



Prokaryotic diversity and potentially pathogenic bacteria in vended foods and environmental samples

Susan W. Muriuki¹ , Michael S. Rengan²  and Nancy L. M. Budambula^{1*} 

Abstract

Purpose: Ready-to-eat fast food vending outlets provide a cheap and readily available food. Foodborne diseases have been previously reported in Embu, Kenya, but data on the prokaryotic metagenome in vended foods is scanty. This study aimed to determine the prokaryotic diversity in fruits, vegetable salad, African sausage, chips (potato fries), fried fish, roasted beef (meat), smokies, samosa, soil, and water collected from food vendors and the surrounding environment in Embu Town and Kangaru Market.

Methods: The study used 454 pyrosequencing, Illumina high-throughput sequencing of 16S rRNA gene in the analysis of total community DNA extracted from samples using the phenol-chloroform method. The 16S rRNA gene variable region (V4-V7) of the extracted DNA was amplified and library construction performed. Sequence analysis was done using QIIME2. Hierarchical clustering of samples, diversity indices, rarefaction curves, and Venn diagrams were generated using the R programming language in R software version 3.6.3.

Results: Bacterial operational taxonomic units (OUTs) were distributed in *Proteobacteria* (52.81%), *Firmicutes* (31.16%), and *Lentisphaerae* (0.001%). The OTUs among archaea were *Candidatus Nitrososphaera* (63.56%) and *Nitrososphaera* spp. (8.77%). *Brucella* spp. and *Bacillus cereus* associated with foodborne diseases were detected. Potential pathogens, *Rickettsia* spp. in risk group 2 and *Brucella* spp. in risk group 3, were detected. Uncultured *Candidatus Koribacter* and *Candidatus Solibacter* were also detected in the food samples. There was a significant difference in the microbial community structure among the sample types ($P < 0.1$).

Conclusion: The results demonstrated the presence of some prokaryotes that are associated with food spoilage or foodborne diseases in vended foods and environmental samples. This study also detected uncultured prokaryotes. The presence of potential pathogens calls for stringent hygiene measures in food vending operations.

Keywords: Prokaryotic metagenome, Vended food, Diversity, 16S rRNA, Total DNA, Kenya

Introduction

Foodborne diseases associated with microbial pathogens, their toxins, or metabolites are a serious global public health problem. Approximately 600 million cases of foodborne diseases and 420,000 deaths occur each year (WHO 2020). Therefore, the assessment of food

matrices for contamination by microbial organisms is important. Conventional methods that have been used in detecting foodborne pathogens rely on growing these microorganisms on synthetic media. These methods are time-consuming and sometimes give prejudiced results as most of the uncultured microorganisms are not detected. Thus, analysis by conventional methods does not give a true representation of the microbial community including potential pathogens in the environment (Law et al. 2015). Molecular methods such as polymerase chain reaction (PCR), denaturing gel electrophoresis

* Correspondence: budambula.nancy@embuni.ac.ke;
nbudambula@yahoo.com

¹Department of Biological Sciences, University of Embu, P.O. Box 6-60100, Embu, Kenya

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

(DGGE), quantitative PCR (qPCR), and loop-mediated isothermal amplification (LAMP) are more reliable in detecting both cultured and uncultured microorganisms in food (Mayo et al. 2014).

Direct metagenome sequencing complements rRNA gene-based characterization by providing information on physiological potential of the microorganisms. Investigation of foodborne pathogens, their source, and associated risks can be done using metagenomics which provides an exhaustive sequencing depth that represents the broader microbial community in the environment (Kergourlay et al. 2015). This technology has proved to be useful in the detection of microorganisms in food and has been previously used to detect *Firmicutes* in tomatoes (Ottesen et al. 2019) and extremely halophilic archaea in food-grade salts (Henriet et al. 2014). Metagenomics identified watercress as the main source of STEC O157 in two concurrent foodborne disease outbreaks (Jenkins et al. 2015). Studies based on next-generation sequencing detected pathogens in Kimchi (a type of fermented food) and in Chinese cabbage (Jung et al. 2011; Kim et al. 2018). Thus, sequencing is important in detecting and characterizing foodborne pathogens (Djenane et al. 2014).

Foodborne disease associated with microbial contamination of food and water in Embu, Kenya, has been previously reported. In 2016, 234 individuals in Embu County who suffered from food poisoning were diagnosed with cholera (George 2016) but there was no application of sequencing technologies in the detection of the causative agent. Whereas metagenomic analysis has been used to investigate prokaryotic diversity in environmental samples such as soil in Kenya (Karanja et al. 2020), application in food safety has not been fully explored. Thus, there could be unculturable potential pathogens that contribute to foodborne diseases in Embu County. Next-generation sequencing (NGS) platform permits the identification and the characterization of the microbial community at a depth of up to millions of sequences per sample, thus allowing the detection and identification of microorganisms that are present in low numbers (Kumaraswamy et al. 2014; Harb and Hong 2017). Metagenomic analysis which involves direct isolation of nucleic acids from samples could generate important information that can be useful in predicting the dynamics of microbial food contamination. Therefore, this study used Illumina high-throughput sequencing to investigate prokaryotic diversity in vended foods, water, and soil from Embu Town and the nearby Kangaru Market. The study was able to determine prokaryotic diversity and detect potentially pathogenic bacteria as well as uncultured prokaryotes in vended foods, water, and soil.

Materials and methods

Study site and sampling

Sampling was carried out in Embu County (latitude: 0° 31' 58.80'' N and longitude: 37° 27' 0.00'') in Kenya. The samples were collected from two sites; the Central Business District of Embu Town and the nearby Kangaru Market which is largely inhabited by university students aged 18–27 years. At the time of study in 2018, there were 324 business entities licensed to sell food in the Central Business District of Embu Town and Kangaru Market. A sample size of $n = 176$ food vendors was determined using the formula described by Israel (2009). Samples were purposively collected from 176 vendors approximately every $\frac{1}{2}$ km distance from each other. Samples comprised fruits, vegetable salad, African sausage (animal intestines filled with beef), potato fries, fried fish, roasted beef, samosa, soil, and water. Since the vended foods were not uniformly stocked by the vendors, they were purchased based on availability. Overall, 21 different sample types were collected. For each sample type, 10 samples were collected in sterile disposable bags while water was drawn directly from the taps into sterile bottles. The soil was scooped from the surrounding environment where the vendors sell their food, to determine whether the contamination could be originating from the surrounding soil. Water was sourced from the water taps where food vendors fetch water for food processing. This was to check whether the contamination could be coming from the water they use and not the food. Water samples included treated water from the Embu Water and Sanitation Company (EWASCO) and irrigation water that is not treated. At the time of sampling, Kangaru Market residents used both irrigation and treated water from EWASCO while those in Embu Town used only treated water from EWASCO. Both irrigation and EWASCO water samples were collected from Kangaru Market, while only EWASCO water samples were collected from Embu Town. The samples were transported to the University of Embu microbiology laboratory at 4°C and processed immediately on arrival. Inocula were prepared from food samples by adding sterile distilled water under aseptic conditions.

Total DNA extraction protocol from fast food and environmental samples

Total DNA was extracted from soil, water, and the selected food samples in duplicate as described by Sambrook et al. (1989) with some modifications. The chemicals that were used in the extraction of total DNA from the samples were purchased from the Inqaba Biotech East Africa Ltd. (IBEA). From each soil sample, 0.5 g of soil was weighed, placed into 2 ml Eppendorf tube, and suspended in 1 ml of sterile water. From the prepared food inoculum, 1 ml was dispensed in 2 ml

Eppendorf tubes. The samples were centrifuged at 13,200 rpm for 10 min, and the supernatant was discarded. All samples formed a visible pellet upon centrifugation. Each pellet was re-suspended in 500 µl of solution A containing 100mM Tris-HCl (pH 8.0) and 100 mM EDTA (pH 8.0), mixed by vortexing and then centrifuged for 1 min after which the supernatant was discarded. The pellet was re-suspended in 200 µl of solution A (100mM Tris-HCl (pH 8.0), 100 mM EDTA (pH 8.0), and 5 µl of lysozyme from a 20 mg/ml solution) and incubated at 37°C for 30 min in a water bath. A lysis buffer of 600 µl (400 mM Tris-HCl (pH 8.0), 60 mM EDTA (pH 8.0), 150 mM NaCl, 1% sodium dodecyl sulfate) was added, and the tubes were allowed to stand for 10 min. A proteinase K of 10 µl (20 mg/ml) was added, and the mixtures were then incubated at 65°C for 55 min in a water bath. An equal volume of chloroform isoamyl alcohol was added to the mixtures which were then centrifuged at 13,200 rpm for 5 min at 4 °C. The supernatant was transferred into new eppendorf tubes. Sodium acetate 150 µl (pH 5.2) and an equal volume of the mixture (supernatant + sodium acetate) of isopropyl alcohol was added, mixed by inverting gently, and incubated overnight at -20°C. The tubes were centrifuged at 13,200 rpm for 10 min after which the supernatant was discarded. The resultant DNA pellets were washed in 300 µl of 70% ethanol. The pellets were centrifuged at 13,200 rpm for 1 min and the supernatant discarded. The DNA pellets were air-dried, lyophilized, and stored at -20°C.

