

Ecological insights into the microbiology of food using metagenomics and its potential surveillance applications

Samuel J. Bloomfield^{1,2}, Falk Hildebrand^{1,2,3}, Aldert L. Zomer⁴, Raphaëlle Palau^{1,2} and Alison E. Mather^{1,2,5,*}

Abstract

A diverse array of micro-organisms can be found on food, including those that are pathogenic or resistant to antimicrobial drugs. Metagenomics involves extracting and sequencing the DNA of all micro-organisms on a sample, and here, we used a combination of culture and culture-independent approaches to investigate the microbial ecology of food to assess the potential application of metagenomics for the microbial surveillance of food. We cultured common foodborne pathogens and other organisms including *Escherichia coli*, *Klebsiella/Raoultella* spp., *Salmonella* spp. and *Vibrio* spp. from five different food commodities and compared their genomes to the microbial communities obtained by metagenomic sequencing following host (food) DNA depletion. The microbial populations of retail food were found to be predominated by psychrotrophic bacteria, driven by the cool temperatures in which the food products are stored. Pathogens accounted for a small percentage of the food metagenome compared to the psychrotrophic bacteria, and cultured pathogens were inconsistently identified in the metagenome data. The microbial composition of food varied amongst different commodities, and metagenomics was able to classify the taxonomic origin of 59% of antimicrobial resistance genes (ARGs) found on food to the genus level, but it was unclear what percentage of ARGs were associated with mobile genetic elements and thus transferable to other bacteria. Metagenomics may be used to survey the ARG burden, composition and carriage on foods to which consumers are exposed. However, food metagenomics, even after depleting host DNA, inconsistently identifies pathogens without enrichment or further bait capture.

Impact Statement

Multiple micro-organisms can be found on food, including those that are dangerous to humans and those that carry antimicrobial resistance genes. Metagenomics involves analyzing all the microbial DNA on a sample, and in this study, we compared the metagenomes of food samples with the genomes of pathogens cultured from the same samples to determine if metagenomics could be used for food pathogen surveillance. Retail food samples were predominated by psychrotrophic bacteria that thrive in the cool temperatures at which food is stored. Pathogens accounted for a small percentage of the food metagenome compared to the psychrotrophic bacteria, and cultured pathogens were inconsistently identified in the metagenome data. For 59% of antimicrobial resistance genes (ARGs) found on food, metagenomics could predict what bacterial genus to which they belonged. However, it was unclear what percentage of ARGs could easily be transferred between bacteria. Metagenomics may be used to survey the burden of ARG on food but is currently unable to identify food pathogens effectively.

Received 06 August 2024; Accepted 22 November 2024; Published 03 January 2025

Author affiliations: ¹Quadram Institute Bioscience, Norwich Research Park, Norwich, UK; ²Centre for Microbial Interactions, Norwich Research Park, Norwich, UK; ³Earlham Institute, Norwich Research Park, Norwich, UK; ⁴Utrecht University, Utrecht, Netherlands; ⁵University of East Anglia, Norwich, UK.

*Correspondence: Alison E. Mather, alison.mather@quadram.ac.uk

Keywords: antimicrobial resistance; food; metagenomes; pathogens.

Abbreviations: AMR, antimicrobial resistance; ARG, antimicrobial resistance gene; BWA, Burrows–Wheeler aligner; ISs, insertion sequences; KEGG, Kyoto Encyclopedia of Genes and Genomes; KO, Kyoto Encyclopedia of Genes and Genomes Orthology; MAG, metagenome-assembled genome; MGSs, metagenome species; SCFA, short-chain fatty acid; ST, sequence type; VFDB, Virulence Finder Database.

Data statement: Fifteen supplementary figures and five supplementary tables are available with the online version of this article.

001337 © 2025 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

DATA SUMMARY

All raw reads generated as part of this study were uploaded to the Sequence Read Archive (SRA) under the following projects: PRJNA849983 (food metagenome Illumina), PRJNA1034280 (chicken metagenome PromethION), PRJNA1107355 (leafy greens, pork and salmon metagenome PromethION), PRJNA939716 (*Salmonella* Illumina), PRJNA1135353 (*Klebsiella* Illumina) and PRJNA1107692 (*Escherichia coli* and *Vibrio* Illumina). The R scripts used to analyse food metagenomes were uploaded to Zenodo: doi.org/10.5281/zenodo.13995543.

INTRODUCTION

Global trade provides the opportunity to consume food from around the world, originating from a wide range of animals and plants and produced using a range of methods. A safe food chain relies on the identification and minimization of food pathogens [1] and the identification of other micro-organisms that may indirectly influence a consumer's health, such as antimicrobial-resistant (AMR) bacteria [2]. New molecular techniques are being developed to identify micro-organisms quickly, such as loop-mediated isothermal amplification to detect food pathogens [3], and at higher resolution, such as metagenomics that can identify micro-organisms at the strain level [4], but their effectiveness needs to be assessed before their utility for food surveillance can be determined.

It is estimated that foodborne pathogens are responsible for 2.4 million infections in the UK each year, including well-known bacterial pathogens such as *Campylobacter* spp., *Salmonella* spp. and *Listeria monocytogenes* [5]. However, not all individuals with foodborne illness will seek medical attention, and fewer still will submit a clinical sample for testing [6, 7], preventing the causative agent and putative sources from being identified. Consumers are at risk when causative agents of foodborne illness remain undetected, typically due to traditional approaches being targeted to specific pathogens (e.g., culture-based, immunoassay and molecular methods) [1, 8] or new foodborne diseases emerging and not being recognized [9]. As such, methods that can identify all micro-organisms on food in a timely, accurate and reproducible manner are required.

Pathogen identity is not the only important factor for food safety, as not all pathogens within the same species represent the same risk to human health. Pathogenic bacteria often contain virulence genes that encode for molecules that help to colonize, move between and evade host defenses, as well as obtain nutrients from their environment [10]; virulence genes are often used to distinguish pathogenic and non-pathogenic strains [11]. For this reason, multiple molecular methods have been developed to identify these virulence genes and the pathogenic strains carrying them [12].

The human burden of foodborne disease is further compounded by AMR. Although many foodborne illnesses are self-limiting, severe infections can occur that may require antimicrobial treatment, which is complicated by AMR. Pathogens and non-pathogenic bacteria on food can also act as a reservoir of antimicrobial resistance genes (ARGs) [13].

Beyond this, metal-tolerance genes are often co-selected with ARGs [14], with metals often used in the agricultural and horticultural industries for their bactericidal and growth-promoting effects [15, 16]. Most studies of metal-tolerant bacteria on food focus on pathogens [17], although many of the non-pathogenic bacteria on food are known to be metal tolerant [18].

The gut microbiota consists of all the micro-organisms in the gastrointestinal tract and is influenced by diet [19]. The gut microbiota contain many metabolic pathways that can utilize and produce metabolites [20]. Collectively, the conserved sequences of reactions that encode for metabolite utilization or production are referred to as metabolic modules [21]. Food varies in chemical compositions, and certain macromolecules in food promote certain micro-organisms in the diet [19]. However, food may also influence the gut microbiota by introducing new strains that colonize the gut.

Previously, we developed a method for depleting host (food) DNA from food and applied it to 109 food samples [22]. In this study, for the 109 samples, we compared the genomes of cultured pathogens to the short-read metagenomes, supplementing a subset of samples with long-read metagenomes, with the purpose of obtaining a deeper understanding of the microbial ecology of foods and evaluating the usefulness of metagenomics in improving food surveillance and safety [23]. We also investigated microbial signatures amongst different food commodities and the ability of metagenomics to predict what bacteria contribute to the ARG reservoir on food.

METHODS

Samples

The 109 food metagenomes analyzed in this study had previously undergone host DNA depletion before they were sequenced on an Illumina NextSeq [22]. They consisted of 28 chicken, 24 leafy green, 15 pork, 18 prawn and 24 salmon samples and were collected at retail throughout Norfolk County of the UK, as previously described [22]. The food samples were collected on 14 sampling trips, and a blank was run for each sampling trip that consisted of buffered peptone water that underwent the same host DNA depletion and sequencing was performed as for the food samples.

Pathogen culturing

The food samples from which the metagenomes were extracted were also cultured for *Escherichia coli*, *Klebsiella/Raoultella* spp., *Salmonella* spp. and, in the case of seafood samples, *Vibrio* spp., as described previously [1]. Multiple isolates from each food sample were selected and analyzed further. For *E. coli*, up to four isolates per sample were analyzed; for *Salmonella*, up to eight isolates per sample were analyzed; and for *Klebsiella/Raoultella* and *Vibrio*, up to two isolates per sample were analyzed.

Genomes were extracted using the Maxwell RSC Cultured Cells DNA Kit (Promega, Madison, WI, USA). Libraries were created using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) and sequenced on a NextSeq 550 System (Illumina) as 150 bp paired-end reads.

