoic acid) or peptide synthesis [e.g., cyclic dipeptides such as cyclo(L-Phe-L-Pro) and cyclo(L-Phe-trans-4-OH-L-Pro)] (Ström et al., 2002; Aunsbjerg et al., 2015; McNair et al., 2018; Leyva Salas et al., 2019; Garnier et al., 2020; Shi and Knøchel, 2021a,b). It should be underlined that with the exception of lactic and acetic acids which are produced in g/kg or g/L

amounts, these molecules are produced in quite low quantities, all of which are at concentrations far from their individual minimum inhibitory concentration (MIC), thus suggesting that they act in synergy or by additive effects. In a recent study, Leyva Salas et al. (2019) used a metabolomics approach coupled with supervised multivariate analysis to investigate 56 antifungal compounds as well as volatiles. It was found that 9 key compounds including acetic acid, 5 aromatic acids, and three volatiles were associated with antifungal activity against *Mucor racemosus* and *Penicillium commune,* although their concentrations were below their respective MICs. Further investigation on *Penicillium roqueforti* and *Mucor circinelloides* revealed that several combinations presented additive (e.g., diacetyl + 3-phenylpropanoic acid, diacetyl + acetic acid) or synergistic effects (diacetyl + octanoic acid, octanoic + 3-phenyl propanoic acids), clearly reinforcing the idea that additive and synergistic effects of antifungal molecules are involved in lactic acid bacteria bioactivity. To go further, future work could include the investigation of more complex mixtures of antifungal compounds at concentrations close to those encountered in biopreserved foods. Moreover, it is not clear how antifungal molecule synthesis and competitive exclusion interact together in different fungal species with various susceptibilities to protective strains.

Biopreservation at the microbiome level

The primary objective of biopreservation is to limit the presence of unwanted microorganisms in food. However, the addition of biopreservatives can have an overall impact on the food microbiome. Omics approaches have helped clarify the complex interactions occurring in the food ecosystem and their impact on the organoleptic properties of foods (**Figure 1**).

Metagenomics

Prior to the availability of high-throughput DNA sequencing (HTS) techniques, DNA fingerprinting techniques as PCR-denaturing gradient gel electrophoresis or PCR-temporal temperature gradient electrophoresis, targeting mainly the V3 region of 16S rDNA, were employed to determine whether bioprotective lactic acid bacteria strains were able to colonise the food matrix and dominate the microbiota (Hu et al., 2008; Saraoui et al., 2017; Zhang et al., 2018). However, discrepancies between these DNA fingerprinting approaches and cultural methods (Ercolini et al., 2010) have highlighted the need for methods of higher resolution. There are two commonly used HTS methods in microbiome research: amplicon sequencing and metagenomic sequencing (Liu et al., 2021). Amplicon sequencing is the most widely used, including in the field of biopreservation. It demonstrated that the three protective cultures *L. rhamnosus* LRH05, *L. sakei* LSK04, and *C. maltaromaticum* CNB06, alone or in combination, were able to colonise cheese and became dominant after storage (Bassi et al., 2020). In shrimp, the results of amplicon sequencing were consistent with the successful colonisation of the food matrix by the biopreservative strains *L. plantarum* and *Lacticaseibacillus casei*, and a reduction in the relative abundance of *Shewanella*, which includes the spoilage species *Shewanella baltica* (Li et al., 2019). In addition, in beef burgers, the analysis of predicted metagenomes revealed that nisin-activated packaging resulted in a reduction in the abundance of specific metabolic pathways related to spoilage (Ferrocino et al., 2015). **Table 2** summarises the different studies using omics approaches to assess the impact of biopreservation on food ecosystems and the main findings.

HTS approaches open up the possibility of going much further than simply answering the question of the implantation success of the protective microorganisms and their effect on the targets. Amplicon sequencing can be used to estimate the impact of biopreservation on food microbiota. When batches of cold-smoked salmon were inoculated with a bioprotective strain of *L. piscium*, the microbiota structure differed significantly between control and biopreserved products after 3 weeks of storage (Leroi et al., 2015). The impact of biopreservation strains can be stronger, as in the case of fermented sausages where 16S rRNA-based analysis revealed a markedly lower microbial diversity of the metabolically active microbial community (Giello et al., 2018). Thus, although special attention is paid to the selection of strains with a narrow antimicrobial spectrum in biopreservation (e.g., bacteriocin-producing bacteria), biopreservation can end up with significant changes in food microbiota structure. Although this side effect can be undesirable, especially in fermented foods, in some circumstances it is possible to take advantage of this broad impact on food microbiota to target spoilage microorganisms, which can encompass a large number of phylogenetic taxa. In dairy products such as low-salt fresh cheeses, spoilage bacteria are mainly psychrotrophic Gram-negative species, including several *Enterobacteriaceae* and *Pseudomonas* spp. (Ledenbach and Marshall, 2009; Spanu et al., 2018; Bassi et al., 2020). The three lactic acid bacterial strains *L. rhamnosus* LRH05, *L. sakei* LSK04, and *C. maltaromaticum* CNB06, added alone or in combination, were found after 5 weeks of storage to be effective in inhibiting the Gram-negative bacteria population of fresh Primo Sale cheese inoculated with a cocktail of 10 bacterial spoilage isolates (Bassi et al., 2020). In Italian fresh filled pasta cheese, the impact of protective cultures at the community level can be used to reduce the initial microbiota associated with raw materials and to confer a competitive advantage on safer or more acceptable bacterial species such as *Leuconostoc mesenteroides,* at the expense of more problematic species such as *Streptococcus uberis* and *Streptococcus parauberis* (Tabanelli et al., 2020).

TABLE 2 Overview of omics approaches used to assess the impact of biopreservation on food ecosystems and main findings.

Omics approach	Methodological details	Food	Main finding	References
PCR-Denaturing gradient gel electrophoresis/Temporal temperature gradient electrophoresis associated or not with band sequencing	V3 region of 16S rDNA	Cooked and peeled shrimp	*Carnobacterium divergens* V41 but not *Lactococcus piscium* CNCM I-4031 used to inoculate shrimp dominated and was associated with reduction of off-flavours	Saraoui et al., 2017
	V3 region of 16S rDNA	Vacuum-packaged beef meat	Bands associated with spoilage bacteria (*Pseudomonas*, Enterobacteriaceae, *Brochothrix. thermosphacta*), but not with LAB, disappeared in samples inoculated with *Latilactobacillus sakei* and *Latilactobacillus curvatus* bioprotective strains	Zhang et al., 2018
	V3 region of 16S rDNA	Vacuum-packed cooked ham	Predominant spoilage LAB were not detected when the bioprotective *Latilactobacillus sakei* B-2 strain was used	Hu et al., 2008
	V6–V8 region of 16S rDNA	Beef cuts packaged in nisin-coated plastic bags	Similar diversity in control and nisin-treated samples although differences were observed with plate counts for *Brochothrix thermosphacta*	Ercolini et al., 2010
DNA sequencing	Pyrosequencing of V3–V4 region of 16S rDNA	Cold-smoked salmon	Different OTU ratios were observed between control and samples inoculated with *Lactococcus piscium* EU2241 (= CNCM I-4031). No correlation with sensory analysis	Leroi et al., 2015
	Illumina sequencing of V3–V4 region of 16S rDNA	Raw/peeled shrimp	*Shewanella baltica* significantly inhibited after co-inoculation with *Lactiplantibacillus plantarum* AB-1 and *Lacticaseibacillus casei* LC	Li et al., 2019
	Illumina sequencing of V3–V4 region of 16S rDNA	St Nectaire-type cheese	Implantation of an inhibitory consortium whose inhibitory activity toward *Escherichia coli* O26:H11 depended on indigenous microbiota composition	Frétin et al., 2020
	Illumina sequencing of V3–V4 region of 16S rDNA and of an internal 280 bp fragment of the *gyr*B gene	Diced cooked ham	Bioprotective activity and implantation of a nisin-producing strain of *Lactococcus lactis* depended on microbiota composition	Chaillou et al., 2022
	Illumina sequencing of V4 region of 16S rDNA	Fresh filled pasta	Cultures of *Lactiplantibacillus plantarum* and *Lacticaseibacillus paracasei* were not dominant but reduced the initial microbiota and gave a competitive advantage to other LAB species	Tabanelli et al., 2020
16S rRNA sequencing	Sequencing of V3–V4 region of 16S rRNA from cDNA	Fermented sausage	Large domination of *Lactobacillaceae* and reduction of bacterial diversity in samples inoculated with protective *Latilactobacillus curvatus* strain	Giello et al., 2018
	Sequencing of V3–V4 region of 16S rRNA from cDNA	Beef burgers in nisin-activated packaging	Lower abundance of some taxa in samples with nisin-activated packaging	Ferrocino et al., 2015
Volatilome analysis	Headspace SPME/GC-MS	Cooked and peeled tropical shrimp	Inhibition of *Brochothrix thermosphacta* by *Lactococcus piscium* CNCM I-4031 correlated with attenuation of off-odours and diminution of some volatile compounds	Fall et al., 2012
	Headspace SPME/GC-MS	Salmon gravlax	6 protective strains exhibited their own volatilome profiles. Quality improvement was not correlated with implantation of protective culture	Wiernasz et al., 2020
	Headspace SPME/GC-MS	Fresh filled pasta	*Lacticaseibacillus rhamnosus* and *Lacticaseibacillus paracasei* influenced the aroma profile with overall acceptability of the product	Tabanelli et al., 2020
	NMR spectroscopy	*in vitro*	Kinetic analysis of 11 major metabolites involved in the metabolism of *Lacticaseibacillus rhamnosus* and *Lacticaseibacillus plantarum*	Ebrahimi et al., 2016
	FTICR-MS	Red wines	No effect on the volatile compounds of a *Metschnikowia pulcherrima* bioprotective strain. Wines produced from bioprotected or sulphited must had different metabolic signatures	Simonin et al., 2020