Amplicon library preparation

PCR amplification of the 16S r RNA gene V4–V7 variable regions was carried out from the extracted DNA using barcoded primers 515F (GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) by the service provider mrdnalab in USA (www.mrdnalab.com Shallowater, TX, USA) as described by Caporaso et al. (2011). The amplification was done in 30 cycles using HotStarTaq Plus Master Mix Kit (Qiagen, USA) under the following conditions; 94°C for 3 min of initial heating, followed by 30 cycles of 94°C for 30 s, 53°C for 40 s, and 72°C for 1 min, after which a final elongation step at 72°C for 5 min was performed. The quality of the PCR products was assessed on 2% agarose gel to determine the success of amplification and the relative intensity of the bands. Purified PCR products were used to prepare the DNA library by following the Illumina TruSeq DNA library preparation protocol (Xia et al. 2014). Sequencing was performed at Molecular Research DNA (www.mrdnalab.com, Shallowater, TX, USA) on a MiSeq 2x300bp Version 3 following the manufacturer's guidelines.

Sequence analysis, taxonomic classification, and data submission

Sequences obtained from the Illumina sequencing platform were depleted of barcodes and primers using a proprietary pipeline (www.mrdnalab.com, MR DNA, Shallowater, TX) developed at the service provider's laboratory. The quality control of sequences was done by trimming and filtering the sequences based on their quality score, followed by clustering them into operational taxonomic units (OTUs) based on a fixed dissimilarity threshold. Microbiome bioinformatics was performed with QIIME 2 2019.410 (Bolyen et al. 2018). Raw sequence data were demultiplexed, and the quality was filtered using the q2-demux plugin followed by denoising with DADA2 (Callahan et al. 2016) via q2-dada2. All amplicon sequence variants (ASVs) were aligned with mafft via q2-phylogeny (Katoh et al. 2002) and used to construct a phylogeny with fasttree2 (Price et al. 2010). Alpha-diversity metrics, observed OTUs and Faith's Phylogenetic Diversity beta (Faith 1992), diversity metrics (weighted UniFrac) (Lozupone et al. 2007), unweighted UniFrac (Lozupone and Knight 2005), Jaccard distance, Bray-Curtis dissimilarity, and principal coordinate analysis (PCoA) were estimated using q2-diversity after samples were rarefied (subsampling without replacement) to 900 sequences per sample. To assign Amplicon Sequence variants (ASVs), q2-feature-classifier and a QIIME 2 plugin for taxonomy classification of marker-gene sequences were used (Bokulich et al., 2018a, b). The q2-feature-classifier plugin supports use of any of the numerous machine-learning classifiers available in scikit-learn for marker gene taxonomy classification and currently provides two alignment-based taxonomy consensus classifiers based on BLAST and VSEARCH. The sequences were classified using the Greengenes 99% OTUs 16S reference sequences for bacterial and fungal communities. The change in direction and magnitude in the first principal co-ordinate axis (PC1) for each sample was computed using q2-longitudinal (McDonald et al. 2012). The sequences were submitted to the National Center for Biotechnology Information (NCBI), U.S.A Sequence Read Archive and were assigned accession number PRJNA669559. The data can be accessed through this link, <https://www.ncbi.nlm.nih.gov/sra/PRJNA669559>.

Data analyses

Diversity indices (Richness, Shannon, Absolute diversity), rarefaction curves, and Venn diagram that compared the shared OTUs between the vended food and the environmental samples were determined from the resulting OTUs using R packages, namely Vegan, Phyloseq, and Ampvis2 R packages versions 2.5.6, 1.30.0, and 2.6.0, respectively, in R version 3.6.3 (2020-02-29). Hundred interaction of rarefaction was computed for each sample to 20,000 sequences using QIIME2 pipeline

version qiime2-2019.10. Chao1, a non-parametric estimation of OTU richness, was used to compare species richness between the data sets and sample types. Vended food and the environmental samples were compared using the Analysis of Similarity (ANOSIM) test, based on Bray-Curtis distance measurements with 999 permutations. Significance was tested at 95% confidence interval ($p = 0.05$). Hierarchical clustering of the samples based on Bray-Curtis dissimilarity was carried out using the R programming language and the Vegan package. Correlation analysis, based on Pearson's correlation coefficient, between the vended food and the structure of the environmental sample was conducted, and significance was tested using the Mantel test in R programming language (C Team 2011). Data analysis 3D PCoA plots were calculated using unweighted UniFrac and Bray-Curtis, which was used to determine the distance between the biological communities. To support OTU-based analysis, taxonomic groups were derived from the number of reads assigned to each taxon at all ranks from domain to species using the exported OTUs output from QIIME2 pipeline Version 2019.10. Barplots were generated using R-package phyloseq software version 1.30.0. Alpha diversity was calculated using R-package Phyloseq ggplot2 version 1.30.0 and version 3.3.2. Beta diversity was determined using R-package Vegan, version 2.5.7. Heat maps were generated using R-package Phyloseq using version 1.30.0. Venn diagrams were generated using R package version 1.6.20.

Research approval

A license to conduct the research was obtained from the National Commission for Science, Technology & Innovation (NACOSTI) in Kenya. Authorization to collect samples was granted by the Public Health Ministry of Embu County.

Results

Taxonomic assignment of prokaryotic sequences from vended foods and environmental samples

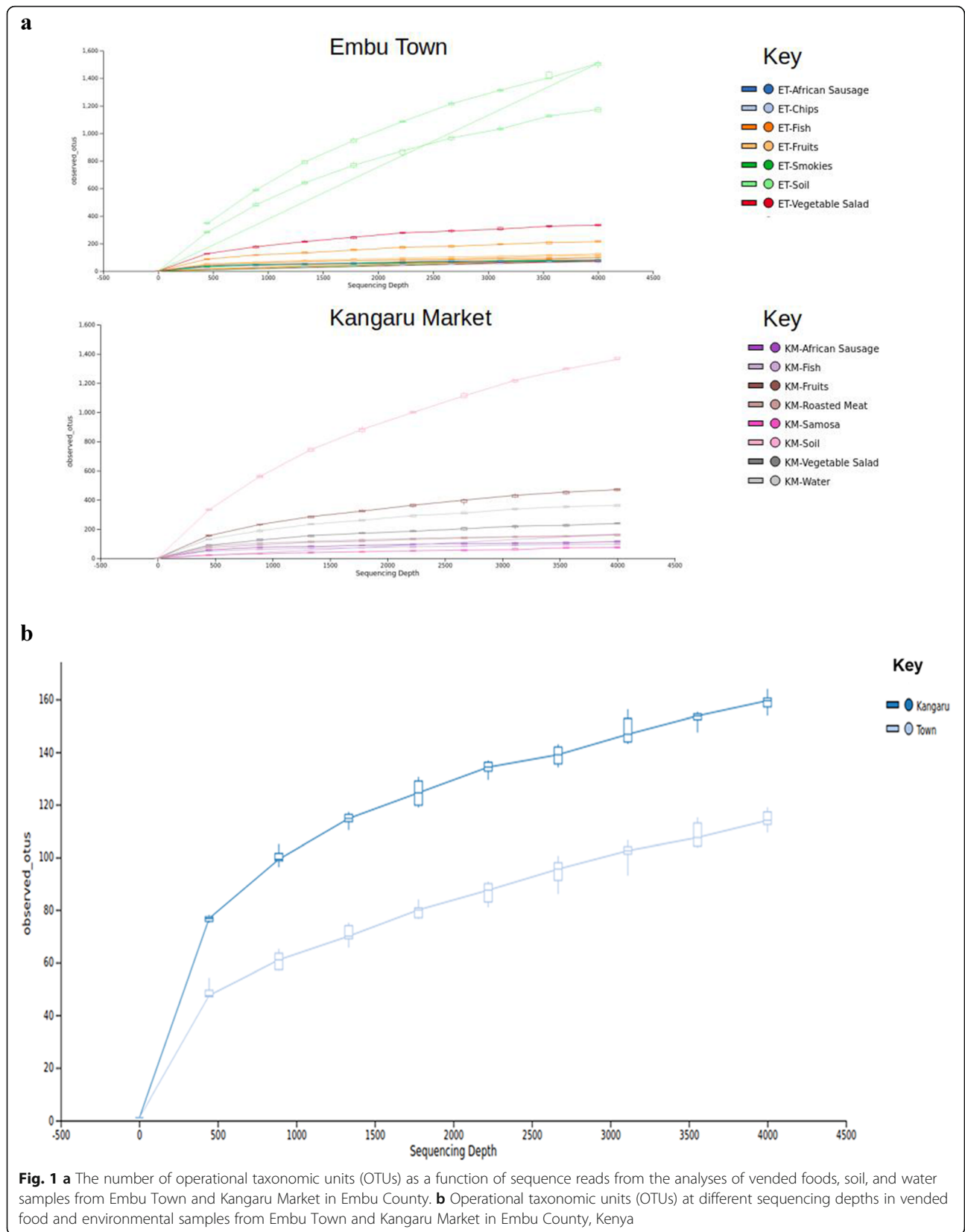
The sequence reads of length >250 bp from Illumina sequencing libraries ranged between 210,536 and 539,981 sequences from both vended food and environmental samples. At the phylum level, the sequence reads contained between 0 and 27,930 OTUs as revealed by the change in the metric of the rarefaction curve. The soil samples had a higher number of observed OTUs compared to food samples while samosa had the least number of observed OTUs (Fig. 1a, b). The samples from Kangaru Market recorded a higher number of OTUs compared to samples from Embu Town.