Genomic Illumina reads were trimmed using fastp v0.19.5 [24] to trim reads with a minimum quality value of 20 and default other parameters. Trimmed genomic reads were assembled using SPAdes v3.11.1 [25] in ‘careful’ mode. The quality of the assemblies was assessed using QUAST v4.6.3 [26] and CheckM v1.1.2 [27] and by aligning reads to the assemblies using the Burrows–Wheeler aligner (BWA) v0.7.17 [28]. Assemblies were accepted if they consisted of less than 500 contigs that were over 500 bp and less than 50 duplicate genes and had a mean read depth of the 4 largest contigs above 30.

MLST v2.16.1 (<https://github.com/tseemann/mlst>) was used to predict the sequence types (STs) of the assembled genomes.

Assembly-independent metagenomics

The raw paired Illumina reads of the 109 food samples and 14 blanks that were previously sequenced at 8 GB per metagenome [22] were processed using fastp similar to genomes. The blanks had been spiked with PhiX, and these reads were filtered out using BBSplit v38.75 [29] and the phiX174 genome (SAMN04281799).

MetaPhlAn v2.6.0 [30] was used to predict the taxonomic origin of the trimmed reads and to identify if reads associated with the cultured pathogens could be detected. In R v4.1.2 [31], phyloseq v1.38.0 (<https://github.com/joey711/phyloseq>) was used to rarefy the samples to the minimum sample total read depth, and ALDEx2 v1.26.0 (https://github.com/ggloor/ALDEx2_dev) was used to identify taxa associated with food commodity using a generalized linear model, and the level of significance was adjusted for multiple comparisons using the Benjamini–Hochberg procedure. A large percentage of the blanks’ reads were classified as *E. coli* reads, likely because the PhiX they were spiked with originated from *E. coli* and PhiX read filtering did not remove all of them.

MetaMLST v1.2.2 [32] was used to identify bacterial STs in food metagenomes using the paired trimmed reads. The STs identified in metagenomes were compared to the STs of pathogens cultured from the same samples.

To identify pathogen genomes within metagenomes, trimmed metagenome Illumina reads were aligned to the assembled genomes using BWA, and the coverage was calculated using SAMtools v1.9 [33]. For each genome, the percentage of the assembly to which metagenomic reads were aligned was calculated.

Genome reconstruction and taxonomic and functional profiling of metagenomes

To investigate the differences in the food microbiota at the species level, MATAFILER was used to process metagenome raw reads, assemble metagenomes and reconstruct and dereplicate metagenome-assembled genomes (MAGs) [34]. Raw metagenomic reads were quality filtered using sdm v1.63 with default parameters [35]. Kraken2 [36] was used to remove metagenome reads associated with hosts using reference genomes of the organisms from which the food originated (Table S1, available in the online version of this article). Host-filtered metagenome reads were assembled using MEGAHIT v1.2.9 [37], and reads were back mapped to the assembly using Bowtie2 v2.3.4.1 [38]. Genes were predicted with Prodigal v2.6.1 with parameters ‘-p meta’ [39], and a gene catalogue clustered at 95% nt identity using MMseqs2 [40]. Matrix operations on the gene catalogue were carried out using rtk [41].

MAGs were calculated using SemiBin2 [42], and their completeness and contamination were estimated using CheckM2 [43]. Using a combination of genome bins (via canopy clusters [44] as implemented in canopy2, <https://github.com/hildebra/canopy2>) and SemiBin2 MAGs, high-quality MAGs (>80% completeness and <5% contamination) [34] were dereplicated using clusterMAGs (<https://github.com/hildebra/clusterMAGs>). The GTDB bacterial core genes were used to guide dereplication. The dereplication procedure uses a canopy clustering approach, requiring at least (weighted) >80% matching in marker genes between two MAGs to collapse into the same species. These dereplicated MAGs are referred to as metagenome species (MGSs). MGSs associated with individual food commodities were identified based on the relative abundance of the MGSs after rarefaction using ALDEx2 as previously described.

To identify functional groups that differed between metagenomes, gene annotation was performed on the metagenome assemblies as described by Frioux *et al.* [45]. Briefly, the gene catalogue was annotated to the eggNOG database [46] using the BLASTP module of Diamond [47]. Higher-level Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) categories were downloaded from the KEGG database [48]. The KO assignments were used to predict conserved sequences of reactions that make up metabolic

modules represented using GOMixer [49, 50]. ALDEx2 was used to determine if food commodity was associated with the relative abundance of metabolic modules after rarefaction as described previously.

KMA v1.2.3t [51] was used to identify ARGs using the ResFinder v3.2 database [52] and virulence genes using the Virulence Finder Database (VFDB) v6.0 [53]. Lapidary v0.5.0 [54] was used to identify metal-tolerance genes using the BacMet database [55]. For AMR, virulence and metal-tolerance genes, 90% identity and 60% coverage cut-offs were used. The number of reads that aligned to these genes was compared to the number of post-host DNA-removed reads. phyloseq was used to rarefy the samples to the minimum sample post-host DNA-removed reads. ALDEx2 was used to look for an association between the genes of interest and food commodity as described previously.

Read depth and pathogen identification comparison

A linear regression model was used to determine if the number of sequenced microbial reads or relative abundance of pathogen was associated with the detection of pathogens. The outcome variable was the percentage of the pathogen genome assembly that reads from the associated metagenome aligned to using BWA, and the explanatory variables were the number of microbial reads (post-Kraken2 host removal), relative abundance of the specific pathogen (calculated by MetaPhlAn2) and the pathogen identity. A partial F-test was used to determine if pathogen identity significantly affected the model.

Long-read analysis

For the 24 food metagenomes that had a DNA concentration above 30 ng μl^{-1} , libraries were formed using the Ligation Sequencing Kit (SQK-LSK110) (Oxford Nanopore Technologies, Oxford, UK) and barcoded using Native Barcoding Expansion 1–12 (EXP-NBD104) and 13–24 (EXP-NBD114) (Oxford Nanopore Technologies). These libraries were sequenced on a PromethION (Oxford Nanopore Technologies) at a target depth of 8 GB per metagenome. Bases were called using Guppy v6.2.1 (<https://nanoporetech.com/document/Guppy-protocol>) with super accuracy basecalling. The raw long reads were trimmed using NanoFilt v0.1.0 (<https://github.com/wdecoster/nanofilt>). Hybrid assemblies with trimmed long and short reads were assembled using OPERA-MS v0.9.0 [56].

ARGs, virulence genes and plasmid replicons were identified in hybrid assemblies using Abricate v1.01 (<https://github.com/tseemann/abricate>) with the ResFinder v3.6, VFDB v2.0.4 and PlasmidFinder v2.1.3 [57] databases, respectively. Metal-tolerance genes were identified using TBLASTN v2.14.0 [58] and the BacMet v2.0 database. For Abricate and TBLASTN, 90% identity and 60% coverage cut-offs were used. For ARG-, metal-tolerance- and virulence-containing contigs, the gene of interest was blocked out and the taxonomic origin of the contigs was predicted using Kraken v2.1.1 [36] and the maxikraken2_1903_140 GB databases. ISEScan v1.7.2.3 [59] was used to identify insertion sequences (ISs), and those classified as ‘complete’ were investigated further. Genes of interest within 10 kbp of ISs were classified as associated with ISs.

The percentage of the metagenome comprised of each gene of interest was calculated by multiplying the gene length by the mean Illumina read depth of the contigs on which the genes were located and dividing by the number of post-host DNA-removed sequenced bases of the metagenome.

To measure the accuracy of contig-based gene taxonomic classifications, the analysis was repeated on the genomes of the bacteria cultured from the same samples as the metagenomes, including masking the genes of interest, except with 90% identity and coverage cut-offs. The bacterial genomes sequenced in this study consisted of well-described pathogens, whilst the metagenomes contained many poorly described micro-organisms. A lower coverage cut-off was used when identifying genes of interest in metagenomes (60%) compared to genomes in case the poorly described pathogens contained genes distantly related to those in the gene databases and would be missed with a stricter coverage cut-off [60].

RESULTS

Metagenome-genome comparisons

Analyses of the metagenomic reads from food samples using MetaPhlAn2 revealed a large amount of variation in the microbial composition of food (Fig. 1); the most abundant genera were *Pseudomonas* (39%), *Acinetobacter* (14%), *Psychrobacter* (6.8%), *Shewanella* (5.6%) and *Carnobacterium* (5.0%). For each sampling run, different batches of buffered peptone water and other reagents were used as sampling took place over a 7-month period, explaining the differences in the microbial composition of the blanks.