While, on the one hand, biopreservation can have an impact on the microbiota, on the other hand, the microbiota can also affect the efficiency of biopreservation. The protective property of a nisin-producing *L. lactis* strain was tested in combination with high pressure under controlled conditions of microbiota composition in reduced-nitrite diced cooked ham (Chaillou et al., 2022). Sterile diced ham cubes were inoculated with two different microbiota collected from cooked hams, together with the protective *L. lactis* strain prior to vacuum packaging and high-pressure treatment. During storage, the two selected cooked ham microbiota were both characterised by a microbial community enriched in high potential spoilage bacterial species (especially *Proteobacteria*). Comparison of the bacterial community composition after 1 month of storage revealed that the protective effect of the *L. lactis*/high-pressure combination is highly dependent on the ham microbiota. Indeed, when ham samples were inoculated with *Pseudomonas* spp. and *Serratia* spp. rich-microbiota, *L. lactis* became dominant (>90% relative abundance). However, when ham samples were instead inoculated with microbiota dominated by *Psychrobacter* sp. and *Vibrio* sp., *L. lactis* was not competitive and *Brochothrix thermosphacta* became dominant (Chaillou et al., 2022). Thus, by interacting with the protective strain, food microbiota can act on the efficiency of biopreservation. By contrast, it is also possible that the microbiota can contribute to biopreservation in concert with the protective strains by acting directly on the unwanted microorganism through its intrinsic protective properties. In uncooked pressed cheese, the protective activity of a consortium comprising three strains belonging to the species *H. alvei, L. plantarum*, and *L. lactis* was dependent on the composition of the microbiota colonising the processed raw milk. The raw milk batches associated with the lowest growth of *E. coli* O26:H11 were characterised by greater relative abundance of lactic acid bacteria, the three *Gammaproteobacteria Acinetobacter, Serratia,* and *Hafnia*, as well as *Macrococcus*. On the other hand, the highest levels of *E. coli* O26:H11 were observed when the milk microbiota was significantly enriched with bacteria from the genera *Ramboutsia, Paeniclostridium*, and *Turicibacter* (Frétin et al., 2020).

Overall, these data were mainly produced thanks to amplicon sequencing. The major drawbacks of this approach are the biases in relative abundances resulting from PCR amplification and differences in the number of ribosomal operons between species (Edgar, 2017), the limited taxonomic resolution, mainly at the genus level even if some efforts have been made in developing specialised databases such as the DAIRYdb (Meola et al., 2019), and the lack of functional information. Amplicon sequencing targeting housekeeping genes such as *gyrB* is also promising for a better identification at the species or even intra-species level, with also less bias for relative abundance determination (Poirier et al., 2018). Investigating food communities by using shotgun metagenomics could help to improve accuracy, especially by gathering information at the species or even the strain level, and could give a global view of the functions involved in the process of biopreservation at the microbial community scale. Furthermore, network analysis of metagenetic and metagenomic data could be used to identify patterns (i.e., co-occurrence and mutual exclusion) in food microbial communities, biopreserved or not. Such an approach would enable hypotheses to be drawn regarding biotic interactions occurring between microorganisms, which could then be tested experimentally, and could be applied to the selection of potential candidates for biopreservation or for a deeper understanding of the impact of selected bioprotective cultures at a community level.

Metabolomics

In addition to their impact on food microorganisms, biopreservation agents are likely to modify the properties of the food matrix and in particular its organoleptic characteristics. Metabolomics approaches allow us to go much further than sensory analyses in the study of the impact of biopreservation agents on the matrix. SPME/GC-MS showed that cooked peeled shrimp contains a reduced amount of unwanted aldehydes and alcohols associated with sensory spoilage when *L. piscium* is used as an inhibitor of *B. thermosphacta* (Fall et al., 2012). Biopreservatives can thus play a positive sensory role by inhibiting a target microorganism responsible for spoilage. However, biopreservation can also negatively impact the matrix and can be responsible for significant changes, highlighting the interest of the polyphasic omics approach to the selection of candidate strains. An exemplary study describes the use of Head Space-Solid Phase MicroExtraction/Gas Chromatography- Mass Spectrometry (HS-SPME/GC-MS) to reveal that the impact of biopreservation candidates varied dramatically depending on the strain considered, and that specific signatures could be associated with each strain. Coupled to microbial community structure investigation by amplicon sequencing, the authors were able to rationally select two strains with the highest protective effect and the lowest sensory activity in salmon gravlax (Wiernasz et al., 2020; **Table 2**). Nuclear Magnetic Resonance (NMR) also has great potential in the study of living organisms, owing to its non-destructive nature, i.e., it can be used for *in vivo* and *in vitro* measurements of biological processes, with no quenching of the metabolism required. NMR spectroscopy can be particularly helpful for the kinetic analysis of the metabolism of protective cultures, as in the development of *in vitro* NMR kinetic measurements of lactic acid bacteria (Ebrahimi et al., 2016; **Table 2**). Such a polyphasic approach was also used successfully in wine to assess biopreservation as an alternative to sulphites (Simonin et al., 2020). The wine metabolome being of high complexity, an ultra-high-resolution mass spectrometric

method -the Fourier transform ion cyclotron resonance mass spectrometry- was used to identify and annotate more than 7,000 molecules. Clustering analysis revealed that even if the biopreservation agent has a molecular impact on the product, it is significantly lower than the winery effect, showing that biopreservation has preserved the typicality of the products, which can be of high relevance for products with protected designation of origin (Simonin et al., 2020; **Table 2**).

Limits of global omics in structured food matrices

Genomic tools based on amplicon sequencing, although powerful for food microbial ecosystem description, may show some limits for biopreserved products. Bioprotective cultures inoculated at a high level are usually dominant, at least at the beginning of storage. As it is generally admitted that taxa representing less than 0.01% of the dominant ones are not detected with metagenetics, the richness of biopreserved food may be underestimated. In the gut, the use of shotgun metagenomics has revealed that the majority of species harbour multiple strains (Ellegaard and Engel, 2016). For instance, despite a low species complexity, the infant gut microbiota exhibits extensive intra-species diversity, with an average of 4.9 strains per subject (Luo et al., 2015). In food, low resolution is a limitation in estimating the impact of protective cultures on subdominant species, such as pathogens usually present at low levels ($<10^2$ CFU/g). This problem is less important for specific spoilage organisms that generally dominate at sensory rejection time. However, the interaction between bacteria can modify the sensory characteristics (Stohr et al., 2001; Joffraud et al., 2006; Silbande et al., 2018), and it may be that underestimated taxa play a role in the deterioration of food. In addition, most efforts have been focused on bacteria, and to a lesser extent on yeasts and moulds, although phages, viruses, or *Archaea* may also be present in food microbiota (Roh et al., 2010; Park et al., 2011).