Bacterial and archaeal taxonomic composition of the vended foods and environmental samples

The OTUs in food and environmental samples were distributed among twenty-three bacterial phyla. The most abundant phyla were *Proteobacteria* (52.81%), *Firmicutes* (31.16%), and *Bacteroidetes* (8.00%) as shown in Fig. 2a. The most predominant genus in soil was *Yersinia*. Among the food samples, the fish had the highest diversity of microorganisms followed by smokies (Fig. 2b). *Clostridiaceae* was most predominant family in fish samples followed by the *Pseudomonaceae*. *Leuconostocaceae* was most predominant in fruits. *Aquaspirillaceae* and the *Enterobacteriaceae* were the most predominant families in roasted beef (Fig. 2c, d). The OTUs recorded in this study included archaea with the most abundant species being *Candidatus Nitrososphaera* (63.56%) and *Nitrososphaera* sp. (8.77%) as shown in Fig. 3a. At the order level, the most abundant archaea were Nitrososphaerales (75.91%) (Fig. 3b). Fruits from Embu Town and Fruits from Kangaru Market recorded different bacterial composition. Vegetable salad from Embu Town and fruits from Kangaru Market had similar archaeal composition. Principle Coordinate Analysis of Jaccard distance tested the differences between the microbial communities from the different samples with statistical significance ($P=0.474$, PERMANOVA, 999 permutations in each test). This indicated that vended food and environmental samples were not distinguishable (Fig. 4). Chips and smokies samples from Embu Town had similar bacterial diversity (Fig. 5a). The genus *Stenotrophomonas* was predominantly detected in smokies from Embu Town while *Citrobacter* was predominantly detected in African sausages and fruits from Embu Town (Fig. 5b). Soil samples from both Kangaru Market and Embu Town had similar archaeal diversity. The most abundant archaeal group was *Candidatus* (74.86%), and it was predominantly recovered from Embu Town soil (Fig. 5c).

Comparison of prokaryotic diversity in vended foods and environmental food samples

Prokaryotic diversity was examined within individual samples and between different samples. From the results output, Chao1 indices showed species richness of data sets while Shannon and inverted Simpson indices reflected the diversity of OTUs in samples. The soil had the highest number of observed OTUs while fish had the highest amount of prokaryotic diversity compared to the rest of the food samples (Fig. 6). Different samples shared common OTUs as illustrated by Venn diagrams. There was an overlap of OTUs between vegetable salad from both Kangaru Market and Embu Town. An overlap of OTUs was also observed in chips from both sites. African sausages from the two sampling sites also exhibited overlapping OTUs (Fig. 7).



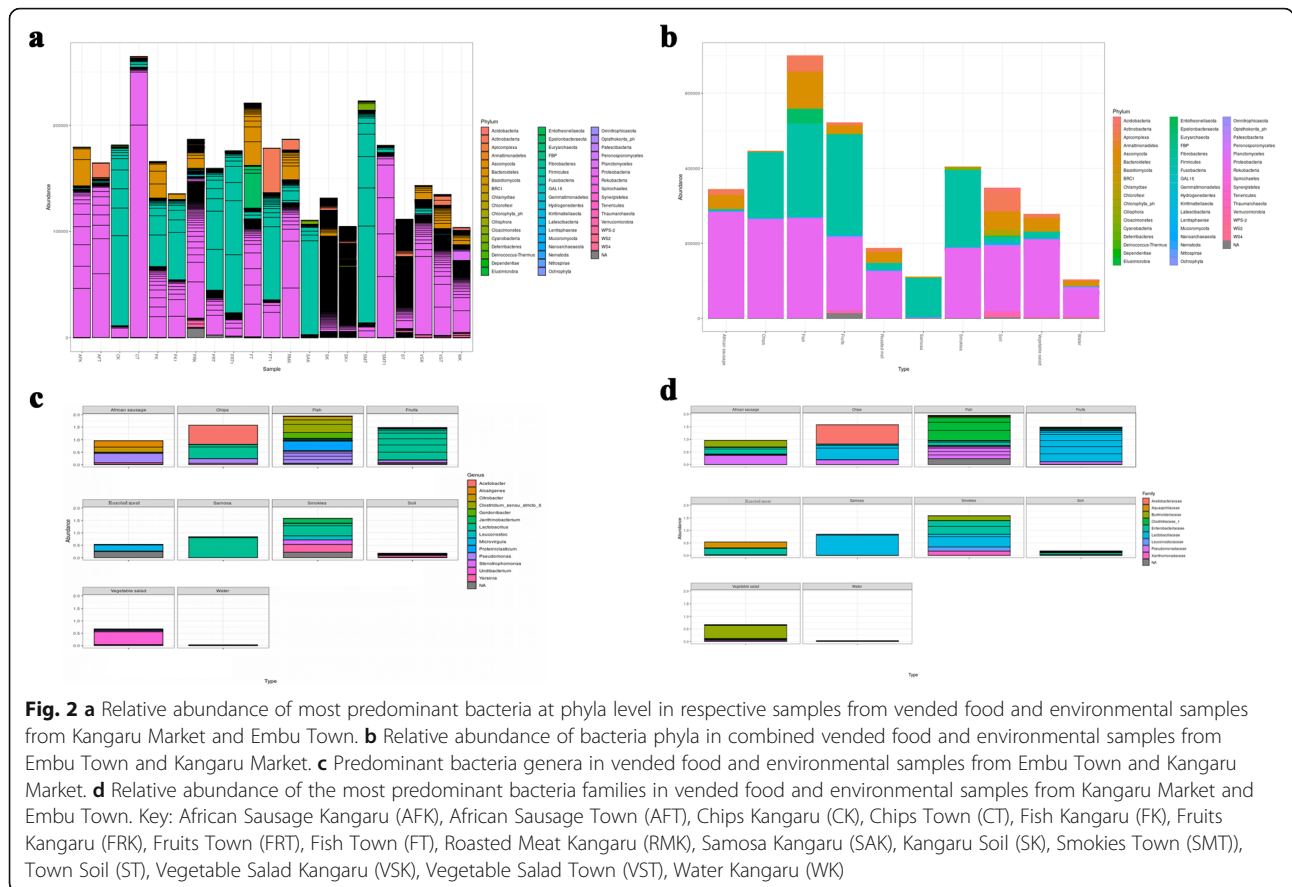


Fig. 2 **a** Relative abundance of most predominant bacteria at phyla level in respective samples from vended food and environmental samples from Kangaru Market and Embu Town. **b** Relative abundance of bacteria phyla in combined vended food and environmental samples from Embu Town and Kangaru Market. **c** Predominant bacteria genera in vended food and environmental samples from Embu Town and Kangaru Market. **d** Relative abundance of the most predominant bacteria families in vended food and environmental samples from Kangaru Market and Embu Town. Key: African Sausage Kangaru (AFK), African Sausage Town (AFT), Chips Kangaru (CK), Chips Town (CT), Fish Kangaru (FK), Fruits Kangaru (FRK), Fruits Town (FRT), Fish Town (FT), Roasted Meat Kangaru (RMK), Samosa Kangaru (SAK), Kangaru Soil (SK), Smokies Town (SMT), Town Soil (ST), Vegetable Salad Kangaru (VSK), Vegetable Salad Town (VST), Water Kangaru (WK)

Potentially pathogenic bacteria from vended foods and environmental samples

Illumina sequencing detected potentially pathogenic bacterial 16S rRNA sequences belonging to *Proteobacteria* and *Firmicutes* (Table 1). The genera *Legionella*, *Aeromonas*, *Staphylococcus*, and *Erysipelothrix* occurred in almost all the samples. A relatively high abundance of the potentially pathogenic bacteria was recorded in the fruit sample from the Kangaru Market at 4.8%. Embu Town samples had a lower relative abundance of sequences belonging to potentially pathogenic bacteria with the soil having 1.43% abundance. Overall, 28 species sequences belonging to 16 genera considered to be potentially pathogenic risk group 2 organisms were detected. *Legionella* spp. were most abundant in water from Kangaru Market. Sequences for *Brucella* spp. and *Coxiella* spp. which are risk group 3 organisms were also detected. The *Brucella* spp. sequences were detected in soil from Embu Town, whereas *Coxiella* spp. were detected in fruit, water, and soil from Kangaru Market (Table 2).