To test the ability of metagenomics to accurately detect pathogens on food, we cultured specific pathogenic bacteria from the 109 food samples. Culturing results varied amongst different food commodities (Table S2), as described in previous work [1]. Of the 109 food samples tested, 62% were positive for *E. coli*, 31% for *Klebsiella/Raoultella* and 2.8% for *Salmonella* through culture. Of the 42 seafood samples tested, 26% were positive for *Vibrio* through culture. For 98% of metagenomes, the relative abundance of pathogens was less than 10% compared to the other micro-organisms in the metagenome, regardless of whether or not the

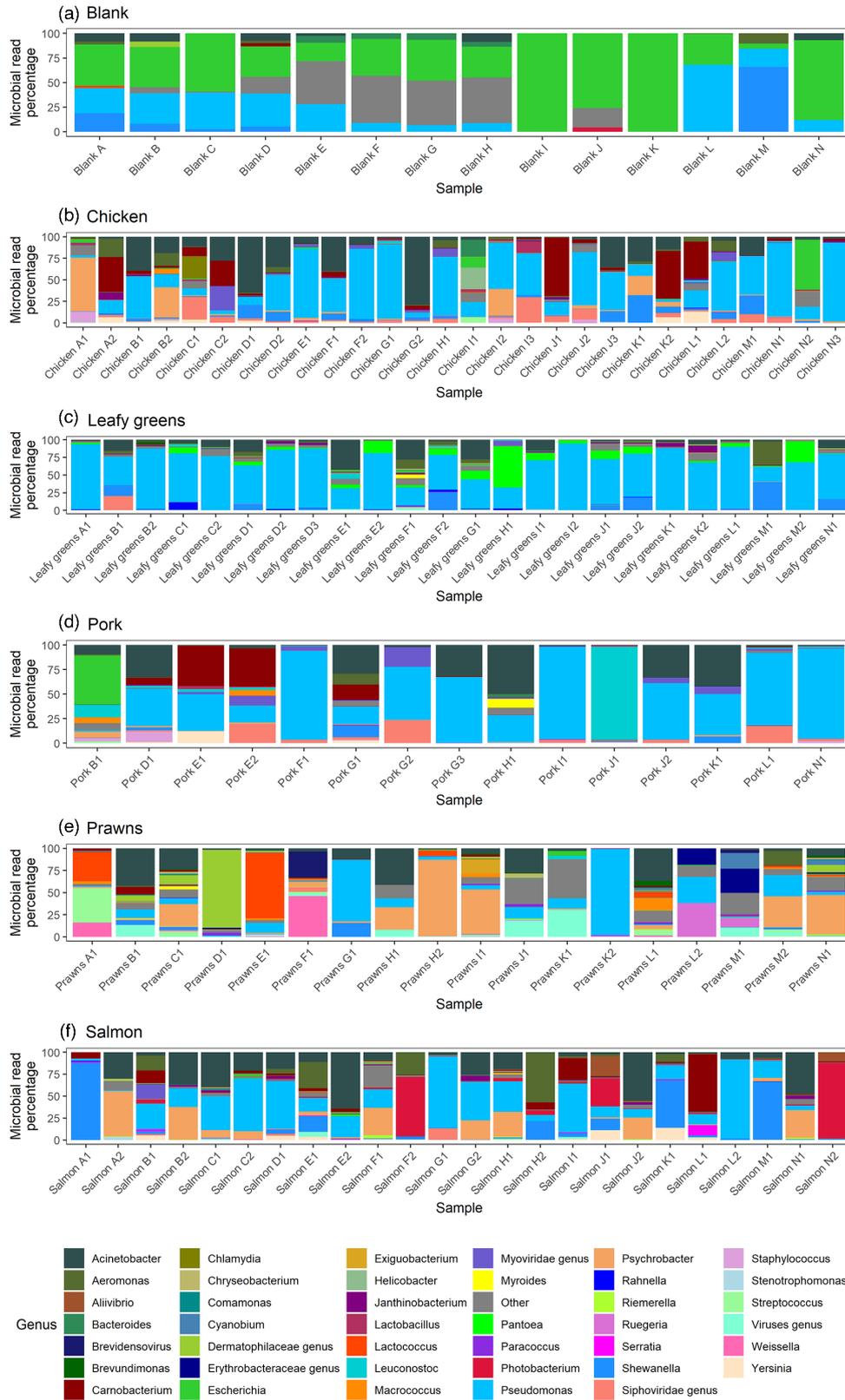


Fig. 1. Barplots of the microbial reads classified at the genus level for the blanks (a), chicken (b), leafy green (c), pork (d), prawn (e) and salmon (f) samples for 109 short-read sequenced metagenomes and 14 blanks. The top 40 most common genera were coloured, and the rest were combined into an 'other' category.

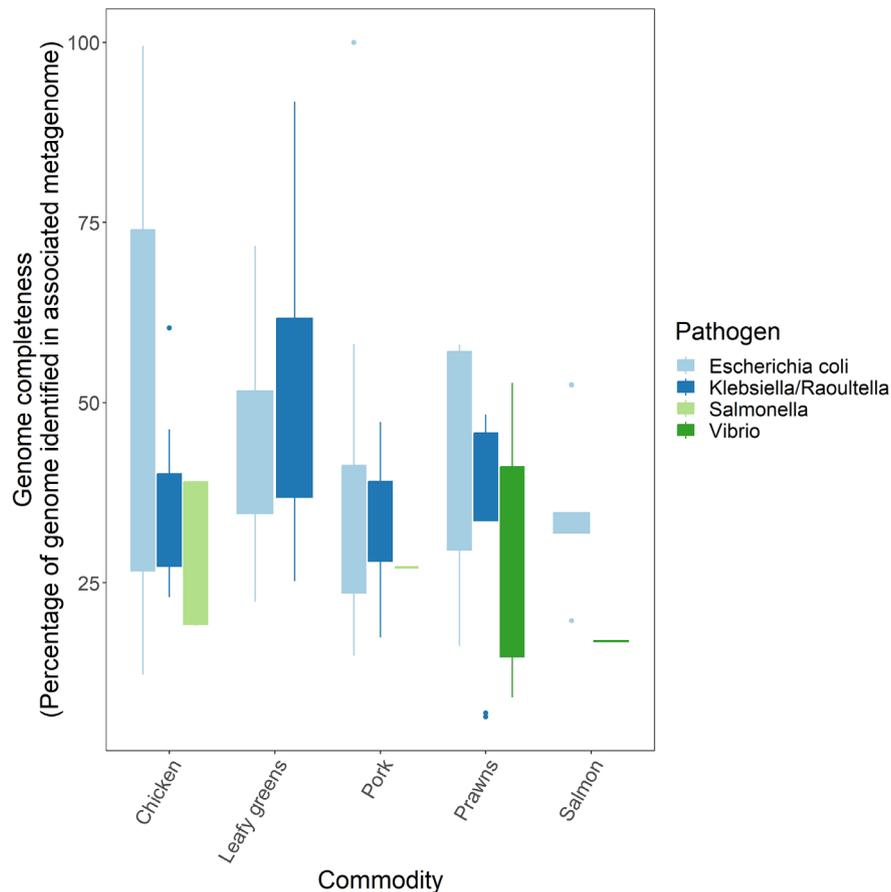


Fig. 2. Percentage of genome assemblies of pathogens cultured from food identified by the associated short-read sequenced metagenome, separated by food commodity and coloured by taxa.

pathogen was recovered from the sample through culture. The exception was a chicken and a pork sample, where the relative abundance of *E. coli* was 49% and 27%, respectively. *E. coli* was cultured from these samples (Fig. S1).

The ability of metagenomes to detect low-abundant pathogens was measured by investigating if MetaPhlAn2 found any reads of the pathogens cultured from food. When compared to culture results, MetaPhlAn2 was 0–38% sensitive and 78–100% specific for the different pathogens investigated (Table S3).

After removing host reads from the food metagenomes, three metagenomes (one chicken and two prawn samples) had too few reads for subsequent assemblies. These samples were excluded from further analysis.

For the isolated pathogens, the associated metagenomes aligned to 6.3–99.9% of their genome assemblies (Fig. 2). The percentage of the genome assembly identified was positively associated with the number of microbial reads sequenced (linear regression: $P=3.3 \times 10^{-4}$) and the relative abundance of the specific pathogen compared to other micro-organisms in the metagenome (linear regression: $P=7.3 \times 10^{-16}$). Pathogen identity was also associated with the percentage of the genome assembly identified (partial F-test: $P=0.011$). The pork metagenome consisting of 27% *E. coli* microbial reads had three *E. coli* isolates cultured from it that belonged to ST 2967, and 99.9% of their assemblies were identified via metagenomic sequencing. The chicken sample consisting of 49% *E. coli* microbial reads had three *E. coli* isolates cultured from it that belonged to different STs, 57, 131 and 8611, and 89.5–99.6% of their assemblies were identified in the metagenome. For the rest of the pathogens cultured from food, 6–98% of their genome assemblies were found in the associated metagenome.