Omics are powerful tools in analysing protective cultures and their interactions at the population level. However, most of these techniques hardly consider microbial population heterogeneity in food systems that can trigger important community functions. Diversification of cell types can originate from genetic variation, ageing, gene expression stochasticity, and environmental condition fluctuation (Bury-Moné and Sclavi, 2017). In structured food matrices such as ground meat or cheese, microorganism microenvironments are highly heterogeneous and vary over time depending upon local microbial activities (Ferrier et al., 2013; Jeanson et al., 2015). Above critical textural levels, individual cells are not able to sediment or swim inside the matrix and eventually grow as large 3D microcolonies (Darsonval et al., 2021; Saint Martin et al., 2022). Microcolony size and sphericity depend on several factors such as local rheological properties, nutrient availability, competing microbiota, and associated interference interactions (Verheyen et al., 2018). Sharp gradients of nutrients and metabolites are generated inside and around the colonies, thus expanding the functional diversification of the local populations. In other structured biosystems such as bacterial biofilms (Lenz et al., 2008; Pérez-Osorio and Franklin, 2008) or gut (Consentino et al., 2021), laser capture microdissection has been put to use in applying omics to defined spatial regions or localised subpopulations of the sample. This could be of interest in deciphering the local behaviour of protective cultures and their targets in distinct and heterogeneous regions of food matrices. The few studies that have approached population functional heterogeneity in real food matrices took advantage of fluorescence microscopy associated with strains reporting the expression of genes of interest by fluorescent proteins (Fleurot et al., 2014; Hernández-Galán et al., 2017). However, food matrices are often opaque to fluorescence microscopy and the density of the population of interest can be too low for reliable quantitative measurements (e.g., bacterial pathogens $< 10^2$ CFU/g). Synthetic microbial ecology approaches, where the complexity of the communities and the factors of influence are reduced to their minimum, but are increased in their controllability, can be used to examine interactions and ecological theories (Connell et al., 2012, 2014; Rothschild, 2016; Hynes et al., 2018). Such approaches can be combined with new creative experimental designs for the study of previously unexplored aspects of bacterial behaviour in spatially structured populations (Connell et al., 2013; Wessel et al., 2013; Bridier et al., 2017). In particular, 3D bioprinting of simplified structured matrices with patterned microbial ecosystems could help study population heterogeneities and interspecies interactions at a single-cell scale in structured matrices (Moon et al., 2016; Kyle, 2018; Gyimah et al., 2021; Krishna Kumar et al., 2021). Target microorganisms could be fluorescently tagged for their geolocalisation and the feeding of spatial models of interactions (with food components and with other microorganisms during growth in the printed matrix; Krishna Kumar et al., 2021). Recently, a transcriptome-imaging approach (par-seqFISH for *parallel sequential fluorescence in situ hybridisation*) was reported to capture gene expression and spatial context within microscale assemblies at a single-cell and molecule resolution and could be put to use in such synthetic ecology approaches (Dar et al., 2021). Biopreservation studies could also benefit from microfluidic approaches to assess small-scale interactions between microorganisms (Burmeister and Grünberger, 2020).

The consideration of population heterogeneity in structured food matrices is starting to be integrated into mathematical spatial modelling and predictive microbiology (Verheyen and Van Impe, 2021). Predictive microbiology is useful to quantify both the impact of biopreservation on the food matrix and to simulate the behaviour (survival, growth, inactivation) of the undesirable target microorganism (Leroy and De Vuyst, 2003;

Koutsoumanis et al., 2004; Couvert et al., 2010; Habimana et al., 2011; Møller et al., 2013). Another important point to consider is that the contamination of food products with pathogens such as *L. monocytogenes* and *E. coli* O157:H7 occurs accidentally and usually at low levels, thus requiring single-cell level approaches. Individual-based modelling combined with microenvironment (pH, aw variabilities) modelling of vacuum-packed cold-smoked salmon was more effective in describing variability in the growth of a few *L. monocytogenes* cells than the traditional population models (Ferrier et al., 2013; Augustin et al., 2015). Such stochastic approaches need to be improved by characterising a wider range of microenvironmental factors, such as the variability of the viscosity within food matrices between liquid and solid states, as well as considering biotic factors, namely food components and food microbial communities. However, they could provide complementary information about the behaviour of unwanted microorganisms at realistic contamination levels.

Conclusion

Omics tools have become essential in the field of biopreservation, both for selecting innovative agents and for studying their effectiveness, their mechanism of action, and their impact on the food ecosystem. These approaches have revealed the diversity and complexity of the molecular mechanisms responsible for the protective activity of biopreservation agents. They have provided fundamental knowledge about biopreservation issues in terms of community description (taxonomy), biotic interactions, and impact on the organoleptic quality of the product. They have also shown that the food microbiota plays a major role in biopreservation by acting positively or negatively (**Figure 1**). It is now clear that the food microbiota must, in the future, be fully integrated into the biopreservation system engineering process to bring the field into the dimension of food microbiome engineering, so that it can play its protective role against pathogens and spoilage microorganisms as well as serve its technological purpose. Moreover, beyond the functioning of the food microbiome, this engineering process must better integrate its interconnections with other microbiomes (soil, water, plant, animal and consumer) to avoid disrupting their functioning and even to contribute to their balance. In this respect, efforts should be pursued to make more extensive use of multi-omics approaches and to combine them with other complementary approaches that take into account the heterogeneity of microorganisms at the cellular, population, and community levels, as well as the heterogeneity of the food matrix.

Author contributions

FB and MZ coordinated the work and consolidated the manuscript. All authors contributed to the design of the review, carried out the bibliographic data search, drafted the manuscript and approved the submitted version.

Funding

This work was financed by the RMT* Actia Florepro, a scientific and technical partnership in the field of biopreservation established and supported by the French Ministry responsible for Food, under the coordination of Actia. All authors of this review are members of the RMT Actia Florepro. *Réseau mixte technologique: Joint Technological Network.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Acin-Albiac, M., Filannino, P., Gobbetti, M., and Di Cagno, R. (2020). Microbial high throughput phenomics: the potential of an irreplaceable omics. *Comput. Struct. Biotechnol. J.* 18, 2290–9. doi: 10.1016/j.csbj.2020.08.010

Agriculture and Economic Development Analysis Division (2013). *The State of Food and Agriculture 2013: Food System for Better Nutrition*. Rome: FAO.

Andreevskaya, M., Johansson, P., Laine, P., Smolander, O.-P., Sonck, M., Rahkila, R., et al. (2015). Genome sequence and transcriptome analysis of meat-spoilage-associated Lactic acid bacterium *Lactococcus piscium* MKFS47. *Appl. Environ. Microbiol.* 81, 3800–11. doi: 10.1128/AEM.00320-15

Augustin, J.-C., Ferrier, R., Hezard, B., Lintz, A., and Stahl, V. (2015). Comparison of individual-based modeling and population approaches for

prediction of foodborne pathogens growth. *Food Microbiol.* 45, 205–15. doi: 10.1016/j.fm.2014.04.006

Aunsbjerg, S. D., Honoré, A. H., Marcussen, J., Ebrahimi, P., Vogensen, F. K., Benfeldt, C., et al. (2015). Contribution of volatiles to the antifungal effect of *Lactobacillus paracasei* in defined medium and yogurt. *Int. J. Food Microbiol.* 194, 46–53. doi: 10.1016/j.ijfoodmicro.2014.11.007

Back, A., Borges, F., Mangavel, C., Paris, C., Rondags, E., Kapel, R., et al. (2015). Recombinant pediocin in *Lactococcus lactis*: increased production by propeptide fusion and improved potency by co-production with PedC. *Microb. Biotechnol.* 9, 466–77. doi: 10.1111/1751-7915.12285

Baltz, R. H. (2019). Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *J. Ind. Microbiol. Biotechnol.* 46, 281–99. doi: 10.1007/s10295-018-2115-4

Barbosa, J., Albano, H., Silva, B., Almeida, M. H., Nogueira, T., and Teixeira, P. (2021). Characterization of a *Lactiplantibacillus plantarum* R23 isolated from Arugula by whole-genome sequencing and its bacteriocin production ability. *IJERPH* 18:5515. doi: 10.3390/ijerph18115515

Bassi, D., Gazzola, S., Sattin, E., Dal Bello, F., Simionati, B., and Cocconcelli, P. S. (2020). Lactic acid bacteria adjunct cultures exert a mitigation effect against spoilage microbiota in fresh cheese. *Microorganisms* 8:E1199. doi: 10.3390/microorganisms8081199

Begrem, S., Ivaniuk, F., Gigout-Chevalier, F., Kolypczuk, L., Bonnetot, S., Leroi, F., et al. (2020). New insight into antimicrobial compounds from food and marine-sourced *Carnobacterium* species through phenotype and genome analyses. *Microorganisms* 8:1093. doi: 10.3390/microorganisms8071093

Ben Said, L., Gaudreau, H., Dallaire, L., Tessier, M., and Fliss, I. (2019). Bioprotective culture: a new generation of food additives for the preservation of food quality and safety. *Ind. Biotechnol.* 15, 138–47. doi: 10.1089/ind.2019.29175.lbs

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., et al. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome* 8:103. doi: 10.1186/s40168-020-00875-0

Besnard, A., Desmasures, N., Voisin-Anastasie, A., Gréau, L., Lelièvre, V., Bré, J.-M., et al. (2021). *Aerococcus* sp. a promising genus as a source of anti-*Salmonella* bioprotective agents for the dairy industry revealed by a miniaturised screening method. *Int. Dairy J.* 116:104949. doi: 10.1016/j.idairyj.2020.104949

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., et al. (2017). antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45, W36–41. doi: 10.1093/nar/gkx319

Borges, F., and Revol-Junelles, A.-M. (2019). *Nouvelles Souches de Carnobacterium Maltaromaticum et Leurs Utilisations. French Patent No FR1911895*. Courbevoie: Institut National de la Propriété Industrielle.