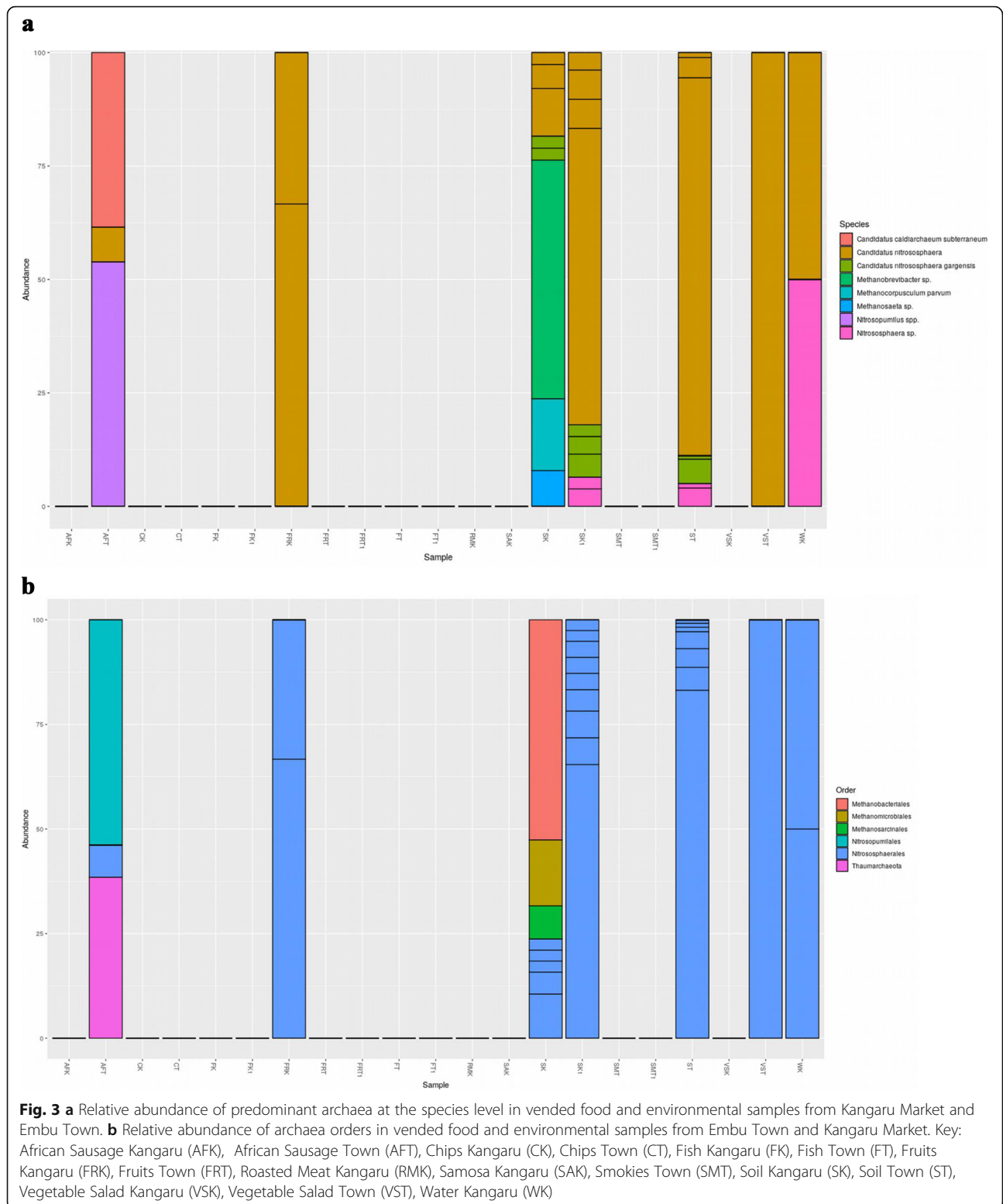
Uncultured prokaryotes in vended foods and environment samples

In species rank, uncultured *Candidatus Solibacter*, *Candidatus Microthrix*, *Candidatus Protochlamydia*, and

Candidatus Kuenenia were detected in vended foods and environmental samples (Table 3).

Discussion and conclusion

Vended foods and environmental samples investigated in the present study were diverse in their microbial community composition. The microbial communities varied depending on the sample type and the site where the samples were collected from. Bacteria formed the most (over 99%) abundant taxa in all the vended foods and environmental samples. This study determined that the most abundant bacteria phyla were *Proteobacteria* (52.81%), *Firmicutes* (31.16%), and *Bacteroidetes* (8.00%). Previous studies have reported the abundance of these phyla in soil and food (Gangwar et al. 2009). *Proteobacteria* pathogens have been detected in drinking water and were associated with biofilm formation in water pipes and leakages (Richards et al. 2018). *Firmicutes* have been reported to adapt to solid food which increases their activity; thus, they can thrive in many solid foods (Hugenholtz et al. 2017). This was evident in the present study, where they thrived in potato fries. Sequences affiliated to archaeal phyla; *Euryarchaeota* (3.36%) and *Thaumarchaeota* (96.64%) were detected in this study. The archaea orders *Methanoarcinales*, *Nitrosopumilales*,



Nitrososphaerales, *Methanomicrobiales*, and *Methanobacteriales* were recovered from vended food and soil samples. The order *Nitrososphaerales* is known to inhabit terrestrial ecosystems, and these prokaryotes must

have been introduced in the soil following surface water runoff and found their way into food following its mis-handling (Kerou and Schleper, 2015a, b; Cavicchioli et al. 2003). *Methanomicrobiales* have been shown to