MetaMLST was used to identify STs in the metagenomes and was able to identify STs of at least 1 bacterial species in 21 of the 109 food samples analyzed (Table S4). However, most of the STs were unknown, with only three known STs identified. The only known pathogen ST identified was *E. coli* ST 353. This was identified in the chicken metagenome where this species comprised 49% of the microbial reads, but none of the *E. coli* isolates cultured from this sample belonged to this ST.

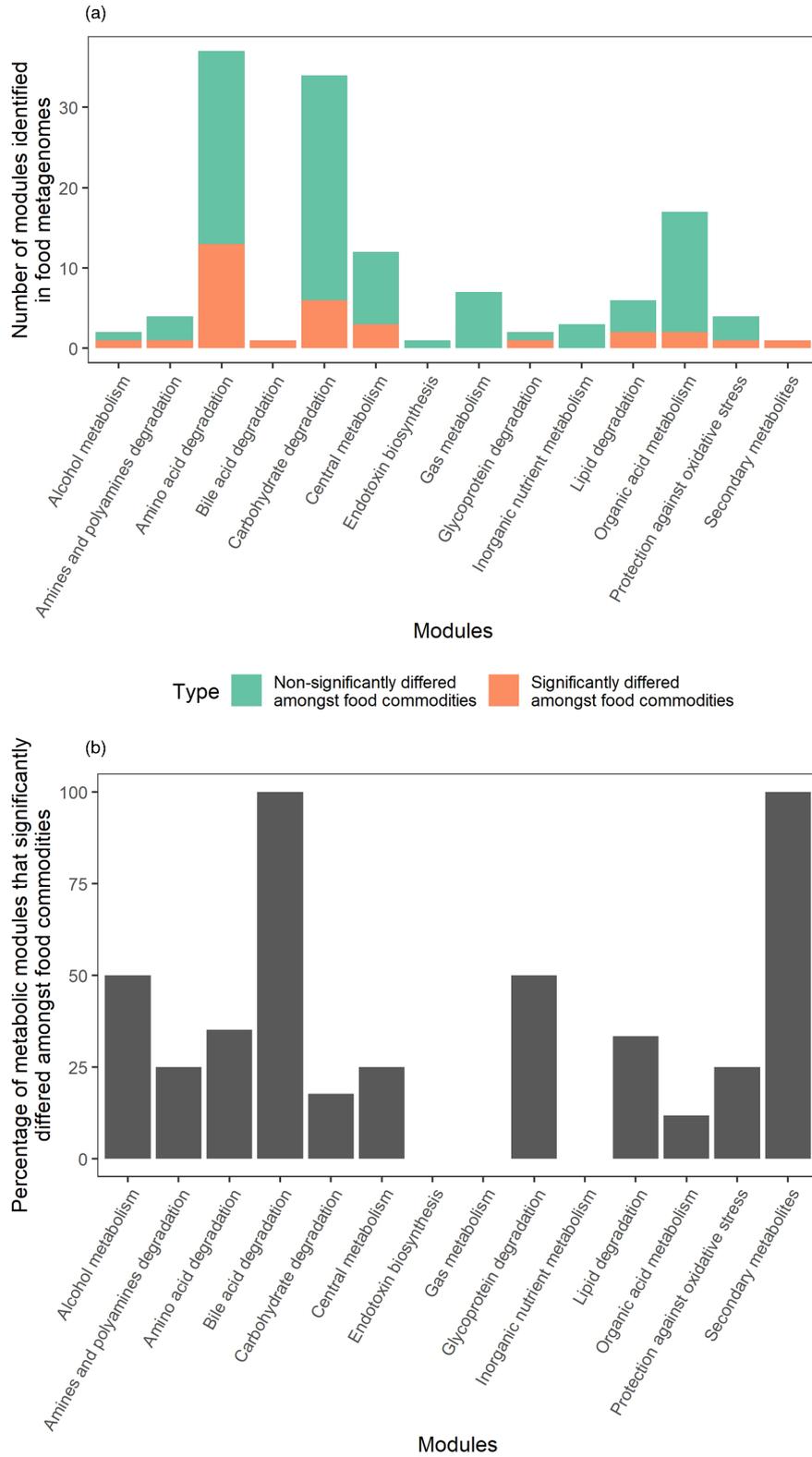


Fig. 3. Number (a) and percentage (b) of metabolic modules identified amongst 106 short-read sequenced food metagenomes whose assemblies passed quality control and that significantly differed between samples from different food commodities.

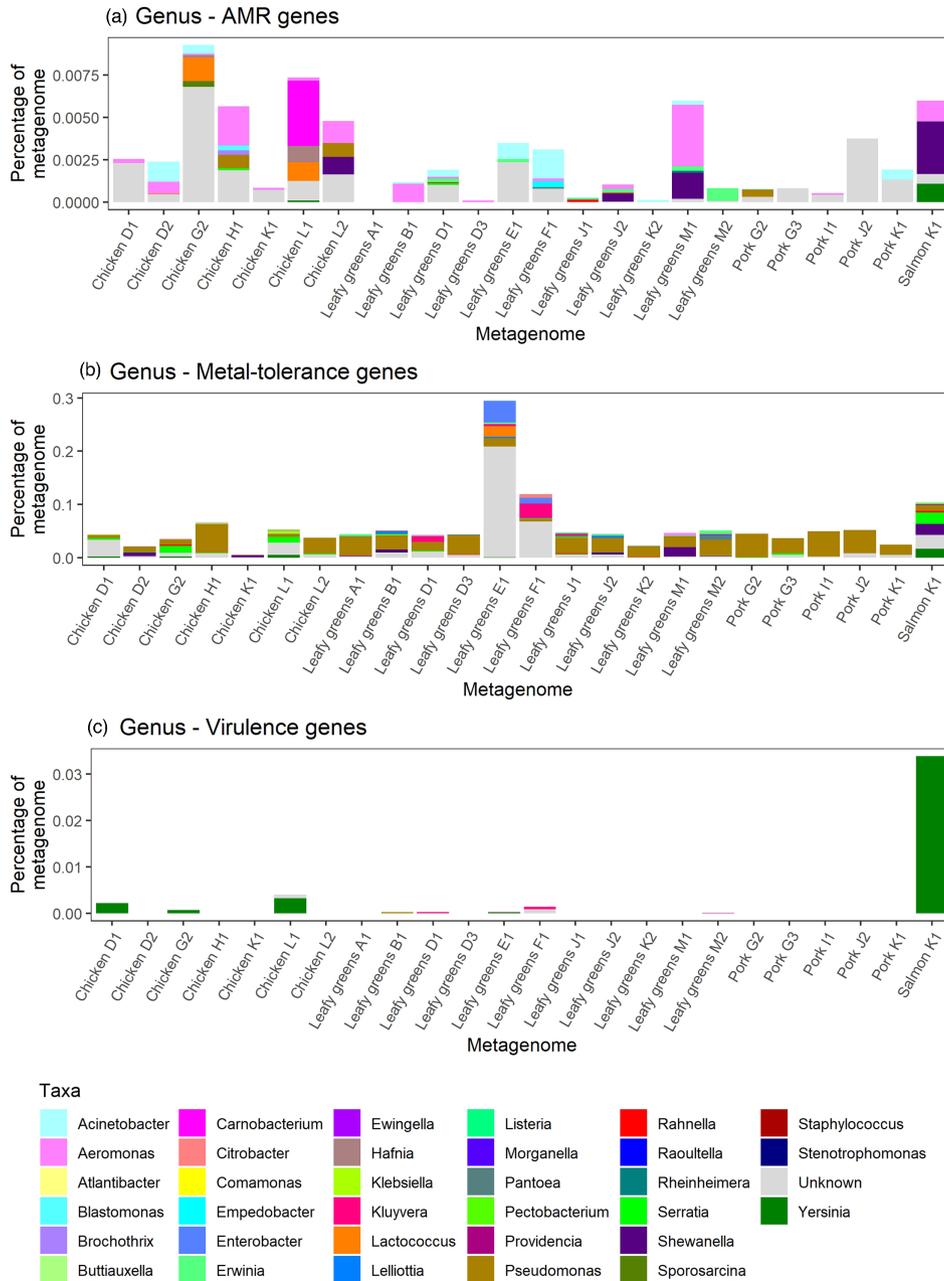


Fig. 4. Percentage of metagenomes comprised of antimicrobial resistance (AMR) (a), metal-tolerance (b) and virulence (c) genes identified in 24 long-read metagenome assemblies and coloured by the predicted taxa at the genus level.