Bourdichon, F., Arias, E., Bückle, A., Bello, F. D., Dubois, A., et al. (2021). The forgotten role of food cultures. *FEMS Microbiol. Lett.* 368:fnab085. doi: 10.1093/femsle/fnab085

Bridier, A., Piard, J.-C., Pandin, C., Labarthe, S., Dubois-Brissonnet, F., and Briandet, R. (2017). Spatial Organization Plasticity as an adaptive driver of surface microbial communities. *Front. Microbiol.* 8:1364. doi: 10.3389/fmicb.2017.01364

Brillet, A., Pilet, M.-F., Prévost, H., Bouttefroy, A., and Leroi, F. (2004). Biodiversity of *Listeria monocytogenes* sensitivity to bacteriocin-producing *Carnobacterium* strains and application in sterile cold-smoked salmon. *J. Appl. Microbiol.* 97, 1029–37. doi: 10.1111/j.1365-2672.2004.02383.x

Burmeister, A., and Grünberger, A. (2020). Microfluidic cultivation and analysis tools for interaction studies of microbial co-cultures. *Curr. Opin. Biotechnol.* 62, 106–15. doi: 10.1016/j.copbio.2019.09.001

Bury-Moné, S., and Sclavi, B. (2017). Stochasticity of gene expression as a motor of epigenetics in bacteria: from individual to collective behaviors. *Res. Microbiol.* 168, 503–14. doi: 10.1016/j.resmic.2017.03.009

Caldeira, C., De Laurentiis, V., Corrado, S., van Holsteijn, F., and Sala, S. (2019). Quantification of food waste per product group along the food supply chain in the European Union: a mass flow analysis. *Resour. Conserv. Recycl.* 149, 479–88. doi: 10.1016/j.resconrec.2019.06.011

Callon, C., Arliguie, C., and Montel, M.-C. (2016). Control of Shigatoxin-producing *Escherichia coli* in cheese by dairy bacterial strains. *Food Microbiol.* 53, 63–70. doi: 10.1016/j.fm.2015.08.009

Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., et al. (2019). Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* 17, 569–86. doi: 10.1038/s41579-019-0222-5

Chaillou, S., Champomier-Vergès, M.-C., Cornet, M., Crutz-Le Coq, A.-M., Dudez, A.-M., Martin, V., et al. (2005). The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. *Nat. Biotechnol.* 23, 1527–33. doi: 10.1038/nbt1160

Chaillou, S., Ramaroson, M., Coeuret, G., Rossero, A., Anthoine, V., Champomier-Vergès, M., et al. (2022). Combination of high pressure treatment at 500 MPa and biopreservation with a *Lactococcus lactis* strain for lowering the bacterial growth during storage of diced cooked ham with reduced nitrite salt. *Microorganisms* 10:456. doi: 10.3390/microorganisms10020456

Chokesajjawatee, N., Santiyanont, P., Chantarasakha, K., Kocharin, K., Thammarongtham, C., Lertampaiporn, S., et al. (2020). Safety assessment of a Nham starter culture *Lactobacillus plantarum* BCC9546 via whole-genome analysis. *Sci. Rep.* 10:10241. doi: 10.1038/s41598-020-66857-2

Comi, G., Andyanto, D., Manzano, M., and Iacumin, L. (2016). *Lactococcus lactis* and *Lactobacillus sakei* as bio-protective culture to eliminate *Leuconostoc mesenteroides* spoilage and improve the shelf life and sensorial characteristics of commercial cooked bacon. *Food Microbiol.* 58, 16–22. doi: 10.1016/j.fm.2016.03.001

Connell, J. L., Kim, J., Shear, J. B., Bard, A. J., and Whiteley, M. (2014). Real-time monitoring of quorum sensing in 3D-printed bacterial aggregates using scanning electrochemical microscopy. *Proc. Natl. Acad. Sci. U S A.* 111, 18255–60. doi: 10.1073/pnas.1421211111

Connell, J. L., Ritschdorff, E. T., Whiteley, M., and Shear, J. B. (2013). 3D printing of microscopic bacterial communities. *Proc. Natl. Acad. Sci. U S A.* 110, 18380–5. doi: 10.1073/pnas.1309729110

Connell, J. L., Whiteley, M., and Shear, J. B. (2012). Sociomicrobiology in engineered landscapes. *Nat. Chem. Biol.* 8, 10–3. doi: 10.1038/nchembio.749

Consentino, L., Rejasse, A., Crapart, N., Bevilacqua, C., and Nielsen-LeRoux, C. (2021). Laser capture microdissection to study *Bacillus cereus* iron homeostasis gene expression during *Galleria mellonella* in vivo gut colonization. *Virulence* 12, 2104–21. doi: 10.1080/21505594.2021.1959790

Cook, P. W., and Nightingale, K. K. (2018). Use of omics methods for the advancement of food quality and food safety. *Anim. Front.* 8, 33–41. doi: 10.1093/af/vfy024

Couvert, O., Pinon, A., Bergis, H., Bourdichon, F., Carlin, F., Cornu, M., et al. (2010). Validation of a stochastic modelling approach for *Listeria monocytogenes* growth in refrigerated foods. *Int. J. Food Microbiol.* 144, 236–42. doi: 10.1016/j.ijfoodmicro.2010.09.024

Cretenet, M., Nouaille, S., Thouin, J., Rault, L., Stenz, L., François, P., et al. (2011). *Staphylococcus aureus* virulence and metabolism are dramatically affected by *Lactococcus lactis* in cheese matrix: *S. aureus* interaction with *L. lactis* in cheese matrix. *Environ. Microbiol. Rep.* 3, 340–51. doi: 10.1111/j.1758-2229.2010.00230.x

Dar, D., Dar, N., Cai, L., and Newman, D. K. (2021). Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science* 373:eabi4882. doi: 10.1126/science.abi4882

Darsonval, M., Grégoire, M., Deschamps, J., and Briandet, R. (2021). "Confocal laser microscopy analysis of *Listeria monocytogenes* biofilms and spatially organized communities," in *Listeria monocytogenes. Methods Mol Biol*, eds E. M. Fox, H. Bierne, and B. Stessl (New York, NY: Springer), 123–36. doi: 10.1007/978-1-0716-0982-8_10

Delavenne, E., Déniel, F., Barbier, G., and Le Blay, G. (2012). Biodiversity of antifungal lactic acid bacteria isolated from raw milk samples from cow, ewe and goat over one-year period. *Int. J. Food Microbiol.* 155, 185–90. doi: 10.1016/j.ijfoodmicro.2012.02.003

Delbes-Paus, C., Dorchies, G., Chaabna, Z., Callon, C., and Montel, M.-C. (2010). Contribution of hydrogen peroxide to the inhibition of *Staphylococcus aureus* by *Lactococcus garvieae* in interaction with raw milk microbial community. *Food Microbiol.* 27, 924–32. doi: 10.1016/j.fm.2010.05.031

Delpech, P., Bornes, S., Alaterre, E., Bonnet, M., Gagne, G., Montel, M.-C., et al. (2015). Staphylococcus aureus transcriptomic response to inhibition by H2O2-producing *Lactococcus garvieae*. *Food Microbiol.* 51, 163–70. doi: 10.1016/j.fm.2015.05.014

Delpech, P., Rifa, E., Ball, G., Nidelet, S., Dubois, E., Gagne, G., et al. (2017). New insights into the anti-pathogenic potential of *Lactococcus garvieae* against *Staphylococcus aureus* based on RNA sequencing Profiling. *Front. Microbiol.* 8:359. doi: 10.3389/fmicb.2017.00359