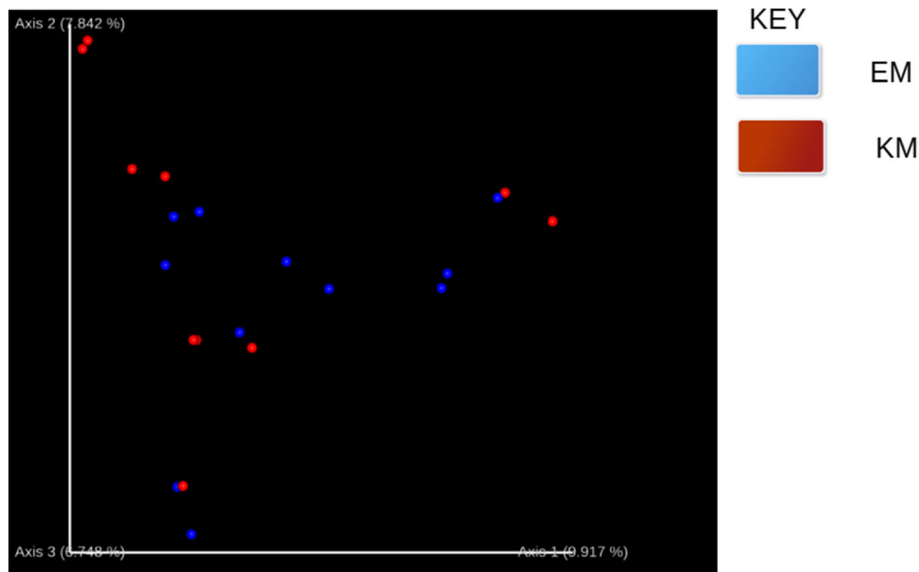


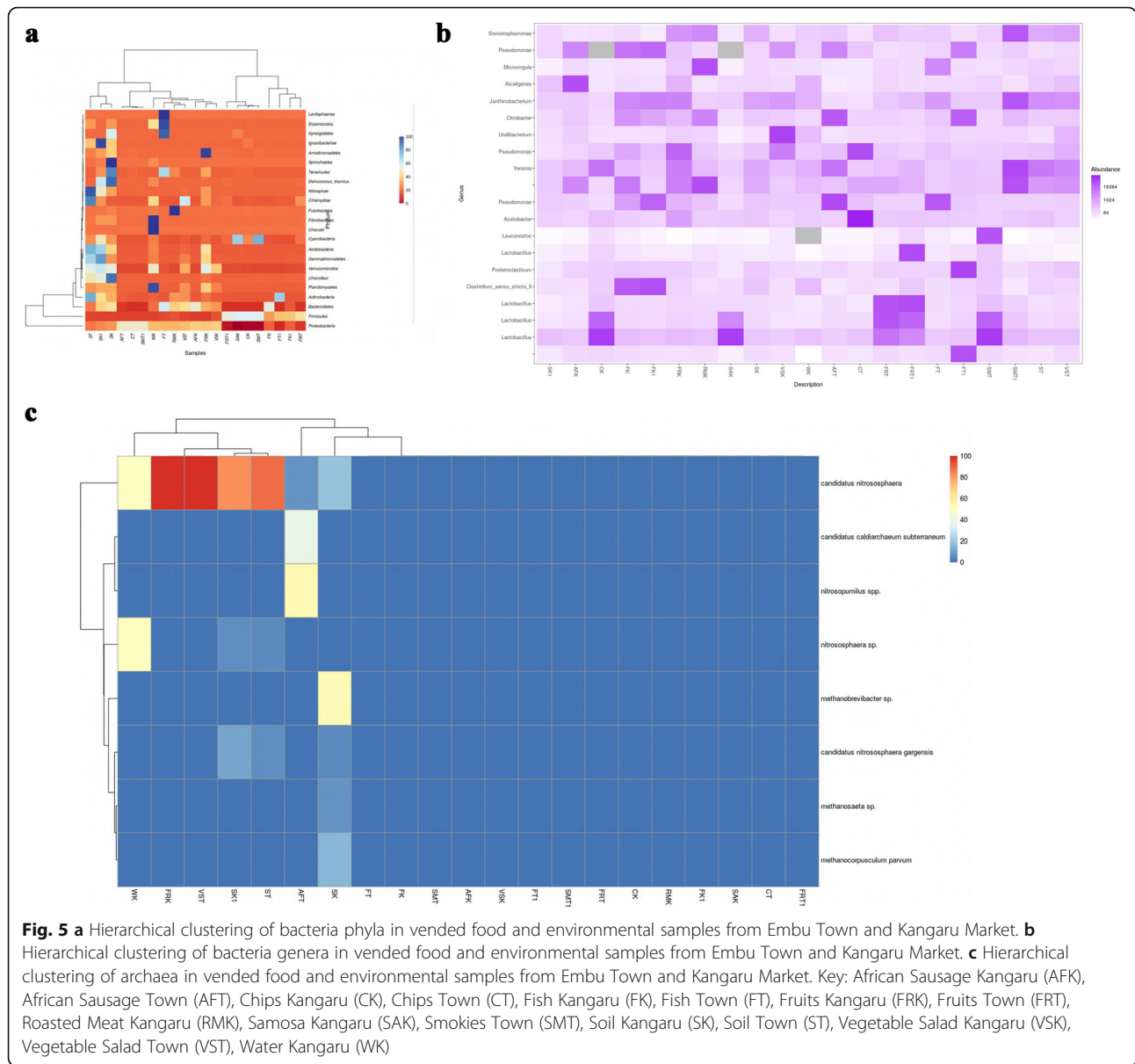
Fig. 4 Principal coordinate analysis of prokaryotic metagenome in vended food and environmental samples from Embu Town and Kangaru Market

inhabit a broad range of anoxic environments (Browne et al. 2017). *Methanobacteriales* prefer anaerobic environments such as marine and freshwater, animal gastrointestinal tracts, and geothermal habitats. The presence of this order in the soil explains the possibility of the introduction of the prokaryote by surface water runoff from animal waste from the surrounding area (Liu 2010).

Prokaryotic diversity varied between sampling sites and among the food samples. Overall, Kangaru Market recorded a higher number of observed OTUs compared to Embu Town, though some food samples from Embu Town had a high number of observed OTUs like fried fish from Embu Town. More variation of prokaryotic diversity was observed in fruits from Kangaru Market compared to those from Embu Town. The variation in the distribution of these species between the sites and the different samples could be due to environmental factors or possibility of careless handling of food by the vendors especially in Kangaru Market. Fried fish samples had the highest diversity of prokaryotes compared to the rest of the food samples. In addition, fried fish from Embu Town had higher diversity compared to fried fish from Kangaru Market. Bacterial contamination of fried fish in the present study could be due to cross-contamination during handling of raw and fried fish, as some vendors use the same hand to handle both (Sifuna and Onyango 2018). Contamination of fried fish could also arise from incomplete cooking; thus, the fish get contaminated by the food they consume or contaminated by the water they live in. Contaminants may also originate from anthropogenic activities and from surviving the cooking process (Donde et al. 2015).

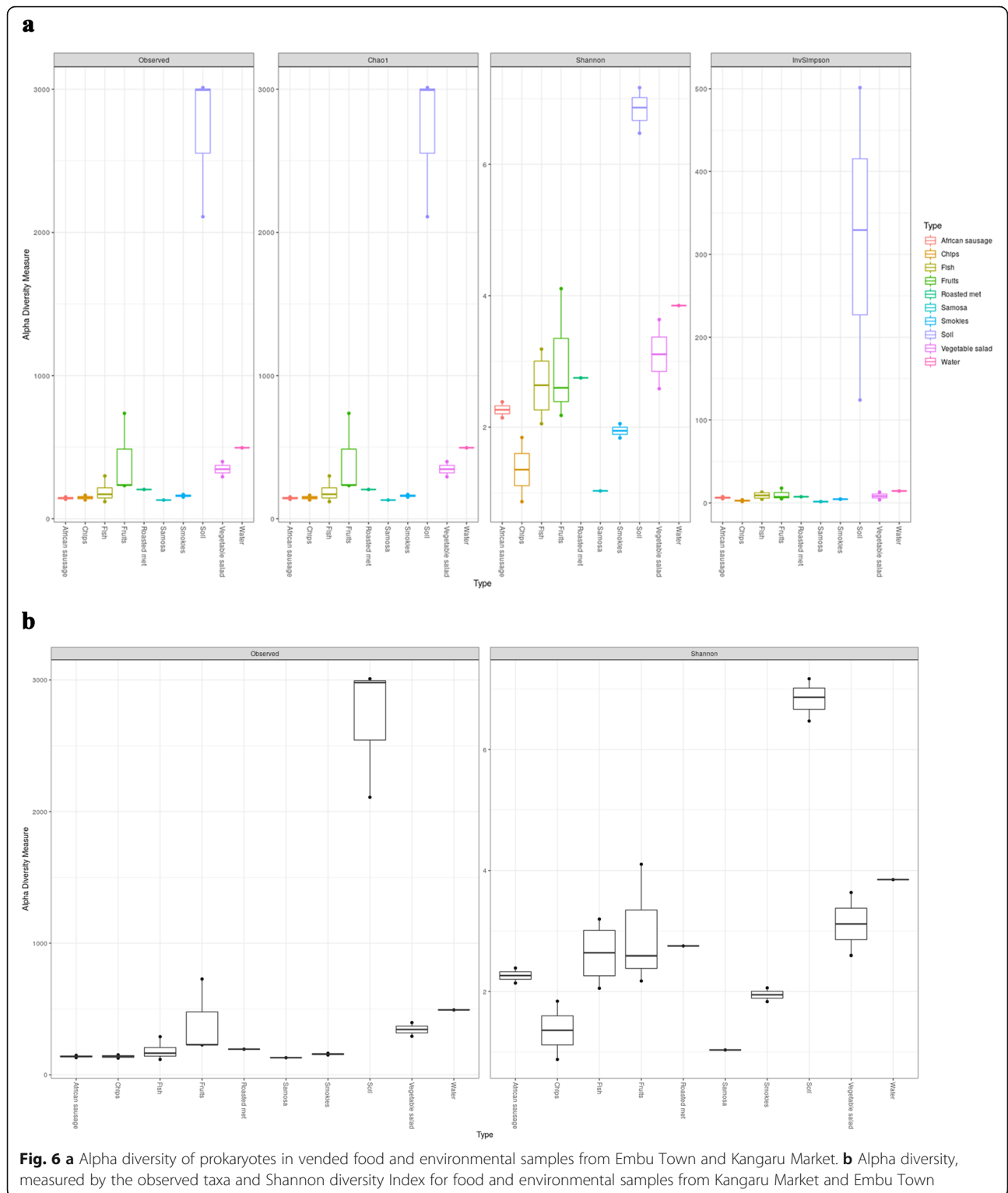
Contamination of street vended smokies which are factory-made could be attributed to the cross-contamination by “kachumbari” (a complementary vegetable salad) or unhygienic handling (Kariuki et al. 2017). The presence of potentially pathogenic and non-pathogenic prokaryotes in the African sausage has been attributed to cross-contamination in previous studies (Karoiki et al. 2018).