Metagenomic microbial composition and metabolic modules

Metagenomic sequences were further processed using MATAFILIER [34] to reconstruct *de novo* MGSs associated with food commodity. MGS analysis identified 111 taxa at the species level across the 106 food metagenomes with sufficient host-removed reads to assemble that were each represented by at least one high-quality reference MAG (>80% completeness and <5% contamination). MGS relative abundance analysis using ALDEX2 identified five taxa that differed between food commodities. Compared to other food commodities, leafy greens were associated with a higher relative abundance of *Pantoea agglomerans*; chicken and pork were associated with a higher relative abundance of *Brochothrix thermosphacta*, *Pseudomonas paraversuta* and *Pseudomonas weihenstephanensis*; and chicken, pork and salmon were associated with a higher relative abundance of *Acinetobacter harbinensis* (Fig. S3). Using non-metric multidimensional scaling, there was some overlap amongst chicken, pork, prawn and salmon samples, whilst leafy green samples clustered separately to the other food commodities (Fig. S4). Additional taxa were

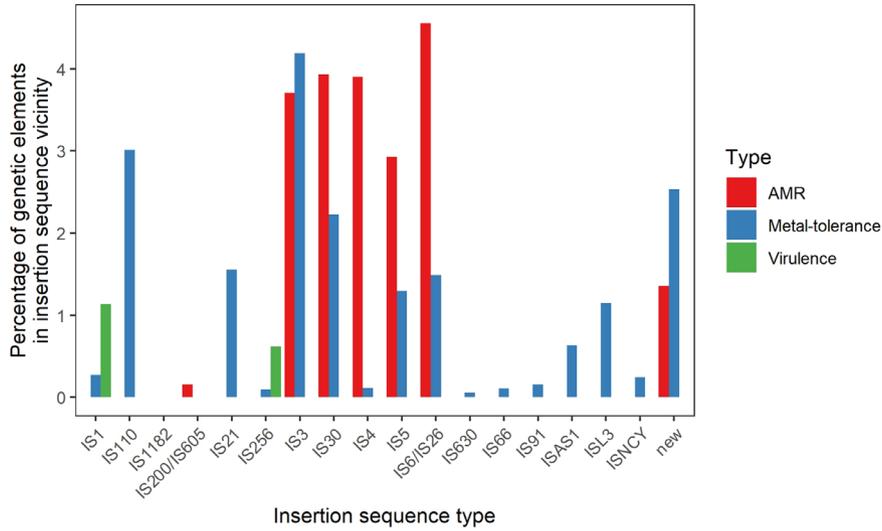


Fig. 5. Percentage of AMR, metal-tolerance and virulence genes in the presence of different insertion sequence types for the 24 long-read food metagenomes.

associated with food commodity when reads were taxonomically classified using MetaPhlan, although MetaPhlan struggled to classify some taxa at the species level (Fig. S5).

Metabolic analysis identified 131 metabolic modules in the food metagenomes. Thirty-two modules significantly differed between food commodities (Fig. S6). The 131 metabolic modules were classified into 14 functional groups, and the metabolic modules that differed amongst food commodities were not associated with functional group (Fisher’s exact test: $P=0.183$) (Fig. 3). Leafy green samples had a higher relative abundance of pectin degradation compared to other food commodities.

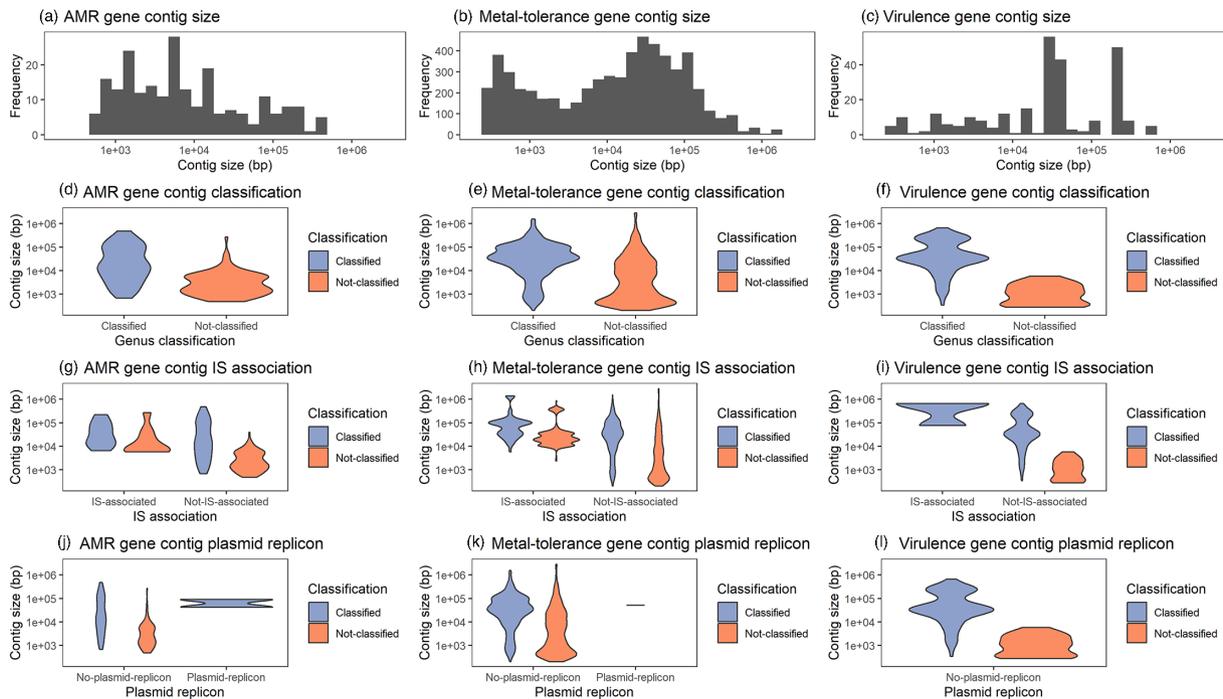


Fig. 6. Size distribution of antimicrobial resistance (AMR) (a, d, g, j), metal-tolerance (b, e, h, k) and virulence (c, f, i, l) gene-containing contigs (a–c) for 24 long-read sequenced food metagenomes, separated by whether they were classified at the genus level (d–f), contained insertion sequences (ISs) (g–i) or contained plasmid replicons (j–l).

ARG analysis identified 368 ARGs across the food metagenomes. Comparisons of ARG concentration within the food metagenome identified three that significantly differed between food commodities (Fig. S7). *blaERP-1* was associated with a higher concentration in leafy greens; *tet(L)* in chicken and pork; and *tet(39)* in chicken, pork and salmon.

Virulence gene analysis identified 922 virulence genes across the food metagenomes. Five virulence genes significantly differed between food commodities: *mrkA*, *mrkB*, *mrkC*, *mrkD* and *mrkF* were associated with a higher concentration in chicken, pork and salmon samples (Fig. S8). These genes are associated with bacterial fimbriae and biofilm formation. The VFDB associates each virulence gene with a specific pathogen; for each pathogen cultured, the culture status was compared with the detection of one or more virulence genes associated with that pathogen. When compared to culture results, virulence gene detection was 0–71% sensitive and 11–94% specific for the pathogens investigated (Table S5).

Metal-tolerance gene analysis identified 413 metal-tolerance genes across the food metagenomes. A total of 125 metal-tolerance genes significantly differed between food commodities, of which 65 were associated with tolerance to multiple compounds (Fig. S9). These genes were associated with tolerance to 61 compounds, most commonly with tolerance to mercury (Fig. S10).

Long-read metagenomes

Long-read sequencing was applied to 24 metagenomes to improve metagenome assembly length and help taxonomically investigate ARGs, along with metal-tolerance and virulence genes. Long-read sequencing was applied to 7 chicken, 11 leafy green, 5 pork and 1 salmon samples. Hybrid long- and short-read assembled metagenomes had a mean N50 value of 17 981 bp (range: 3760–41 102 bp) and a mean largest contig of 1 258 817 bp (range: 365 137–4 020 681 bp).

ARGs comprised 0–0.009% of the metagenome, metal-tolerance genes 0.005–0.3% of the metagenome and virulence genes 0–0.03% of metagenomes (Fig. 4). The taxonomic origin was predicted for 59% of AMR, 68% of metal-tolerance and 96% of virulence genes at the genus level, but greater percentages were predicted at higher taxon levels (Fig. S11). The genera most commonly found with ARGs were *Aeromonas* (18%), *Shewanella* (9.4%), *Acinetobacter* (9.2%) and *Carnobacterium* (6.0%). Amongst the 24 hybrid metagenomes, 66 unique ARGs were identified and 35 were identified in multiple contigs. Of those identified multiple times, eight were consistently found with the same genus and four were found with different genera (Fig. S12). The genus most commonly found with metal-tolerance genes was *Pseudomonas* (41%), and that most commonly found with virulence genes was *Yersinia* (93%).