Duhutrel, P., Bordat, C., Wu, T.-D., Zagorec, M., Guerquin-Kern, J.-L., and Champomier-Vergès, M.-C. (2010). Iron sources used by the nonpathogenic lactic acid bacterium *Lactobacillus sakei* as revealed by electron energy loss spectroscopy and secondary-ion mass spectrometry. *Appl. Environ. Microbiol.* 76, 560–5. doi: 10.1128/AEM.02205-09

Ebrahimi, P., Larsen, F. H., Jensen, H. M., Vogensen, F. K., and Engelsen, S. B. (2016). Real-time metabolomic analysis of lactic acid bacteria as monitored by in vitro NMR and chemometrics. *Metabolomics* 12, 1–17. doi: 10.1007/s11306-016-0996-7

Edgar, R. C. (2017). UNBIAS: an attempt to correct abundance bias in 16S sequencing, with limited success. *bioRxiv* 2017:124149. doi: 10.1101/124149

Eijsink, V. G. H., and Axelsson, L. (2005). Bacterial lessons in sausage making. *Nat. Biotechnol.* 23, 1494–5. doi: 10.1038/nbt1205-1494

El Kheir, S. M., Cherrat, L., Awussi, A. A., Ramia, N. E., Taha, S., Rahman, A., et al. (2018). High-throughput identification of candidate strains for biopreservation by using bioluminescent *Listeria monocytogenes*. *Front. Microbiol.* 9:1883. doi: 10.3389/fmicb.2018.01883

Ellegaard, K. M., and Engel, P. (2016). Beyond 16S rRNA community profiling: intra-species diversity in the gut microbiota. *Front. Microbiol.* 7:1475. doi: 10.3389/fmicb.2016.01475

Elsser-Gravesen, D., and Elsser-Gravesen, A. (2013). "Biopreservatives," in *Biotechnology of Food and Feed Additives Advances in Biochemical Engineering/Biotechnology*, eds H. Zorn and P. Czermak (Berlin: Springer), 29–49. doi: 10.1007/10_2013_234

Ercolini, D., Ferrocino, I., La Storia, A., Mauriello, G., Gigli, S., Masi, P., et al. (2010). Development of spoilage microbiota in beef stored in nisin activated packaging. *Food Microbiol.* 27, 137–43. doi: 10.1016/j.fm.2009.09.006

Even, S., Charlier, C., Nouaille, S., Ben Zakour, N. L., Cretenet, M., Cousin, F. J., et al. (2009). *Staphylococcus aureus* virulence expression is impaired by *Lactococcus lactis* in mixed cultures. *Appl. Environ. Microbiol.* 75, 4459–72. doi: 10.1128/AEM.02388-08

Fall, P. A., Pilet, M. F., Leduc, F., Cardinal, M., Duflos, G., Guérin, C., et al. (2012). Sensory and physicochemical evolution of tropical cooked peeled shrimp inoculated by *Brochothrix thermosphacta* and *Lactococcus piscium* CNCM I-4031 during storage at 8°C. *Int. J. Food Microbiol.* 152, 82–90. doi: 10.1016/j.ijfoodmicro.2011.07.015

FAO (2014). *Food Wastage Footprint: Full Cost-Accounting: Final Report*. Rome: FAO.

FAO (2019). *State of Food and Agriculture 2019. Moving Forward on Food Loss and Waste Reduction*. Rome: FAO.

Ferrier, R., Hezard, B., Lintz, A., Stahl, V., and Augustin, J.-C. (2013). Combining individual-based modeling and food microenvironment descriptions to predict the growth of *Listeria monocytogenes* on smear soft cheese. *Appl. Environ. Microbiol.* 2013:13. doi: 10.1128/AEM.01311-13

Ferrocino, I., Greppi, A., Storia, A. L., Rantsiou, K., Ercolini, D., and Cocolin, L. (2015). Impact of nisin-activated packaging on microbiota of beef burgers during storage. *Appl. Environ. Microbiol.* 2015:15. doi: 10.1128/AEM.03093-15

Finnegan, M., Linley, E., Denyer, S. P., McDonnell, G., Simons, C., and Maillard, J. Y. (2010). Mode of action of hydrogen peroxide and other oxidizing agents: differences between liquid and gas forms. *J. Antimicrob. Chemother.* 65, 2108–15. doi: 10.1093/jac/dkq308

Fleurot, I., Aigle, M., Fleurot, R., Darrigo, C., Hennekinne, J.-A., Gruss, A., et al. (2014). Following pathogen development and gene expression in a food ecosystem: the case of a *Staphylococcus aureus* isolate in cheese. *Appl. Environ. Microbiol.* 80, 5106–15. doi: 10.1128/AEM.01042-14

Frétin, M., Chassard, C., Delbès, C., Lavigne, R., Rifa, E., Theil, S., et al. (2020). Robustness and efficacy of an inhibitory consortium against *E. coli* O26:H11 in raw milk cheeses. *Food Control* 115:107282. doi: 10.1016/j.foodcont.2020.107282

Galia, W., Leriche, F., Cruveiller, S., Garnier, C., Navratil, V., Dubost, A., et al. (2017). Strand-specific transcriptomes of Enterohemorrhagic *Escherichia coli* in response to interactions with ground beef microbiota: interactions between microorganisms in raw meat. *BMC Genom.* 18:574. doi: 10.1186/s12864-017-3957-2

Gálvez, A., López, R. L., and Ben Omar, N. (2007). Bacteriocin-based strategies for food biopreservation. *Int. J. Food Microbiol.* 120, 51–70. doi: 10.1016/j.ijfoodmicro.2007.06.001

Garnier, L., Penland, M., Thierry, A., Maillard, M.-B., Jardin, J., Coton, M., et al. (2020). Antifungal activity of fermented dairy ingredients: identification of antifungal compounds. *Int. J. Food Microbiol.* 322:108574. doi: 10.1016/j.ijfoodmicro.2020.108574

Giello, M., La Storia, A., De Filippis, F., Ercolini, D., and Villani, F. (2018). Impact of *Lactobacillus curvatus* 54M16 on microbiota composition and growth of *Listeria monocytogenes* in fermented sausages. *Food Microbiol.* 72, 1–15. doi: 10.1016/j.fm.2017.11.003

Guillier, L., Stahl, V., Hezard, B., Notz, E., and Briandet, R. (2008). Modelling the competitive growth between *Listeria monocytogenes* and biofilm microflora of smear cheese wooden shelves. *Int. J. Food Microbiol.* 128, 51–7. doi: 10.1016/j.ijfoodmicro.2008.06.028

Gyimah, N., Scheler, O., Rang, T., and Pardy, T. (2021). Can 3D printing bring droplet microfluidics to every lab?—A systematic review. *Micromachines* 12:339. doi: 10.3390/mi12030339

Habimana, O., Guillier, L., Kulakauskas, S., and Briandet, R. (2011). Spatial competition with *Lactococcus lactis* in mixed-species continuous-flow biofilms inhibits *Listeria monocytogenes* growth. *Biofouling* 27, 1065–72. doi: 10.1080/08927014.2011.626124

Hebert, E. M., Saavedra, L., Taranto, M. P., Mozzi, F., Magni, C., Nader, M. E. F., et al. (2012). Genome sequence of the bacteriocin-producing *Lactobacillus curvatus* strain CRL705. *J. Bacteriol.* 194, 538–9. doi: 10.1128/JB.06416-11

Hernández-Galán, L., Cattenoz, T., Le Feunteun, S., Canette, A., Briandet, R., Le-Guin, S., et al. (2017). Effect of dairy matrices on the survival of *Streptococcus thermophilus*, *Brevibacterium aurantiacum* and *Hafnia alvei* during digestion. *Food Res. Int.* 100, 477–88. doi: 10.1016/j.foodres.2017.07.044

Hibbing, M. E., Fuqua, C., Parsek, M. R., and Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* 8, 15–25. doi: 10.1038/nrmicro2259

Gänzle, M. G., Höltzel, A., Walter, J., Jung, G., and Hammes, W. P. (2000). Characterization of reutericyclin produced by *Lactobacillus reuteri* LTH2584. *Appl. Environ. Microbiol.* 66, 4325–33. doi: 10.1128/AEM.66.10.4325-4333.2000

HPLE (2014). "Food losses and waste in the context of sustainable food systems," in *A report by the high level panel of experts on food security and nutrition. HLPE Report 8*, (Rome: FAO).