The most abundant bacteria recovered across all samples were *Alcaligenes feacalis*, *Lactobacillus perolens*, *Pseudomonas* spp., *Citrobacter freundii*, *Clostridium* spp., and *Acetobacter* spp. Many of the detected bacteria are either normal flora, spoilage microorganisms, or potential pathogens and thus might predispose the public to food poisoning. Multidrug resistance has been reported to occur in *Citrobacter freundii*; thus, its presence in food should inform policy regarding proper handling of food to avoid cases of antibiotic resistance arising from consumption of food contaminated with these strains in the future (Liu et al. 2018). Sequences of *Bacillus cereus* were recovered from the soil in this study. It has been previously been detected in food and has been associated with food poisoning following the formation of endospores and toxin production (Griffiths and Schraft 2017). *Hafnia* spp. sequences were detected in all samples. Commensal *Hafnia* spp. have been shown to reduce food intake and fat mass in people with obese conditions (Legrand et al. 2020). The presence of this bacterium in food could be of benefit to man. *Pseudomonas* spp. were detected in all samples, some *Pseudomonas* spp strains, have been associated with food spoilage while others are pathogenic and have been



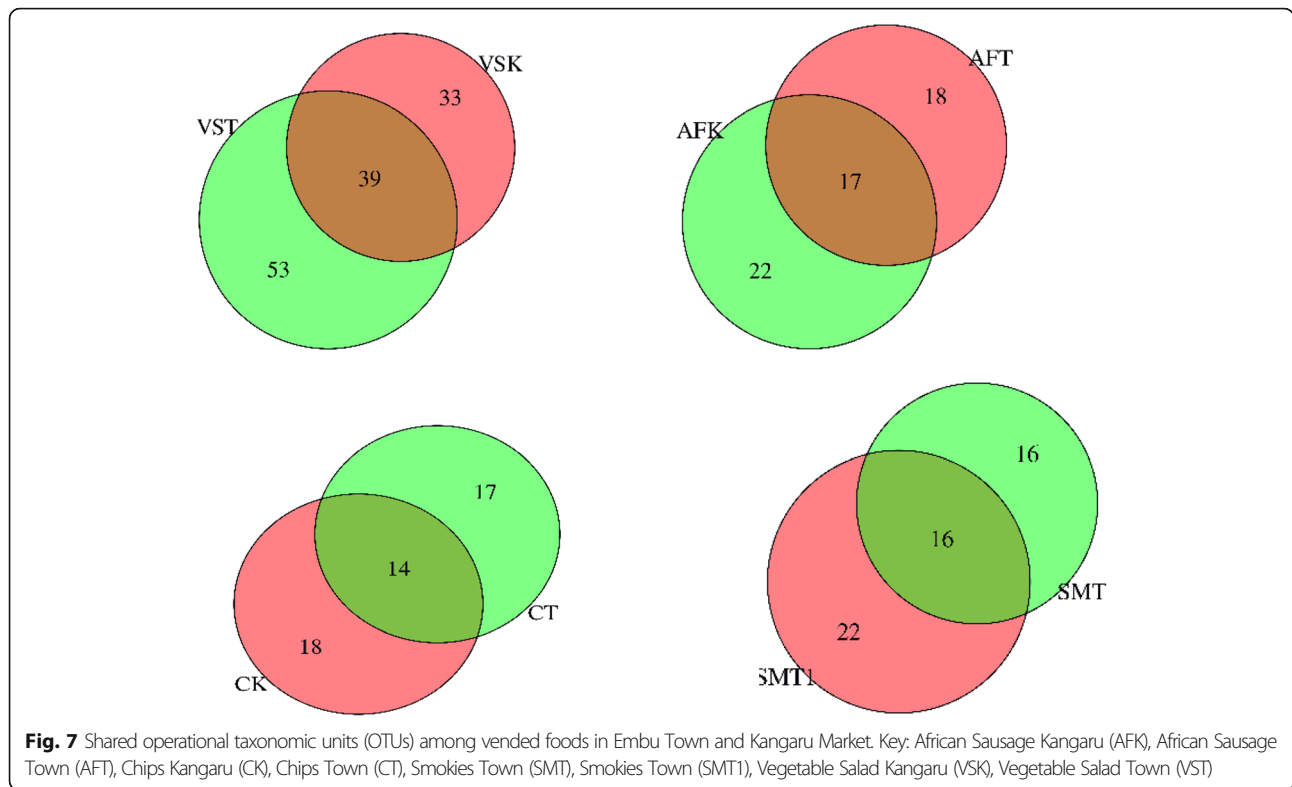
associated with cases of foodborne disease outbreaks (Fakhkhari et al. 2020). The present study detected *Brucella* spp., which belongs to risk group 3 in the soil. Some strains of *Brucella* spp. have been previously associated with food poisoning; thus, the public needs to be informed to observe proper hygiene measures so as to avoid the introduction of the bacterium in food (Garcell et al. 2016). *Clostridium* spp. sequences were recovered from fried fish samples from Kangaru Market. Some *Clostridium* strains are toxigenic and have been associated with cases of foodborne disease outbreaks (Rodriguez et al. 2016). *Coxiella* spp., which belong to risk group 3 were detected in the present study. Some species in this genus have been associated with dermatitis outbreaks in Italy (Raele et al. 2018). *Staphylococcus*

pasteuri, which belongs to risk group 2 was detected in almost all food samples. Its presence in food has been associated with bacteremia in patients diagnosed with acute leukemia, catheter-associated urinary tract infection, and endocarditis (Ramnarain et al. 2019). The presence of *S. pasteuri* in almost all food samples is therefore of concern. *Legionella pneumophila* a clinically important pathogen that was detected in most of the food samples in this study has been associated with severe pneumonia legionnaires disease. It has been previously detected in freshwater where it freely replicates in protozoa (Mendis et al. 2015). In the present study, it was isolated from soil and food samples to which it could have been introduced from the environment.



Most of the prokaryotes in the present study were detected from the soil. This is expected since the soil is known to harbour millions of microorganisms as previously revealed by next-generation sequencing (Vestergaard et al. 2017). In general, fruits from the Kangaru

Market with 260 OTUs and vegetable salads from Embu Town with 200 OTUs had the highest number of observed OTUs among the food samples. Fresh produce such as fruits and vegetables are often exposed to contamination by microbial communities as a result of



contact with irrigation water, manure, or soil. Microorganisms such as *Lactobacillus pentosaceus*, *Weissella cibaria*, and *Lactobacillus plantarum* have been previously detected in vegetables, and some of these bacteria have been associated with food spoilage (Peng et al. 2018).

Prokaryotes that have not yet been cultured were detected in this study and included; *Candidatus Solibacter*, *Candidatus Microthrix*, *Candidatus Koribacter*, *Candidatus Protochlamydia*, and *Candidatus Kuenenia*. Most of these prokaryotes were detected from soil samples; however, *Candidatus Koribacter*, *Candidatus Microthrix*,

Table 1 Sequences of potentially pathogenic bacteria in food and environmental samples from Embu Town and Kangaru Market

Genus	Embu Town	Kangaru Market	Total
<i>Geodermatophilus</i>	253	52	305
<i>Leptospira</i>	0	5	5
<i>Legionella</i>	21	929	950
<i>Rickettsia</i>	6	1357	1363
<i>Mycoplasma</i>	0	4	4
<i>Aeromonas</i>	61	638	699
<i>Coxiella</i>	1	26	27
<i>Brucella</i>	8	0	8
<i>Treponema</i>	0	3	3
<i>Staphylococcus</i>	97	119	216
<i>Acholeplasma</i>	2	56	58
<i>Erysipelothrix</i>	2	134	136
<i>Mycobacterium</i>	14	5	39
Total potential pathogens	465	3336	3801
Total sequences	305774	282682	588456
Potential pathogens percentages	1.98	12	13.98

Table 2 Potentially pathogenic bacteria detected in food and environmental samples from Embu Town and Kangaru Market