Amongst the 24 hybrid assembled metagenomes, 173–1303 (mean=680) ISs and 0–47 (mean=12) plasmid replicons were identified. Using a 10 000 bp cut-off, 16% (sample range: 0–49%) of ARGs were associated with ISs, as were 13% (sample range: 0–77%) of metal-tolerance genes and 1.8% (sample range: 0–100%) of virulence genes. ARGs were most commonly associated with IS3 (3.7%; sample range: 0–39%), IS4 (3.9%; sample range: 0–25%), IS5 (2.9%; sample range: 0–36%), IS6/IS26 (4.6%; sample range: 0–15%) and IS30 (3.9%; sample range: 0–17%) (Fig. 5). 6.0% (sample range: 0–52%) of ARGs, 0.07% of metal-tolerance genes (sample range: 0–0.77%) and no virulence genes were found on the same contig as a plasmid replicon.

The ARG-containing contigs were 479–485 710 bp (mean: 37 227 bp) in length, metal-tolerance gene-containing contigs were 202–2 849 978 bp (mean: 65 141 bp) in length and virulence gene-containing contigs were 273–665 242 bp (mean: 85 980 bp) in length (Fig. 6). Above 10 000 bp, the origin of most contigs was predicted at the genus level. There was no association between taxonomic classification at the genus level and being associated with an IS for ARG-containing (Fisher's exact test: $P=0.53$), metal-tolerance-containing (Fisher's exact test: $P=0.18$) or virulence-containing (Fisher's exact test: $P=1.0$) contigs.

To validate the taxonomic classification of metagenomic elements of interest, the analysis was repeated on the cultured isolate genomes where the taxonomy was known. None of the genome contigs containing AMR, metal-tolerance or virulence genes were misclassified, but only 17% of ARGs, 11% of metal-tolerance genes and 12% of virulence genes were classified at the genus level (Figs S13–S15). However, the percentage of genes classified at the genus level differed between the pathogens cultured for the genes of interest investigated (Fisher's exact test: $P<1\times 10^{-8}$, 1×10^{-8} and 1×10^{-8} , respectively), and larger percentages of genes were classified at higher taxon levels.

DISCUSSION

Metagenomics has the potential to revolutionize food safety diagnostics, allowing for the identification and typing of pathogens and tracking AMR in a matter of hours or at most days [23]. However, by comparing the genomes of pathogens cultured from food samples to the metagenomic detection of the same pathogens, we demonstrate that it is currently not practical in all scenarios. Its ability to detect the presence of pathogens or genetic traits, which may be in low abundance, must be evaluated before it should be applied to food surveillance.

A wide range of micro-organisms were found on food that varied in abundance, but the predominating micro-organisms belonged to the *Acinetobacter*, *Carnobacterium*, *Pseudomonas*, *Psychrobacter* and *Shewanella* genera. These bacteria are mostly psychrotrophic, able to replicate on food when refrigerated. We also investigated the ability of metagenomics to detect the

presence of particular bacteria relevant to food safety (*E. coli*, *Klebsiella/Raoultella*, *Salmonella* and *Vibrio*), whose presence was confirmed through a culture-based approach. A pork and a chicken sample had a large percentage of *E. coli* reads in their metagenome, and from the *E. coli* cultured from these samples, we were able to identify >99% of the genomes of the *E. coli* cultured from the pork sample and >99% of the genome of one of the *E. coli* isolate genomes cultured from the chicken sample. The chicken sample had a diverse population of *E. coli* present, as indicated by the multiple *E. coli* STs we cultured from the sample, and our ST analysis of the metagenomic reads indicated that we did not isolate the most prominent *E. coli* ST. Therefore, if there is a diverse population of a bacterial species, then metagenomics will likely identify the most prominent strain, which may be missed if culturing methodologies involve an enrichment step and isolating a small number of colonies. However, for the remaining samples, the cultured pathogens comprised less than 10% of the micro-organism reads on food, and the percentage of their genomes identified in metagenomes varied from 6 to 98%. In addition, investigating low-abundant pathogens in metagenomes by the presence of any reads classified as belonging to pathogens was not a very sensitive approach when compared to culture results, nor was searching metagenomes for virulence genes associated with pathogens. Therefore, our results indicate that efficient identification and metagenomic characterization of pathogens on food require either (a) deeper metagenomic sequencing or (b) specific enrichment of pathogens or their DNA.

The number of microbial reads sequenced in a food metagenome is dependent on the overall sequencing depth and the amount of host DNA removed prior to sequencing. The food metagenomes analysed in this study had undergone host DNA depletion, but the method used worked better for certain food commodities over others [22], leaving variation in the number of microbial reads. The percentage of the pathogen genomes identified in the associated metagenomes was related to the number of microbial reads sequenced. However, the linear regression results indicate that 4.1×10^6 additional microbial reads need to be sequenced in the metagenome to increase the percentage of pathogen assembly in the metagenome by 1%. Therefore, the read depth would need to be magnitudes higher to ensure the sensitive identification of pathogens on food. Previously, shotgun metagenomics was used to detect *Brucella* in host DNA-depleted raw milk samples [61]. However, the raw food commodities in our study likely have higher concentrations of non-pathogenic micro-organisms than raw milk, explaining its inability to detect culturable pathogens in many samples. Leonard *et al.* [62] were able to detect Shiga toxin-producing *E. coli* on spinach, but only after an 8-h enrichment that significantly altered the microbiota. Therefore, on food samples with a large concentration of non-pathogens present, pathogen enrichment by growing the pathogens or rejecting host DNA through approaches such as Oxford Nanopore adaptive sequencing [63] is required. However, both enrichment approaches will alter the relative abundance of micro-organisms in the microbiota, preventing analyses that rely on abundance.

Virulence genes encode microbial components that are often required for pathogens to cause infections [10]. The short-read analysis identified 922 virulence genes in the food metagenomes, but only 5 significantly differed amongst food commodities, which were associated with chicken, pork and salmon. These virulence genes are associated with *Klebsiella pneumoniae* in the VFDB but have been isolated from other bacteria, such as *Pseudomonas* from food [64]. For 24 metagenomes, we also had long-read sequences supplementing the short-read sequences. Most virulence genes in these samples were associated with *Yersinia*, in concordance with our previous study that demonstrated the presence of *Yersinia enterocolitica* on 76–80% of meat and seafood samples [65], but almost all isolates belonged to the non-pathogenic biotype 1A. These virulence genes could also belong to other *Yersinia* species, such as *Yersinia intermedia* or *Yersinia ruckeri* for which we identified MGSs in some of the food metagenomes, but no human clinical infections have been found associated with these species [66].

Pathogens are not the only risk to humans on food, as food can also carry micro-organisms that are resistant to antimicrobial drugs. Hybrid metagenome analysis predicted that most ARGs which could be associated with specific bacterial hosts originated from potential opportunistic pathogens (e.g. *Acinetobacter* and *Aeromonas*) and environmental organisms (e.g. *Carnobacterium* and *Shewanella*). However, there were a large number of ARGs whose taxonomic origin we could not predict, and when the approach was applied to the genomes of cultured bacteria, there was a lot of variation between the percentage of genes classified and the pathogen cultured, suggesting that this approach may be better at predicting the origin of ARGs belonging to some taxa compared to others. ARG databases are also biased towards culturable and/or pathogenic bacteria [67], and these databases need to be extended to be relevant to other bacterial species. The taxonomic predictions also do not take into consideration if ARGs are transferable to other bacteria, which is relevant for the assessment of potential risk to human health.

One way to evaluate the potential mobility of ARGs is to assess their proximity to ISs. These genetic elements facilitate the movement of genes around and between genomes, including ARGs [68]. In the 24 metagenomes which had long-read sequences in addition to short-read sequences, 16% of ARGs were associated with ISs and the most common IS types were IS3, IS4, IS5, IS6/IS26 and IS30. Most of these IS types have been found to be highly or variably associated with ARGs in bacterial pathogens and in animal and environmental metagenomes [69]. In addition, 6% of ARGs were found on plasmid replicon-containing contigs. The small number of ARGs associated with ISs and plasmids could suggest that only a small percentage of the ARGs on food are transmissible to other bacteria but is more likely due to the fragmented metagenome assemblies. Even with long-read sequencing, metagenomic analysis could not assemble all chromosomes and plasmids

into single contigs, leaving many of the ARGs on small contigs, and neither their bacterial host nor genomic location (chromosome, plasmid or other mobile genetic element) could be determined. However, even if all plasmids were assembled into single contigs, we may still struggle to identify the origin of the plasmid if it is associated with a wide range of bacteria. High-throughput chromosome conformation capture techniques crosslink sections of DNA that interact with each other before sequencing, allowing the identification of chromosomes and plasmids from the same cell [70], and may be required for plasmid taxonomic classification on food. When looking at specific ARGs, we identified some that were unlikely to be mobile, such as the *blaA* and *vatF* genes, which were predicted to be from *Yersinia* and are found in the chromosome of this bacterial genus [71, 72]. However, four ARGs were associated with multiple bacterial genera and therefore likely to be mobile. Overall, this suggests that at least 32% (range: 0–100% amongst food samples) of ARGs in the food metagenome are mobile, as they are either associated with an IS, plasmid replicon or multiple genera. However, there was also a sample bias in this analysis, as long-read sequencing requires a higher DNA concentration than short-read sequencing, preventing long-read sequencing of samples with low DNA concentrations, as was observed for most seafood samples examined in this study [22].

Non-pathogenic bacteria on food are not just potential AMR carriers; they may also be beneficial to human health [73]. There is a large amount of variation in the nutrients and macromolecules found in food [74], which likely leads to the selection of micro-organisms that contain metabolic modules that allow them to utilize these nutrients and macromolecules. Most of the foods analyzed here are likely to be cooked prior to consumption, limiting the exposure of consumers to the micro-organisms identified. However, for leafy greens, there is likely to be little processing, and diets rich in leafy greens are likely to expose consumers to micro-organisms associated with specific carbohydrate degradation modules, such as pectin degradation. Pectins are polysaccharides found in many fruits and vegetables that are indigestible by human enzymes [75] but can be fermented in the gastrointestinal tract into products such as short-chain fatty acid (SCFA) [76], mostly in the large intestine [77]. Along with being a type of dietary fibre, pectins promote gastrointestinal health by promoting pectinolytic micro-organisms that can ferment pectin [78] and the immune-modulatory role of the SCFA produced [79]. Leafy greens may therefore be beneficial to the human diet by providing consumers with pectin and the micro-organisms to utilize them. Further work is required to identify the micro-organisms responsible for these modules on food and if they can colonize the gastrointestinal tract.

Outside of health risks and benefits, micro-organisms on food can also cause spoilage [80]. *Pseudomonas* is one of the main microbial causes of food spoilage [81] and is the predominant bacterial genus found on retail food [22]. Previously, we tested 32 food samples for *Pseudomonas*, including 8 that had their metagenomes sequenced and were included as part of this study, and identified a diverse population of this genus on food [64]. Metal-tolerance genes were most commonly associated with *Pseudomonas*, but this is likely because of how prevalent *Pseudomonas* is on food. In this study, we identified *Pseudomonas* species significantly differed between food commodities. Future work is required to determine how food spoilage alters the populations of different *Pseudomonas* species on food.

Consumer health relies on safe food. The microbial communities on retail food are predominated by psychrotrophic bacteria that benefit from the cool temperatures at which food is stored. Therefore, consistent identification of pathogens that are not adapted to thrive in cold conditions and thus comprise a small percentage of the overall microbial community on food will require methods that can enrich relevant bacteria and/or selectively sequence their genetic elements. The microbial composition of food varies amongst different commodities, and food metagenomics can predict the taxonomic origin of many ARGs on food at the genus level, but further work is required to determine what percentage of the ARGs on food is associated with mobile genetic elements and thus transferable to other bacteria. Food metagenomics could then be used to investigate what AMR bacteria consumers are exposed to on food, but currently, it is unable to consistently identify pathogens without enrichment.

Funding information

This project was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent project BBS/E/F/000PR10348 (Theme 1, Epidemiology and Evolution of Pathogens in the Food Chain), BBSRC Institute Strategic Programme Microbes and Food Safety BB/X011011/1 and its constituent project BBS/E/F/000PR13634 (Theme 1, Microbial threats from foods in established and evolving food systems), BBSRC Institute Strategic Programme Food Microbiome and Health BB/X011054/1 and its constituent project BBS/E/F/000PR13631, Earlham Institute ISP Decoding Biodiversity BX011089/1, projects BBS/E/ER/230002A, European Research Council H2020 StG (erc-stg-948219, EPYC) and Food Standards Agency (FSA) project FS101185 through an FSA Fellowship to A.E.M. This work was also supported in part by BBSRC grant BB/V01823X/1. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Acknowledgements

We thank the Quadram Institute Bioscience core facility sequencing and bioinformatic teams, laboratory technicians, media preparation and laboratory managers for their assistance in the project and the CLIMB-BIG-DATA-computing servers, an infrastructure supported by the UKRI Medical Research Council grant MR/T030062/1.

Author contributions

S.J.B. and A.E.M. conceived and designed the study. S.J.B. and R.P. collected the data. S.J.B., F.H. and A.L.Z. performed the formal analysis. S.J.B. and A.E.M. wrote the original draft. All reviewed and edited the manuscript.

Conflicts of interest

International Patent Application No. PCT/GB2023/050906 entitled 'Determination and quantification of the microbial communities and antimicrobial resistance genes on food' in the name of Quadram Institute Bioscience has been filed (priority date 05/04/2022) and is currently in the international phase. The authors declare no other competing interests.

References

- Janecko N, Zamudio R, Palau R, Bloomfield SJ, Mather AE. Repeated cross-sectional study identifies differing risk factors associated with microbial contamination in common food products in the United Kingdom. *Food Microbiol* 2023;111:104196.
- Sun C, Wang Y, Ma S, Zhang S, Liu D, et al. Surveillance of antimicrobial resistance in *Escherichia coli* and enterococci from food products at retail in Beijing, China. *Food Control* 2021;119:107483.
- Shang Y, Ye Q, Cai S, Wu Q, Pang R, et al. Loop-mediated isothermal amplification (LAMP) for rapid detection of *Salmonella* in foods based on new molecular targets. *LWT* 2021;142:110999.
- Liao H, Ji Y, Sun Y. High-resolution strain-level microbiome composition analysis from short reads. *Microbiome* 2023;11:183.
- Holland D, Mahmoudzadeh N. Foodborne disease estimates for the United Kingdom in 2018. In: *Food Standards Agency*. 2020. pp. 1–27.
- Imhoff B, Morse D, Shiferaw B, Hawkins M, Vugia D, et al. Burden of self-reported acute diarrheal illness in FoodNet surveillance areas, 1998–1999. *Clin Infect Dis* 2004;38 Suppl 3:S219–26.
- Voetsch AC, Van Gilder TJ, Angulo FJ, Farley MM, Shallow S, et al. FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. *Clin Infect Dis* 2004;38 Suppl 3:S127–34.
- Vinayaka AC, Ngo TA, Kant K, Engelsmann P, Dave VP, et al. Rapid detection of *Salmonella enterica* in food samples by a novel approach with combination of sample concentration and direct PCR. *Biosens Bioelectron* 2019;129:224–230.
- Smith JL, Fratamico PM. Emerging and re-emerging foodborne pathogens. *Foodborne Pathog Dis* 2018;15:737–757.
- Casadevall A, Pirofski L-A. Virulence factors and their mechanisms of action: the view from a damage-response framework. *J Water Health* 2009;7 Suppl 1:S2–S18.
- Melton-Celsa AR. Shiga toxin (Stx) classification, structure, and function. *Microbiol Spectr* 2014;2:1–13.
- Sánchez-Chica J, Correa MM, Aceves-Diez AE, Castañeda-Sandoval LM. A novel method for direct detection of *Bacillus cereus* toxin genes in powdered dairy products. *Int Dairy J* 2020;103:104625.
- Rolain J-M. Food and human gut as reservoirs of transferable antibiotic resistance encoding genes. *Front Microbiol* 2013;4:1–10.
- Vats P, Kaur UJ, Rishi P. Heavy metal-induced selection and proliferation of antibiotic resistance: a review. *J Appl Microbiol* 2022;132:4058–4076.
- Espinosa CD, Stein HH. Digestibility and metabolism of copper in diets for pigs and influence of dietary copper on growth performance, intestinal health, and overall immune status: a review. *J Anim Sci Biotechnol* 2021;12:13.
- Jesus V, Martins D, Branco T, Valério N, Neves MGPMS, et al. An insight into the photodynamic approach versus copper formulations in the control of *Pseudomonas syringae* pv. *actinidiae* in kiwi plants. *Photochem Photobiol Sci* 2018;17:180–191.
- Bearson BL, Trachsel JM, Shippy DC, Sivasankaran SK, Kerr BJ, et al. The role of *Salmonella* genomic Island 4 in metal tolerance of *Salmonella enterica* serovar I 4,[5],12:l:- Pork outbreak isolate USDA15WA-1. *Genes* 2020;11:1291.
- Syed A, Elgorban AM, Bahkali AH, Eswaramoorthy R, Iqbal RK, et al. Metal-tolerant and siderophore producing *Pseudomonas* fluorescence and *Trichoderma* spp. improved the growth, biochemical features and yield attributes of chickpea by lowering Cd uptake. *Sci Rep* 2023;13:4471.
- Cronin P, Joyce SA, O'Toole PW, O'Connor EM. Dietary fibre modulates the gut microbiota. *Nutrients* 2021;13:1–22.
- Brial F, Chilloux J, Nielsen T, Vieira-Silva S, Falony G, et al. Human and preclinical studies of the host-gut microbiome co-metabolite hippurate as a marker and mediator of metabolic health. *Gut* 2021;70:2105–2114.
- Muto A, Kotera M, Tokimatsu T, Nakagawa Z, Goto S, et al. Modular architecture of metabolic pathways revealed by conserved sequences of reactions. *J Chem Inf Model* 2013;53:613–622.
- Bloomfield SJ, Zomer AL, O'Grady J, Kay GL, Wain J, et al. Determination and quantification of microbial communities and antimicrobial resistance on food through host DNA-depleted metagenomics. *Food Microbiol* 2023;110:104162.
- Billington C, Kingsbury JM, Rivas L. Metagenomics approaches for improving food safety: a review. *J Food Prot* 2022;85:448–464.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–1055.
- Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 2009;25:1754–1760.
- Bushnell B. BBMap: a fast, accurate, splice-aware aligner. In: *9th Annual Genomics of Energy & Environment Meeting*. Walnut Creek, 2014. p. 1.
- Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 2015;12:902–903.
- R core team. R: a language and environment for statistical computing. In: *R Foundation for Statistical Computing*. 2019. p. 1.
- Zolfo M, Tett A, Jousson O, Donati C, Segata N. MetaMLST: multi-locus strain-level bacterial typing from metagenomic samples. *Nucleic Acids Res* 2017;45:1–10.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
- Hildebrand F, Gossmann TI, Frioux C, Özkurt E, Myers PN, et al. Dispersal strategies shape persistence and evolution of human gut bacteria. *Cell Host Microbe* 2021;29:1167–1176.
- Özkurt E, Fritscher J, Soranzo N, Ng DYK, Davey RP, et al. LotuS2: an ultrafast and highly accurate tool for amplicon sequencing analysis. *Microbiome* 2022;10:176.
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257.
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;31:1674–1676.

38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
39. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;11:1–11.
40. Steingger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017;35:1026–1028.
41. Saary P, Forslund K, Bork P, Hildebrand F. RTK: efficient rarefaction analysis of large datasets. *Bioinformatics* 2017;33:2594–2595.
42. Pan S, Zhao X-M, Coelho LP. SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. *Bioinformatics* 2023;39:i21–i29.
43. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023;20:1203–1212.
44. McCallum A, Nigam K, Ungar LH. Efficient clustering of high-dimensional data sets with application to reference matching. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2000. pp. 169–178.
45. Frioux C, Ansorge R, Özkurt E, Ghassemi Nedjad C, Fritscher J, et al. Enterosignatures define common bacterial guilds in the human gut microbiome. *Cell Host Microbe* 2023;31:1111–1125.
46. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;47:D309–D314.
47. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 2021;18:366–368.
48. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27–30.
49. Darzi Y, Falony G, Vieira-Silva S, Raes J. Towards biome-specific analysis of meta-omics data. *ISME J* 2016;10:1025–1028.
50. Vieira-Silva S, Falony G, Darzi Y, Lima-Mendez G, Garcia Yunta R, et al. Species-function relationships shape ecological properties of the human gut microbiome. *Nat Microbiol* 2016;1:16088.
51. Clausen P, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 2018;19:307.
52. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67:2640–2644.
53. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. *Nucleic Acids Res* 2016;44:D694–7.
54. Bloomfield SJ, Zomer AL, Mather AE. Lapidary: identifying and reporting amino acid sequences in metagenomes using sequence reads and Diamond. *bioRxiv* 2024;1. DOI: 10.1101/2024.03.25.586564.
55. Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DGJ. BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res* 2014;42:D737–43.
56. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat Biotechnol* 2019;37:937–944.
57. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58:3895–3903.
58. Gertz EM, Yu Y-K, Agarwala R, Schäffer AA, Altschul SF. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol* 2006;4:1–14.
59. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 2017;33:3340–3347.
60. Lal Gupta C, Kumar Tiwari R, Cytryn E. Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes. *Environ Int* 2020;138:105667.
61. Grütze J, Gwida M, Deneke C, Brendebach H, Projahn M, et al. Direct identification and molecular characterization of zoonotic hazards in raw milk by metagenomics using *Brucella* as a model pathogen. *Microbial Genomics* 2021;7:1–15.
62. Leonard SR, Mammel MK, Lacher DW, Elkins CA. Application of metagenomic sequencing to food safety: detection of shiga toxin-producing *Escherichia coli* on fresh bagged spinach. *Appl Environ Microbiol* 2015;81:8183–8191.
63. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, et al. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol* 2022;23:1–27.
64. Bloomfield SJ, Palau R, Holden ER, Webber MA, Mather AE. Genomic characterization of *Pseudomonas* spp. on food: implications for spoilage, antimicrobial resistance and human infection. *BMC Microbiol* 2024;24:1–12.
65. Palau R, Bloomfield SJ, Jenkins C, Greig DR, Jorgensen F, et al. *Yersinia enterocolitica* biovar 1A: an underappreciated potential pathogen in the food chain. *Int J Food Microbiol* 2024;412:110554.
66. Punsalang AJ, Edinger R, Nolte FS. Identification and characterization of *Yersinia intermedia* isolated from human feces. *J Clin Microbiol* 1987;25:859–862.
67. Gschwind R, Ugarcina Perovic S, Weiss M, Petitjean M, Lao J, et al. ResFinderFG v2.0: a database of antibiotic resistance genes obtained by functional metagenomics. *Nucleic Acids Res* 2023;51:W493–W500.
68. Bloomfield S, Duong VT, Tuyen HT, Campbell JI, Thomson NR, et al. Mobility of antimicrobial resistance across serovars and disease presentations in non-typhoidal *Salmonella* from animals and humans in Vietnam. *Microbial Genomics* 2022;8:1–13.
69. Razavi M, Kristiansson E, Flach C-F, Larsson DGJ. The association between insertion sequences and antibiotic resistance genes. *mSphere* 2020;5:1–12.
70. Stalder T, Press MO, Sullivan S, Liachko I, Top EM. Linking the resistome and plasmidome to the microbiome. *ISME J* 2019;13:2437–2446.
71. Stock I, Wiedemann B. Natural antimicrobial susceptibilities and biochemical profiles of *Yersinia enterocolitica*-like strains: *Y. frederiksenii*, *Y. intermedia*, *Y. kristensenii* and *Y. rohdei*. *FEMS Immunol Med Microbiol* 2003;38:139–152.
72. Seoane A, García Lobo JM. Identification of a streptogramin A acetyltransferase gene in the chromosome of *Yersinia enterocolitica*. *Antimicrob Agents Chemother* 2000;44:905–909.
73. Ejtahed HS, Mohtadi-Nia J, Homayouni-Rad A, Niafar M, Asghari-Jafarabadi M, et al. Effect of probiotic yogurt containing *Lactobacillus acidophilus* and *Bifidobacterium lactis* on lipid profile in individuals with type 2 diabetes mellitus. *J Dairy Sci* 2011;94:3288–3294.
74. Balakrishna Y, Manda S, Mwambi H, van Graan A. Identifying nutrient patterns in South African foods to support national nutrition guidelines and policies. *Nutrients* 2021;13:1–17.
75. Bailoni L, Schiavon S, Pagnin G, Tagliapietra F, Bonsembiante M. Quantitative evaluation of pectins in the dietary fibre of 24 foods. *Ital J Anim Sci* 2005;4:49–58.
76. Elshahed MS, Miron A, Aprotosoae AC, Farag MA. Pectin in diet: interactions with the human microbiome, role in gut homeostasis, and nutrient-drug interactions. *Carbohydr Polym* 2021;255:117388.
77. Holloway WD, Tasman-Jones C, Maher K. Pectin digestion in humans. *Am J Clin Nutr* 1983;37:253–255.
78. Chung WSF, Meijerink M, Zeuner B, Holck J, Louis P, et al. Prebiotic potential of pectin and pectic oligosaccharides to promote anti-inflammatory commensal bacteria in the human colon. *FEMS Microbiol Ecol* 2017;93:1–9.

79. Bartolomaeus H, Balogh A, Yakoub M, Homann S, Markó L, *et al.* Short-chain fatty acid propionate protects from hypertensive cardiovascular damage. *Circulation* 2019;139:1407–1421.
80. Papadopoulou OS, Iliopoulos V, Mallouchos A, Panagou EZ, Chorianopoulos N, *et al.* Spoilage potential of *Pseudomonas* (*P. fragi*, *P. putida*) and LAB (*Leuconostoc mesenteroides*, *Lactobacillus sakei*) strains and their volatilome profile during storage of sterile pork meat using GC/MS and data analytics. *Foods* 2020;9:1–17.
81. Kumar H, Franzetti L, Kaushal A, Kumar D. *Pseudomonas fluorescens*: a potential food spoiler and challenges and advances in its detection. *Ann Microbiol* 2019;69:873–883.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org