Hu, P., Xu, X. L., Zhou, G. H., Han, Y. Q., Xu, B. C., and Liu, J. C. (2008). Study of the *Lactobacillus sakei* protective effect towards spoilage bacteria in vacuum packed cooked ham analyzed by PCR-DGGE. *Meat Sci.* 80, 462–9. doi: 10.1016/j.meatsci.2008.01.011

Hynes, W. F., Chacón, J., Segrè, D., Marx, C. J., Cady, N. C., and Harcombe, W. R. (2018). Bioprinting microbial communities to examine interspecies interactions in time and space. *Biomed. Phys. Eng. Express* 4:55010. doi: 10.1088/2057-1976/aad544

Iacumin, L., Cappellari, G., Colautti, A., and Comi, G. (2020). *Listeria monocytogenes* survey in cubed cooked ham packaged in modified atmosphere and bioprotective effect of selected lactic acid bacteria. *Microorganisms* 8:898. doi: 10.3390/microorganisms8060898

Jameson, J. E. (1962). A discussion of the dynamics of *salmonella* enrichment. *J. Hyg.* 60, 193–207. doi: 10.1017/s0022172400039462

Jeanson, S., Floury, J., Gagnaire, V., Lortal, S., and Thierry, A. (2015). Bacterial colonies in solid media and foods: a review on their growth and interactions with the micro-environment. *Front. Microbiol.* 6:1284. doi: 10.3389/fmicb.2015.01284

Joffraud, J.-J., Cardinal, M., Cornet, J., Léon, S., Gigout, F., et al. (2006). Effect of bacterial interactions on the spoilage of cold-smoked salmon. *Int. J. Food Microbiol.* 112, 51–61. doi: 10.1016/j.ijfoodmicro.2006.05.014

Jones, R. J., Wiklund, E., Zagorec, M., and Tagg, J. R. (2010). Evaluation of stored lamb bio-preserved using a three-strain cocktail of *Lactobacillus sakei*. *Meat Sci.* 86, 955–9. doi: 10.1016/j.meatsci.2010.07.023

Koutsoumanis, K. P., Kendall, P. A., and Sofos, J. N. (2004). A comparative study on growth limits of *Listeria monocytogenes* as affected by temperature, pH and aw when grown in suspension or on a solid surface. *Food Microbiol.* 21, 415–22. doi: 10.1016/j.fm.2003.11.003

Krishna Kumar, R., Meiller-Legrand, T. A., Alcinesio, A., Gonzalez, D., Mavridou, D. A. I., Meacock, O. J., et al. (2021). Droplet printing reveals the importance of micron-scale structure for bacterial ecology. *Nat. Commun.* 12:857. doi: 10.1038/s41467-021-20996-w

Kyle, S. (2018). 3D printing of bacteria: the next frontier in bofabrication. *Trends Biotechnol.* 36, 340–1. doi: 10.1016/j.tibtech.2018.01.010

Laursen, M. F., Bahl, M. I., Licht, T. R., Gram, L., and Knudsen, G. M. (2015). A single exposure to a sublethal pediocin concentration initiates a resistance-associated temporal cell envelope and general stress response in *Listeria monocytogenes*. *Environ. Microbiol.* 17, 1134–51. doi: 10.1111/1462-2920.12534

Ledenbach, L. H., and Marshall, R. T. (2009). "Microbiological spoilage of dairy products," in *Compendium of the Microbiological Spoilage of Foods and Beverages*, eds W. H. Sperber and M. P. Doyle (New York, NY: Springer), 41–67. doi: 10.1007/978-1-4419-0826-1_2

Leisner, J. J., Laursen, B. G., Prevost, H., Drider, D., and Dalgaard, P. (2007). *Carnobacterium:* positive and negative effects in the environment and in foods. *FEMS Microbiol. Rev.* 31, 592–613. doi: 10.1111/j.1574-6976.2007.00080.x

Lenz, A. P., Williamson, K. S., Pitts, B., Stewart, P. S., and Franklin, M. J. (2008). Localized gene expression in *Pseudomonas aeruginosa* biofilms. *Appl. Environ. Microbiol.* 74, 4463–71. doi: 10.1128/AEM.00710-08

Leroi, F., Cornet, J., Chevalier, F., Cardinal, M., Coeuret, G., Chaillou, S., et al. (2015). Selection of bioprotective cultures for preventing cold-smoked salmon spoilage. *Int. J. Food Microbiol.* 213, 79–87. doi: 10.1016/j.ijfoodmicro.2015.05.005

Leroy, F., and De Vuyst, L. (2003). A combined model to predict the functionality of the bacteriocin-producing *Lactobacillus sakei* strain CTC 494. *Appl. Environ. Microbiol.* 69, 1093–9. doi: 10.1128/AEM.69.2.1093-1099.2003

Leyva Salas, M., Mounier, J., Maillard, M.-B., Valence, F., Coton, E., and Thierry, A. (2019). Identification and quantification of natural compounds produced by antifungal bioprotective cultures in dairy products. *Food Chem.* 301:125260. doi: 10.1016/j.foodchem.2019.125260

Leyva Salas, M., Mounier, J., Valence, F., Coton, M., Thierry, A., and Coton, E. (2017). Antifungal microbial agents for food biopreservation-A Review. *Microorganisms* 5:5030037. doi: 10.3390/microorganisms5030037

Leyva Salas, M., Thierry, A., Lemaître, M., Garric, G., Harel-Oger, M., Chatel, M., et al. (2018). Antifungal activity of lactic acid bacteria combinations in dairy mimicking models and their potential as bioprotective cultures in pilot scale applications. *Front. Microbiol.* 9:1787. doi: 10.3389/fmicb.2018.01787

Li, J., Yang, X., Shi, G., Chang, J., Liu, Z., and Zeng, M. (2019). Cooperation of lactic acid bacteria regulated by the AI-2/LuxS system involve in the biopreservation of refrigerated shrimp. *Food Res. Int.* 120, 679–87. doi: 10.1016/j.foodres.2018.11.025

Lin, X. B., Lohans, C. T., Duar, R., Zheng, J., Vederas, J. C., Walter, J., et al. (2015). Genetic determinants of reutericyclin biosynthesis in *Lactobacillus reuteri*. *Appl. Environ. Microbiol.* 2015:14. doi: 10.1128/AEM.03691-14

Liu, Y.-X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., et al. (2021). A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* 12, 315–30. doi: 10.1007/s13238-020-00724-8

Luo, C., Knight, R., Siljander, H., Knip, M., Xavier, R. J., and Gevers, D. (2015). ConStrains identifies microbial strains in metagenomic datasets. *Nat. Biotechnol.* 33, 1045–52. doi: 10.1038/nbt.3319

Marché, L., Saraoui, T., Remenant, B., Zagorec, M., Prévost, H., Delbarre-Ladrat, C., et al. (2017). Complete genome sequence of *Lactococcus piscium* CNCM I-4031, a bioprotective strain for seafood products. *Genome Announc.* 5:16. doi: 10.1128/genomeA.01510-16

Martinez, R. C. R., Alvarenga, V. O., ávaro-Trindade, C. S., Sant'Ana, A., and de, S. (2016). Assessment of the inhibitory effect of free and encapsulated commercial nisin (Nisaplin®), tested alone and in combination, on *Listeria monocytogenes* and *Bacillus cereus* in refrigerated milk. *LWT - Food Sci. Technol.* 68, 67–75. doi: 10.1016/j.lwt.2015.12.027

McNair, L. K. F., Siedler, S., Vinther, J. M. O., Hansen, A. M., Neves, A. R., Garrigues, C., et al. (2018). Identification and characterization of a new antifungal peptide in fermented milk product containing bioprotective *Lactobacillus* cultures. *FEMS Yeast Res.* 18:foy094. doi: 10.1093/femsyr/foy094

Meola, M., Rifa, E., Shani, N., Delbès, C., Berthoud, H., and Chassard, C. (2019). DAIRYdb: a manually curated reference database for improved taxonomy annotation of 16S rRNA gene sequences from dairy products. *BMC Genom.* 20:560. doi: 10.1186/s12864-019-5914-8

Metivier, A., Pilet, M.-F., Dousset, X., Sorokine, O., Anglade, P., Zagorec, M., et al. (1998). Divercin V41, a new bacteriocin with two disulphide bonds produced by *Carnobacterium divergens* V41: primary structure and genomic organization. *Microbiology* 144, 2837–44. doi: 10.1099/00221287-144-10-2837

Møller, C. O. A., Ilg, Y., Aabo, S., Christensen, B. B., Dalgaard, P., and Hansen, T. B. (2013). Effect of natural microbiota on growth of *Salmonella* spp. in fresh pork – A predictive microbiology approach. *Food Microbiol.* 34, 284–95. doi: 10.1016/j.fm.2012.10.010