Species	Food or environmental sample		Risk group ^a
	Embu Town	Kangaru market	
<i>Geodermatophilus spp.</i>	Chips T, African sausage T, fruits T, soil T	Water K, soil K,	2
<i>Leptospira idonii</i>	-	Water K, soil K	2
<i>Legionella spp.</i> , <i>Legionella pneumophila</i> , <i>Legionella wadsworthii</i>	Chips T, fruits T, vegetable salad T, fish T, soil T	Fruits K, water K, fish K, fish K, soil K	2
<i>Rickettsia sp.</i> , <i>Rickettsia spp.</i> , <i>Rickettsia canadensis</i>	African sausage T, fruits T, vegetable salad T	Fruits K, water K, fish K, roasted beef K, Soil K	2
<i>Mycoplasma alligatoris</i>	-	Soil K	2
<i>Aeromonas salmonicida</i> , <i>Aeromonas sp.</i>	All	All	2
<i>Thiobacillus spp.</i>	Vegetable salad T, soil T	Soil K	2
<i>Coxiella spp.</i>	African sausage T	Fruits K, water K, samosa K, soil K	2
<i>Brucella spp.</i>	Soil T	-	2
<i>Treponema spp.</i>	-	Soil K	3
<i>Staphylococcus pasteurii</i>	Chips T, African sausage T, fruits T, roasted beef K, vegetable salad T	Fruits K, water K, samosa K, fish K, roasted beef K, soil K	3
<i>Acholeplasma hippikon</i> , <i>Acholeplasma vituli</i> , <i>Acholeplasma oculi</i>	Chips T, fish T	Samosa K, soil K, soil K	2
<i>Acholeplasma sp.</i> , <i>Acholeplasma laidlawii</i>	Chips T, fish T	Samosa K, soil K, soil K	2
<i>Erysipelothrix spp.</i> , <i>Erysipelothrix tonsillarum</i>	Soil T	Fruits K, water K, fish K, roasted beef K, fish K, soil K	2
<i>Mycobacterium sacrum</i> , <i>Mycobacterium spp.</i>	Vegetable salad T, Soil T	Fruits K, water K, soil K	3

^aNIH Appendix B: Classification of human agents based on hazard (2016)

and *Candidatus Solibacter* were detected from food samples. The abundance of this genus in the soil and sludge wastewater has been documented, and *Candidatus Microthrix* has been shown to dominate activated sludge (McIlroy et al. 2013). To our knowledge, so far, no study has reported on detection of *Candidatus*

groups in food or on their pathogenicity. In the present study, these prokaryotes may have been introduced to food from the environment. A cholera outbreak caused by *Vibrio cholerae* has been previously reported in Embu, but there are no studies that have reported on its detection in food using metagenomic techniques. So far, the reported cases in Embu County of foodborne disease have been associated with *V. cholera* (George 2016). The present study did not detect *V. cholera* sequences in soil, water, or food samples.

Table 3 Uncultured prokaryotes detected in vended foods and environmental samples from Embu Town and Kangaru Market

No	Uncultured prokaryote	Sample	Location
1.	<i>Candidatus Koribacter</i>	Fruits Soil Soil	Kangaru Market Kangaru Market Embu Town
2.	<i>Candidatus Kuenenia</i>	Soil Soil	Kangaru Market Embu Town
3.	<i>Candidatus Microthrix</i>	African sausage Soil Soil	Kangaru Market Kangaru Market Embu Town
4.	<i>Candidatus Protochlamydia</i>	Soil	Embu Town
5.	<i>Candidatus Solibacter</i>	Fruit Chips Smokies Vegetable salad Fish Soil Soil	Kangaru Market Kangaru Market Embu town Kangaru Market Embu Town Kangaru Market Embu Town

Some of the microorganisms that were detected are associated with food spoilage or foodborne diseases. *S. pasteurii* and *R. canadensis*, which belong to risk group 2, and *Brucella* spp. which belong to risk group 3 were detected in the study. The isolation and characterization of uncultured prokaryotes are important so as to verify whether the strains in the present study could be of public health concern. The development of culture techniques for these prokaryotes will facilitate isolation and understanding of their biology. There was a lot of similarity in the diversity of microorganisms between the environmental and food samples; thus, cross-contamination cannot be ruled out. Data from the current study indicates the need for proper handling of food, to avoid cases of food poisoning and associated disease outbreaks.

Acknowledgements

The provision of laboratory space and technical support by University of Embu is appreciated.

Authors' contributions

SWM conceptualized the study, collected and analysed the samples, and participated in the data analysis, interpretation of the results, and manuscript drafting. MSR participated in the data analysis, interpretation of the results, and manuscript drafting. NLMB participated in the design of the study, data collection, data analysis, interpretation of the results, and manuscript preparation and provided overall supervision of the project. The authors have approved the final submission of this manuscript for publication.

Funding

This study was funded by the University of Embu <https://www.embuni.ac.ke/>. The funding was awarded to Ms. Susan W. Muriuki. The University management had no role in study design, data collection, data analysis, or decision to publish.

Availability of data and materials

All the data from this study has been provided in the manuscript. Sequence data is available at NCBI accession number PRJNA669559. The data can be accessed using this link, <https://www.ncbi.nlm.nih.gov/sra/PRJNA669559>

Declarations**Ethics approval and consent to participate**

Not applicable

Consent for publication

Not applicable

Competing interests

Africa Genomics Center and Consultancy is a commercial entity that provides NGS sequencing and bioinformatic support. Sequencing in this study was done in the USA at Molecular Research DNA (www.mrdalab.com, Shallowater, TX, USA). No payments were made by the University of Embu to the authors to Africa Genomics Center and Consultancy. The participation of one author (M.S.R.) who is an employee of Africa Genomics Center and Consultancy in data analysis of this study was a capacity building outreach gesture from the management of Africa Genomics Center and Consultancy

Author details

¹Department of Biological Sciences, University of Embu, P.O. Box 6-60100, Embu, Kenya. ²The Africa Genomics Center and Consultancy, P. O. Box 381-00517, Nairobi, Kenya.

Received: 12 May 2021 Accepted: 28 June 2021

Published online: 12 July 2021

References

- Bokulich NA, Dillon MR, Zhang Y, Rideout JR, Bolyen E, Li H, Caporaso JG (2018a) q2-longitudinal: longitudinal and paired-sample analyses of microbiome data. *MSystems* 3(6). <https://doi.org/10.1128/mSystems.00219-18>
- Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Caporaso JG (2018b) Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6(1):1–17. <https://doi.org/10.1186/s40168-018-0470-z>
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Caporaso JG (2018) QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science (No. e27295v1). *PeerJ Preprints*. <https://doi.org/10.7287/peerj.preprints.27295v2>
- Browne P, Tamaki H, Kyripides N, Woyke T, Goodwin L, Imachi H, Cadillo-Quiroz H (2017) Genomic composition and dynamics among Methanomicrobiales predict adaptation to contrasting environments. *ISME J* 11(1):87–99. <https://doi.org/10.1038/ismej.2016.104>
- C Team C (2011) nlme: linear and nonlinear mixed effects models. In: R package version 3.1-104R Development Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna URL <http://www.R-project.org.2012>
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13(7):581–583. <https://doi.org/10.1038/nmeth.3869>
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci* 108(Supplement 1):4516–4522
- Cavicchioli R, Curmi PM, Saunders N, Thomas T (2003) Pathogenic archaea: do they exist? *Bioessays* 25(11):1119–1128. <https://doi.org/10.1002/bies.10354>
- Djenane D, Yanguela J, Roncales P (2014) A review and future potential approach for *Campylobacter* control in retail poultry meats. *Afr J Microbiol Res* 8(5):4041–4052
- Donde OO, Wairimu AM, Shivoga AW, Trick GC, Creed FI (2015) Faecal bacterial contamination of borehole water between points-of-access and points-of-use in Naivasha, Kenya; Public health implication. *Egerton J Sci Technol* 13:165–184
- Faith DP (1992) Systematics and conservation: on predicting the feature diversity of subsets of taxa. *Cladistics* 8(4):361–373. <https://doi.org/10.1111/j.1096-0031.1992.tb00078.x>
- Fakhkhari P, Tajeddin E, Azimirad M, Salmanzadeh-Ahrabi S, Abdi-Ali A, Nikmanesh B, Alebouyeh M (2020) Involvement of *Pseudomonas aeruginosa* in the occurrence of community and hospital acquired diarrhea, and its virulence diversity among the stool and the environmental samples. *Int J Environ Health Res*:1, 1–11, 11. <https://doi.org/10.1080/09603123.2020.1726300>
- Gangwar P, Alam SI, Bansod S, Singh L (2009) Bacterial diversity of soil samples from the western Himalayas, India. *Can J Microbiol* 55(5):564–577. <https://doi.org/10.1139/W09-011>
- Garcell HG, Garcia EG, Pueyo PV, Martín IR, Arias AV, Serrano RNA (2016) Outbreaks of brucellosis related to the consumption of unpasteurized camel milk. *J Infect Public Health* 9(4):523–527. <https://doi.org/10.1016/j.jiph.2015.12.006>
- George G (2016) Notes from the field: ongoing cholera outbreak—Kenya, 2014–2016. *Morbidity and Mortality Weekly Report* 65. <https://doi.org/10.15585/mmwr.mm6503a7>
- Griffiths MW, Schraft H (2017) *Bacillus cereus* food poisoning. In: *Foodborne diseases*, 3rd ed. Academic, pp 395–405. <https://doi.org/10.1016/B978-0-12-85007-2.00020-6>
- Harb M, Hong PY (2017) Molecular-based detection of potentially pathogenic bacteria in membrane bioreactor (MBR) systems treating municipal wastewater: a case study. *Environ Sci Pollut Res* 24(6):5370–5380. <https://doi.org/10.1007/s11356-016-8211-y>
- Henriet O, Fourmentin J, Delincé B, Mahillon J (2014) Exploring the diversity of extremely halophilic archaea in food-grade salts. *Int J Food Microbiol* 191:36–44. <https://doi.org/10.1016/j.jifoodmicro.2014.08.019>
- Hugenholtz F, Ritari J, Nylund L, Davids M, Satokari R, De Vos WM (2017) Feasibility of metatranscriptome analysis from infant gut microbiota: adaptation to solid foods results in increased activity of firmicutes at six months. *Int J Microbiol* 2017:1–9. <https://doi.org/10.1155/2017/9547063>
- Israel DS (2009) Using published tables using formulas to calculate a sample size using a census for small populations
- Jenkins C, Dallman TJ, Launders N, Willis C, Byrne L, Jorgensen F, McLaughlin J (2015) Public health investigation of two outbreaks of Shiga toxin-producing *Escherichia coli* O157 associated with consumption of watercress. *Appl Environ Microbiol* 81(12):3946–3952. <https://doi.org/10.1128/AEM.04188-14>
- Jung JY, Lee SH, Kim JM, Park MS, Bae JW, Hahn Y, Jeon CO (2011) Metagenomic analysis of kimchi, a traditional Korean fermented food. *Appl Environ Microbiol* 77(7):2264–2274. <https://doi.org/10.1128/AEM.02157-10>
- Karanja EN, Fliessbach A, Adamtey N, Kambura AK, Musyoka M, Fiaboe K, Mwirichia R (2020) Diversity and structure of prokaryotic communities within organic and conventional farming systems in central highlands of Kenya. *PLoS one* 15(8):e0236574. <https://doi.org/10.1371/journal.pone.0236574>
- Kariuki EN, Waitheza NZ, Wanzala P (2017) Bacteriological contamination of street foods among street food vendors in githurai and gikomba markets-Nairobi county, Kenya. *Int J Innov Res Adv Stud* 4:337–346
- Karoki WH, Karanja DN, Bebor LC, Njagi LW (2018) Isolation, characterization, and quantification of bacteria from African sausages sold in Nairobi County, Kenya. *Int J Food Sci* 2018:1–9. <https://doi.org/10.1155/2018/3861265>
- Katoh K, Misawa K, Kuma KI, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14):3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kergourlay G, Taminiau B, Daube G, Vergès MCC (2015) Metagenomic insights into the dynamics of microbial communities in food. *Int J Food Microbiol* 213:31–39. <https://doi.org/10.1016/j.jifoodmicro.2015.09.010>

- Kerou M, Schleper C (2015a) Nitrososphaera. In: Bergey's manual of systematics of archaea and bacteria, pp 1–10. <https://doi.org/10.1002/9781118960608.gbm01294>
- Kerou M, Schleper C (2015b) "Nitrososphaera," Bergey's Man. Syst Archaea Bact 25:1–10
- Kim D, Hong S, Kim YT, Ryu S, Kim HB, Lee JH (2018) Metagenomic approach to identifying foodborne pathogens on Chinese cabbage. *J Microbiol Biotechnol* 28(2):227–235. <https://doi.org/10.4014/jmb.1710.10021>
- Kumaraswamy R, Amha YM, Anwar MZ, Henschel A, Rodriguez J, Ahmad F (2014) Molecular analysis for screening human bacterial pathogens in municipal wastewater treatment and reuse. *Environ Sci Technol* 48(19):11610–11619. <https://doi.org/10.1021/es502546t>
- Law JWF, Ab Mutalib NS, Chan KG, Lee LH (2015) Rapid methods for the detection of foodborne bacterial pathogens: principles, applications, advantages and limitations. *Front Microbiol* 5:770
- Legrand R, Lucas N, Dominique M, Azhar S, Deroissart C, Le Solliec MA, Fetissovo SO (2020) Commensal *Hafnia alvei* strain reduces food intake and fat mass in obese mice—a new potential probiotic for appetite and body weight management. *Int J Obes* 44(5):1041–1051. <https://doi.org/10.1038/s41366-019-0515-9>
- Liu LH, Wang NY, Wu AYJ, Lin CC, Lee CM, Liu CP (2018) *Citrobacter freundii* bacteremia: risk factors of mortality and prevalence of resistance genes. *J Microbiol Immunol Infect* 51(4):565–572. <https://doi.org/10.1016/j.jmii.2016.08.016>
- Liu Y (2010) Methanobacteriales. In: Handbook of Hydrocarbon and Lipid Microbiology. https://doi.org/10.1007/978-3-540-77587-4_43
- Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71(12):8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>
- Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73(5):1576–1585. <https://doi.org/10.1128/AEM.01996-06>
- Mayo B, Rachid C, Alegria A, Leite A, Peixoto R, Delgado S (2014) Impact of next generation sequencing techniques in food microbiology. *Curr Genom* 15(4):293–309. <https://doi.org/10.2174/1389202915666140616233211>
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Hugenholtz P (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 6(3):610–618. <https://doi.org/10.1038/ismej.2011.139>
- McIlroy SJ, Kristiansen R, Albertsen M, Karst SM, Rossetti S, Nielsen JL, Nielsen PH (2013) Metabolic model for the filamentous *Candidatus Microthrix parvicella* based on genomic and metagenomic analyses. *ISME J* 7(6):1161–1172. <https://doi.org/10.1038/ismej.2013.6>
- Mendis N, McBride P, Faucher SP (2015) Short-term and long-term survival and virulence of *Legionella pneumophila* in the defined freshwater medium Fraquil. *PLoS one* 10(9):e0139277. <https://doi.org/10.1371/journal.pone.0139277>
- Ottesen A, Ramachandran P, Reed E, Gu G, Gorham S, Ducharme D, Brown E (2019) Metagenome tracking biogeographic agroecology: phytobiota of tomatoes from virginia, maryland, north carolina and california. *Food Microbiol* 79:132–136. <https://doi.org/10.1016/j.fm.2018.12.001>
- Peng Q, Jiang S, Chen J, Ma C, Huo D, Shao Y, Zhang J (2018) Unique microbial diversity and metabolic pathway features of fermented vegetables from Hainan, China. *Front Microbiol* 9:399. <https://doi.org/10.3389/fmicb.2018.00399>
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS one* 5(3):e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Raele DA, Galante D, Pugliese N, La Salandra G, Lomuto M, Cafiero MA (2018) First report of *Coxiella burnetii* and *Borrelia burgdorferi* sensu lato in poultry red mites, *Dermanyssus gallinae* (Mesostigmata, Acari), related to urban outbreaks of dermatitis in Italy. *New Microb New Infect* 23:103–109. <https://doi.org/10.1016/j.nmni.2018.01.004>
- Ramnarain J, Yoon J, Runnegar N (2019) *Staphylococcus pasteurii* infective endocarditis: a case report. *IDCases* 18:e00656. <https://doi.org/10.1016/j.idcr.2019.e00656>
- Richards CL, Broadway SC, Eggers MJ, Doyle J, Pyle BH, Camper AK, Ford TE (2018) Detection of pathogenic and non-pathogenic bacteria in drinking water and associated biofilms on the Crow Reservation, Montana, USA. *Microbial Ecol* 76(1):52–63. <https://doi.org/10.1007/s00248-015-0595-6>
- Rodriguez C, Taminiau B, Van Broeck J, Delmée M, Daube G (2016) *Clostridium difficile* in food and animals: a comprehensive review. *Adv Microbiol Infect Dis Public Health*:65–92. https://doi.org/10.1007/5584_2016_27
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (No. Ed. 2). Cold spring harbor laboratory press, New York
- Sifuna AW, Onyango DM (2018) Source attribution of *Salmonella* and *Escherichia coli* contaminating lake victoria fish in Kenya
- Vestergaard G, Schulz S, Schöler A, Schloter M (2017) Making big data smart—how to use metagenomics to understand soil quality. <https://doi.org/10.1007/s00374-017-1191-3>
- WHO, Food safety. 2020. Available: <https://www.who.int/news-room/fact-sheets/detail/food-safety.2020>. Accessed: 01-May-2020.
- Xia Y, Wang Y, Fang HH, Jin T, Zhong H, Zhang T (2014) Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. *Sci Rep* 4(1):1–9. 10.1038/srep06708

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