Moon, S., Fritz, I. L., Singer, Z. S., and Danino, T. (2016). Spatial control of bacteria using screen printing. *3D Print Addit. Manuf* 3, 194–203. doi: 10.1089/3dp.2016.0040

Nogueira Viçosa, G., Vieira Botelho, C., Botta, C., Bertolino, M., Fernandes de Carvalho, A., Nero, L. A., et al. (2019). Impact of co-cultivation with *Enterococcus faecalis* over growth, enterotoxin production and gene expression of *Staphylococcus aureus* in broth and fresh cheeses. *Int. J. Food Microbiol.* 308:108291. doi: 10.1016/j.ijfoodmicro.2019.108291

Nouaille, S., Even, S., Charlier, C., Le Loir, Y., Cocaign-Bousquet, M., and Loubière, P. (2009). Transcriptomic response of *Lactococcus lactis* in mixed culture with *Staphylococcus aureus*. *Appl. Environ. Microbiol.* 75, 4473–82. doi: 10.1128/AEM.02653-08

Nouaille, S., Rault, L., Jeanson, S., Loubière, P., Le Loir, Y., and Even, S. (2014). Contribution of *Lactococcus lactis* reducing properties to the downregulation of a major virulence regulator in *Staphylococcus aureus*, the *agr* system. *Appl. Environ. Microbiol.* 80, 7028–35. doi: 10.1128/AEM.02287-14

Orihuel, A., Terán, L., Renaut, J., Planchon, S., Valacco, M. P., Masias, E., et al. (2019). Physiological and proteomic response of *Escherichia coli* O157:H7 to a bioprotective lactic acid bacterium in a meat environment. *Food Res. Int.* 125:108622. doi: 10.1016/j.foodres.2019.108622

Orihuel, A., Terán, L., Renaut, J., Vignolo, G. M., De Almeida, A. M., Saavedra, M. L., et al. (2018). Differential proteomic analysis of lactic acid bacteria-*Escherichia coli* O157:H7 interaction and its contribution to bioprotection strategies in Meat. *Front. Microbiol.* 9:1083. doi: 10.3389/fmicb.2018.01083

Ortiz-Rivera, Y., Sánchez-Vega, R., Gutiérrez-Méndez, N., León-Félix, J., Acosta-Muñiz, C., and Sepulveda, D. R. (2017). Production of reuterin in a fermented milk product by *Lactobacillus reuteri*: inhibition of pathogens, spoilage microorganisms, and lactic acid bacteria. *J. Dairy Sci.* 100, 4258–68. doi: 10.3168/jds.2016-11534

Park, E. J., Kim, K. H., Abell, G. C. J., Kim, M. S., Roh, S. W., and Bae, J. W. (2011). Metagenomic analysis of the viral communities in fermented foods. *Appl. Environ. Microbiol.* 77, 1284–91. doi: 10.1128/AEM.01859-10

Pérez-Osorio, A. C., and Franklin, M. J. (2008). Isolation of RNA and DNA from biofilm samples obtained by laser capture microdissection microscopy. *CSH Protoc.* 2008:rot5065. doi: 10.1101/pdb.prot5065

Pinilla, C. M. B., Stincone, P., and Brandelli, A. (2021). Proteomic analysis reveals differential responses of *Listeria monocytogenes* to free and nanoencapsulated nisin. *Int. J. Food Microbiol.* 346:109170. doi: 10.1016/j.ijfoodmicro.2021.109170

Poirier, S., Coeuret, G., Champomier-Vergès, M.-C., and Chaillou, S. (2018). Draft genome sequences of nine strains of *Brochothrix thermosphacta, Carnobacterium divergens, Lactobacillus algidus, Lactobacillus fuchuensis, Lactococcus piscium, Leuconostoc gelidum* subsp. *gasicomitatum, Pseudomonas lundensis,* and *Weissella viridescens,* a collection of psychrotrophic species involved in meat and seafood spoilage. *Genome Announc.* 6, e479–418. doi: 10.1128/genomeA.00479-18

Ramaroson, M., Guillou, S., Rossero, A., Rezé, S., Anthoine, V., Moriceau, N., et al. (2018). Selection procedure of bioprotective cultures for their combined use with High Pressure Processing to control spore-forming bacteria in cooked ham. *Int. J. Food Microbiol.* 276, 28–38. doi: 10.1016/j.ijfoodmicro.2018.04.010

Rathod, N. B., Nirmal, N. P., Pagarkar, A., Özogul, F., and Rocha, J. M. (2022). Antimicrobial impacts of microbial metabolites on the preservation of fish and fishery products: a review with current knowledge. *Microorganisms* 10:773. doi: 10.3390/microorganisms10040773

Remenant, B., Borges, F., Cailliez-Grimal, C., Revol-Junelles, A.-M., Marché, L., Lajus, A., et al. (2016). Draft genome sequence of *Carnobacterium divergens* V41, a bacteriocin-producing strain. *Genome Announc.* 4, e1109–16. doi: 10.1128/genomeA.01109-16

Riedel, C. U., Monk, I. R., Casey, P. G., Morrissey, D. O., Sullivan, G. C., Tangney, M., et al. (2007). Improved luciferase tagging system for *Listeria monocytogenes* allows real-time monitoring in vivo and in vitro. *Appl. Environ. Microbiol.* 73, 3091–4. doi: 10.1128/AEM.02940-06

Rimaux, T., Vrancken, G., Vuylsteke, B., De Vuyst, L., and Leroy, F. (2011). The pentose moiety of adenosine and inosine is an important energy source for the fermented-meat starter culture *Lactobacillus sakei* CTC 494. *Appl. Environ. Microbiol.* 77, 6539–50. doi: 10.1128/AEM.00498-11

Roh, S. W., Kim, K. H., Nam, Y. D., et al. (2010). Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. *ISME J.* 4, 1–16. doi: 10.1038/ismej.2009.83

Rothschild, L. J. (2016). Synthetic biology meets bioprinting: enabling technologies for humans on Mars (and Earth). *Biochem. Soc. Trans.* 44, 1158–64. doi: 10.1042/BST20160067

Saint Martin, C., Darsonval, M., Grégoire, M., Caccia, N., Midoux, L., Berland, S., et al. (2022). Spatial organisation of *Listeria monocytogenes* and *Escherichia coli* O157:H7 cultivated in gel matrices. *Food Microbiol.* 103:103965. doi: 10.1016/j.fm.2021.103965

Saraoui, T., Cornet, J., Guillouet, E., Pilet, M. F., Chevalier, F., Joffraud, J.-J., et al. (2017). Improving simultaneously the quality and safety of cooked and peeled shrimp using a cocktail of bioprotective lactic acid bacteria. *Int. J. Food Microbiol.* 241, 69–77. doi: 10.1016/j.ijfoodmicro.2016.09.024

Saraoui, T., Fall, P. A., Leroi, F., Antignac, J.-P., Chéreau, S., and Pilet, M. F. (2016). Inhibition mechanism of *Listeria monocytogenes* by a bioprotective bacteria *Lactococcus piscium* CNCM I-4031. *Food Microbiol.* 53, 70–8. doi: 10.1016/j.fm.2015.01.002

Saraoui, T., Leroi, F., Chevalier, F., Cappelier, J.-M., Passerini, D., and Pilet, M.-F. (2018). Bioprotective Effect of *Lactococcus piscium* CNCM I-4031 against *Listeria monocytogenes* growth and virulence. *Front. Microbiol.* 9:1564. doi: 10.3389/fmicb.2018.01564

Schlusselhuber, M., Godard, J., Sebban, M., Bernay, B., Garon, D., Seguin, V., et al. (2018). Characterization of milkisin, a novel lipopeptide with antimicrobial properties produced by *Pseudomonas* sp. UCMA 17988 isolated from bovine raw milk. *Front. Microbiol.* 9:1030. doi: 10.3389/fmicb.2018.01030

Schlusselhuber, M., Godard, J., Sebban, M., Bernay, B., Garon, D., Seguin, V., et al. (2020). Corrigendum: characterization of milkisin, a novel lipopeptide with antimicrobial properties produced by *Pseudomonas* sp. UCMA 17988 isolated from bovine raw milk. *Front. Microbiol.* 11:1323. doi: 10.3389/fmicb.2020.01323

Shi, C., and Knøchel, S. (2021a). Inhibitory effects of binary combinations of microbial metabolites on the growth of tolerant *Penicillium roqueforti* and *Mucor circinelloides*. *LWT* 149:112039. doi: 10.1016/j.lwt.2021.112039

Shi, C., and Knøchel, S. (2021b). Susceptibility of dairy associated molds towards microbial metabolites with focus on the response to diacetyl. *Food Control* 121:107573. doi: 10.1016/j.foodcont.2020.107573

Shi, C., and Maktabdar, M. (2022). Lactic acid bacteria as biopreservation against spoilage molds in dairy products – A review. *Front. Microbiol.* 12:819684. doi: 10.3389/fmicb.2021.819684

Siedler, S., Balti, R., and Neves, A. R. (2019). Bioprotective mechanisms of lactic acid bacteria against fungal spoilage of food. *Curr. Opin. Biotechnol.* 56, 138–46. doi: 10.1016/j.copbio.2018.11.015

Silbande, A., Cornet, J., Cardinal, M., Chevalier, F., Rochefort, K., Smith-Ravin, J., et al. (2018). Characterization of the spoilage potential of pure and mixed cultures of bacterial species isolated from tropical yellowfin tuna (*Thunnus albacares*). *J. Appl. Microbiol.* 124, 559–71. doi: 10.1111/jam.13663

Simonin, S., Roullier-Gall, C., Ballester, J., Schmitt-Kopplin, P., Quintanilla-Casas, B., Vichi, S., et al. (2020). Bio-protection as an alternative to sulphites: impact on chemical and microbial characteristics of red wines. *Front. Microbiol.* 11:1308. doi: 10.3389/fmicb.2020.01308

Spanu, C., Piras, F., Mocci, A. M., Nieddu, G., De Santis, E. P. L., and Scarano, C. (2018). Use of *Carnobacterium* spp protective culture in MAP packed Ricotta fresca cheese to control *Pseudomonas* spp. *Food Microbiol.* 74, 50–6. doi: 10.1016/j.fm.2018.02.020

Stiles, M. E. (1996). Biopreservation by lactic acid bacteria. *Antonie van Leeuwenhoek* 70, 331–45. doi: 10.1007/BF00395940

Stincone, P., Comerlato, C. B., and Brandelli, A. (2021). Proteomic analysis of *Listeria monocytogenes* exposed to free and nanostructured antimicrobial lipopeptides. *Mol. Omics* 17, 426–37. doi: 10.1039/D0MO00178C

Stohr, V., Joffraud, J. J., Cardinal, M., and Leroi, F. (2001). Spoilage potential and sensory profile associated with bacteria isolated from cold-smoked salmon. *Food Res. Int.* 34, 797–806. doi: 10.1016/S0963-9969(01)00101-6

Ström, K., Sjögren, J., Broberg, A., and Schnürer, J. (2002). *Lactobacillus plantarum* MiLAB 393 produces the antifungal cyclic dipeptides cyclo(L -Phe- L -Pro) and cyclo(L -Phe- *trans* -4-OH- L -Pro) and 3-phenyllactic acid. *Appl. Environ. Microbiol.* 68, 4322–7. doi: 10.1128/AEM.68.9.4322-4327.2002

Tabanelli, G., Barbieri, F., Campedelli, I., Venturini, M. C., Gardini, F., and Montanari, C. (2020). Effects of bioprotective cultures on the microbial community during storage of Italian fresh filled pasta. *Food Control* 115:107304. doi: 10.1016/j.foodcont.2020.107304

Truchado, P., Elsser-Gravesen, A., Gil, M. I., and Allende, A. (2020). Post-process treatments are effective strategies to reduce *Listeria monocytogenes* on the surface of leafy greens: a pilot study. *Int. J. Food Microbiol.* 313:108390. doi: 10.1016/j.ijfoodmicro.2019.108390

van Heel, A. J., de Jong, A., Song, C., Viel, J. H., Kok, J., and Kuipers, O. P. (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* 46, W278–81. doi: 10.1093/nar/gky383

Verheyen, D., and Van Impe, J. F. M. (2021). The inclusion of the food microstructural influence in predictive microbiology: state-of-the-art. *Foods* 10:2119. doi: 10.3390/foods10092119

Verheyen, D., érez-Rodríguez, F., Baka, M., Skåra, T., and Van Impe, J. F. (2018). Effect of food microstructure on growth dynamics of *Listeria monocytogenes* in fish-based model systems. *Int. J. Food Microbiol.* 283, 7–13. doi: 10.1016/j.ijfoodmicro.2018.05.032

Verplaetse, E., André-Leroux, G., Duhutrel, P., Coeuret, G., Chaillou, S., Nielsen-Leroux, C., et al. (2020). Heme uptake in *Lactobacillus sakei* evidenced by a new energy coupling factor (ECF)-like transport system. *Appl. Environ. Microbiol.* 2020:19. doi: 10.1128/AEM.02847-19

Viçosa, G. N., Botta, C., Ferrocino, I., Bertolino, M., Ventura, M., Nero, L. A., et al. (2018). *Staphylococcus aureus* undergoes major transcriptional reorganization during growth with *Enterococcus faecalis* in milk. *Food Microbiol.* 73, 17–28. doi: 10.1016/j.fm.2018.01.007

Wang, X., Wang, S., and Zhao, H. (2019). Unravelling microbial community diversity and succession of Chinese Sichuan sausages during spontaneous fermentation by high-throughput sequencing. *J. Food Sci. Technol.* 56, 3254–63. doi: 10.1007/s13197-019-03781-y

Warnecke, T., and Gill, R. T. (2005). Organic acid toxicity, tolerance, and production in *Escherichia coli* biorefining applications. *Microb. Cell Fact* 4:25. doi: 10.1186/1475-2859-4-25

Wessel, A. K., Hmelo, L., Parsek, M. R., and Whiteley, M. (2013). Going local: technologies for exploring bacterial microenvironments. *Nat. Rev. Microbiol.* 11, 337–48. doi: 10.1038/nrmicro3010

Wiernasz, N., Cornet, J., Cardinal, M., Pilet, M.-F., Passerini, D., and Leroi, F. (2017). Lactic acid bacteria selection for biopreservation as a part of hurdle technology approach applied on seafood. *Front. Mar. Sci.* 4:119. doi: 10.3389/fmars.2017.00119

Wiernasz, N., Leroi, F., Chevalier, F., Cornet, J., Cardinal, M., Rohloff, J., et al. (2020). Salmon Gravlax biopreservation with lactic acid bacteria: a polyphasic approach to assessing the impact on organoleptic properties, microbial ecosystem and volatilome composition. *Front. Microbiol.* 10:3103. doi: 10.3389/fmicb.2019.03103

Windholtz, S., Redon, P., Lacampagne, S., Farris, L., Lytra, G., Cameleyre, M., et al. (2021). Non-*Saccharomyces* yeasts as bioprotection in the composition of red wine and in the reduction of sulfur dioxide. *LWT* 149:111781. doi: 10.1016/j.lwt.2021.111781

World Health Organization (2015). *WHO Estimates of the Global Burden of Foodborne Diseases: Foodborne Disease Burden Epidemiology Reference Group 2007-2015*. Geneva: World Health Organization.

Wu, W., Deng, G., Liu, C., Gong, X., Ma, G., Yuan, Q., et al. (2020). Optimization and multiomic basis of phenyllactic acid overproduction by *Lactobacillus plantarum*. *J. Agric. Food Chem.* 68, 1741–9. doi: 10.1021/acs.jafc.9b07136

Yi, L., Luo, L., and Lü, X. (2018). Efficient exploitation of multiple novel bacteriocins by combination of complete genome and peptidome. *Front. Microbiol.* 9:1567. doi: 10.3389/fmicb.2018.01567

Zagorec, M., and Champomier-Vergès, M.-C. (2017). *Lactobacillus sakei*: a starter for sausage fermentation, a protective culture for meat products. *Microorganisms* 5:56. doi: 10.3390/microorganisms5030056

Zdenkova, K., Alibayov, B., Karamonova, L., Purkrtova, S., Karpiskova, R., and Demnerova, K. (2016). Transcriptomic and metabolic responses of *Staphylococcus aureus* in mixed culture with *Lactobacillus plantarum*, *Streptococcus thermophilus* and *Enterococcus durans* in milk. *J. Ind. Microbiol. Biotechnol* 43, 1237–47. doi: 10.1007/s10295-016-1794-y

Zhang, Y., Zhu, L., Dong, P., Liang, R., Mao, Y., Qiu, S., et al. (2018). Bio-protective potential of lactic acid bacteria: effect of *Lactobacillus sakei* and *Lactobacillus curvatus* on changes of the microbial community in vacuum-packaged chilled beef. *Asian Austral. J. Anim. Sci.* 31, 585–94. doi: 10.5713/ajas.17.0540

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership