Editon Consortium Journal of Physical and Applied Sciences [ISSN: 2710: 060X]

Issue no: 1 | **Vol no:** 4 | **May** 2025: 1-19

Establishing and characterising the genetic diversity profile of bacteria colonising fresh produce sold in open-air markets in Juja

Ian Mwangi [(1)]
Johnstone Neondo [(2)]
Alfrick Makori [(3)]
Eddy Odari [(4)]

Jomo Kenyatta University of Agriculture and Technology, Kenya
Main author email: mwangi.njuguna1@students.jkuat.ac.ke

Article History
Received: 2025-03-16
Accepted: 2025-04-15
Published: 2025-05-16

Cite this article in APA
Mwangi, I., Neondo, J., Makori, A., & Odari, E. (2025). Establishing and characterising the genetic diversity profile of bacteria colonising fresh produce sold in open-air markets in Juja. *Editon consortium journal of physical and applied sciences*, *5*(1), 1-19. https://doi.org/10.51317/ecjpas.v4i1.573

Abstract
Increase in the consumption of fresh produce and changes in their production/distribution chains, coupled with the complex biology of bacterial pathogens, limit the usefulness of conventional processing and chemical sanitising methods used in preventing foodborne outbreaks. The microbial quality of nine types of fresh produce obtained from the two selected open markets was determined by both standard quantitative and NGS techniques. Purposive sampling technique was used in this cross-sectional study design to collect fresh produce items based on their tendency to be consumed raw or with minimal processing from the two selected open markets. Standard laboratory microbe culturing techniques were used to detect the presence of faecal coliform *E. coli* and foodborne pathogen *Salmonella paratyphi.* DNA was extracted from the surfaces of samples and 16S rDNA sequences were used to analyse the diversity of microbiomes found on the fresh produce using QIIME II software. Members of the Enterobacteriaceae family were in high proportional abundances, and pathogens belonging to this family were detected in the fresh produce. Bla-TEM, is one of the most important genes encoding ESBLs predominantly in the Enterobacteriaceae family, and was prevalent in the fresh produce resistome. Findings of this study provided the much-needed genomic information about pathogenic bacteria contaminating fresh produce sold in the open market that will guide the development and deployment of reliable control/management strategies against foodborne outbreaks.

Key words: Bacteria, diversity profile, fresh produce, open air markets.

INTRODUCTION

Microbial ecology and environmental microbiology have experienced remarkable expansion over the last few decades, and as a result, natural microbial communities are now recognised to be substantially more diverse than previously believed. From the late 1990s onward, there has been an increased awareness of the natural microbiota associated with plants, but much of the early research in this field concentrated on employing genetic techniques to define the variety of microbial assemblages in soils and waterways. The importance of the plant microbiome is starting to be investigated, whether in the context of its role in plant growth and disease resistance or in the context of human consumption of plant material. This is similar to the research focus on determining the make-up of the complex microbiome of animals (and especially mammals) and its role in health and disease. In research that has tended to concentrate on pathogen detection and survival in fresh produce vegetables, the composition of this microbiome has largely gone unexplored. However, the richness of this produce-associated bacterial community is starting to become apparent with the use of next-generation 16S rRNA gene sequencing techniques.

Foods may put customers at an increased risk of contracting a foodborne illness because they are frequently not put through the necessary processes of processing to guarantee the proper removal or inactivation of harmful microorganisms before ingestion. As a result, reports of disease outbreaks linked to RTE (ready-to-eat) fruit and vegetable consumption have significantly increased in recent years, and information on these occurrences is frequently not readily available. For food producers to create effective mitigation strategies, it is essential to identify the kind and source of microbial contamination in these commodities. A wide variety of produce products have been linked to human disease outbreaks all over the world, and some products are more frequently implicated in these outbreaks than others. For instance, leafy greens like lettuce and spinach, as well as fresh herbs like parsley and basil, are known to be potential sources of bacterial infections. The consumption of leafy vegetables has been linked to a number of significant, high-profile multi-state outbreaks in the USA, such as the 2006 E. coli O157:H7 outbreak, which was linked to the consumption of bagged spinach and resulted in almost 200 cases of food poisoning and three fatalities. An international outbreak of Salmonella infection linked to tainted basil from Israel that infected at least 51 people from England, Wales, Scotland, Denmark, the Netherlands, and the USA was discovered in 2007 by a microbiological study of fresh herbs sold at retail in the UK. Freshly cut and whole melons have also been connected to a number of outbreaks in a number of nations. With 63 confirmed instances of food illness in Europe in late 2011/early 2012, watermelon from Brazil was linked to a multi-nation Salmonella infection outbreak (Byrne et al., 2014). The WHO in 2010 (Summary, accessed 2020) estimated 600 million foodborne illnesses, of which 420,000 were fatal. The WHO also reported that an estimated 18 million illnesses were caused by bacteria diarrheal disease agents, particularly enteropathogenic *Escherichia coli* and non-typhoidal *Salmonella enterica*. The threat of foodborne illnesses derived from fresh produce has resulted in a substantial burden to public health and has rightly become a concern.

LITERATURE REVIEW

Microbial Contamination of Fresh Produce

As developing environmental contaminants that pose a rising risk to human health, pathogenic bacteria and antibiotic resistance genes (ARGs) are thought to be present in fresh vegetables (Yin et al., 2022). However, nothing is known about the frequency of pathogens in the phyllosphere of fresh vegetables or how ARGs are related to bacteria that cause disease. Both as endophytes within plant tissue and as epiphytes on the plant surface, plants are home to a varied bacterial community. Others may be human pathogens, while other plant-associated bacteria behave as plant pathogens or encourage plant growth. Numerous sources provide fresh food, which is frequently consumed raw. As a result, it must be handled carefully from the farm to the table to prevent contamination of the product by the farmer, shipper, processor, food service operators, merchants, or customers. As a result, it is sensitive to foodborne pathogen contamination. Production, harvesting, processing, and distribution activities all frequently involve contamination. Fresh produce, rather than more conventional carriers like poultry, pork, and shellfish, is one of the main food sources

of foodborne illness outbreaks (CDC, 2022). ARB in the environment have been detected in fresh produce and consequently attributed to disease outbreaks. Data from the Centres for Disease Control and Prevention (CDC) show that fresh produce accounted for 46 per cent of all foodborne illnesses and 25 per cent of deaths in U.S. outbreaks. Additionally, according to CDC data, produce-related outbreaks from 1998 to 2016 totalled between 30 and 60 per year, resulting in 900 to 3000 illnesses. Reports from the United States FDA, CDC, and state and local officials estimated that 30 - 60 per cent of outbreaks were linked to fresh produce, especially leafy greens, which accounted for 10 - 40 per cent of produce-related illnesses. The threat of foodborne illnesses derived from fresh produce has resulted in a substantial burden on public health and has rightly become a concern.

It is crucial to evaluate the natural microbiome of edible plants because some microbial assemblages may lessen the possibility of disease colonisation or survival. For instance, lettuce with a more diversified endophyte population has shown lower levels of Salmonella enterica colonisation. Although the exact mechanism for this pathogen reduction with greater endophyte variety is unknown, it's possible that increased overall diversity may enhance the possibility of antagonists to *S. enterica* being present. This idea is further supported by a study that looked at the viability of E. coli O157:H7 on romaine lettuce and found that the diversity of bacteria in the phyllosphere differed between plants with culturable *E. coli* O157:H7 cells and plants where the E. coli was no longer viable. It has been demonstrated in lettuce and alfalfa sprouts that native plant-associated microorganisms can compete with potential human pathogens like *Salmonella* species and *E. coli* O157:H7 indicating that even in the absence of specific antagonistic interactions, natural phyllosphere and endophytic communities may limit the presence and abundance of pathogenic bacteria by merely outcompeting them in the living plant. Therefore, understanding these societies' organisational structures may offer insights into disease outbreaks brought on by products and perhaps inspire the creation of methods for calculating the likelihood of these outbreaks. Even a potential strategy to lessen enteropathogen contamination of fresh fruit has been put forth: the deliberate insertion of competitive native microbiota.

Fresh produce has been connected to microbial contamination at various stages in the supply chain. Fresh produce's microbiological quality is routinely examined for indicator organisms and pathogens, the most frequent of which are the aerobic plate count (APC), total coliforms (TC), and generic E. coli. A good indicator of the microbiological safety of fresh fruit and the environment in which the produce is grown and processed is the absence of coliforms/E. coli. Microbial contamination at the pre-harvest stage includes the growing field. Johannessen et al. (2002) postulated that plant surface flora mirrors the environmental flora of the field in which it is grown, leading to microbial contamination at the pre-harvest stage. This therefore implies the occurrence of pathogenic microorganisms on fresh produce derived from contaminated irrigation water, agricultural soil and untreated or incompletely composted manure. Pathogenic microbes have also compromised the quality of fresh produce in the farm-to-table continuum at the post-harvest stage. Possible sources of microbial contaminate can be ascribed to (i) poor construction and maintenance of sanitation facilities and drainage system; (ii) unchecked source and quality of water being on the fresh produce; (iii) disease carrying agents such as flies in proximity of the fresh produce; (iv) uncontrolled conditions such as temperature and humidity during temporal storage of the fresh produce; and inter alia (v) general disregard for hygiene during transaction processes. However, Amoah et al. (2007) concluded that sometimes the microbial contamination of crops acquired during pre-harvest may be too high to mask any significant contribution of poor handling and transport of the product. Ndiege et al. (2017) also derived no significant association between observation of hygiene and the microbial burden on food. However, Harris et al. (2018) inferred a significant increase in microbial contamination of fresh produce held in retail markets compared to wholesale markets. This relationship suggests that the transport process or conditions at the retail markets confer contamination on the produce. Tatsika described the inefficiency of household decontamination methods, but what further aggravates the risk concern is that some fresh produce is consumed raw or lightly processed.

Since there are many stages in the cultivation, harvesting, and processing of open fresh produce, contamination can be introduced at any of these sites and then transferred to the customer. Prior to now, it was believed that the post-harvest wash procedure was adequate to get rid of contamination picked up in the field. As a result, many studies have been conducted on evaluating or creating efficient sanitisers. As information grew, it became clear that post-harvest washing in commercial settings had a limited ability to eliminate contamination and, at worst, may even result in cross-contamination incidents (Barrera et al., 2012). This is the current concept in place to protect fresh fruit from contamination in the field and to reduce cross-contamination during post-harvest handling. Even good agricultural practices (GAP) are insufficient to guarantee that human infections are not introduced into the supply chain for fresh food because it is difficult to prevent contamination in fields or greenhouses. Applying post-harvest decontamination measures, which can either replace or enhance post-harvest washing, is a more efficient means of control.

Instead of directly controlling infections of the decay pathogens within the produce, this is generally accomplished by sanitising wash water, produce surfaces, equipment, and storage areas (Feliziani et al., 2016). The most commonly used disinfectant is chlorine, which is sprayed on or dipped in water. After sanitisation, the product may be treated with one or more fungicides, which leave behind a residue that prevents pathogens that cause decay from infecting later or evading the sanitiser's effects (Feliziani et al., 2016). Sanitisers are also frequently used to reduce the contamination of produce with viruses that pose a threat to human health. Pathogens of fungal degradation are very different from pathogens of humans. Because the plant serves as their principal food source, plant infections can grow quickly inside and digest the host tissue, in contrast to Salmonella spp., Listeria spp., Escherichia coli, and other human pathogens and viruses. The control of human diseases manifests as a decrease in colony-forming units since their populations are made up of single cells. On the other hand, fungal post-harvest infections start out as isolated propagules before evolving into a networked fungal mass deep inside the host. The best way to measure their control is a decrease in the proportion of infected specific produce items (Feliziani et al., 2016).

Laser scanning confocal microscopy, epifluorescence microscopy, and the recovery of viable cells from the interior tissues of plants were used to demonstrate the transmission of Escherichia coli O157:H7 from manure-contaminated soil to lettuce plants and irrigation water to lettuce plants. E. coli O157:H7 moved to interior regions in plant tissue, where it was inaccessible and hence shielded from the action of sanitising agents (Solomon et al., 2002). Experiments show that E. coli O157:H7 can move throughout the edible part of the lettuce plant after entering through the root system (Solomon et al., 2002). Escherichia coli O157:H7 has been isolated from fresh produce more frequently in recent years, including leaf lettuce, cantaloupes, apples, and bean sprouts (Ackers et al., 1998). According to one theory, the plant gets polluted when it is grown in fields that have badly treated manure as fertiliser (Beuchat, 1999). According to epidemiological data, E. coli O157:H7 is shed asymptomatically in the faeces and may be present in up to 8.3 per cent of dairy and beef cattle. Current manure-handling recommendations include composting the manure before using it as fertiliser on a field (Food and Drug Administration, 1998). Even thorough adherence to the recommendation may result in the application of manure containing culturable E. coli O157:H7 to production fields because studies have shown that E. coli O157:H7 can survive for an extended period of time in manure maintained under a variety of circumstances.

Flood irrigation using contaminated water from cow excrement or contact with polluted surface runoff are two additional ways that E. coli O157:H7 may be spread (Ackers et al., 1998). In addition, research has shown that the pathogen can survive for long periods of time in water (Chalmers et al., 2000) and has been related to a number of recent E. coli O157:H7 outbreaks.

Metagenomic Approach to Characterisation of Bacterial community

The study of culturable and unculturable microorganisms has been enabled by metagenomics through the analysis of genomic data obtained

directly from an environmental sample, which provides knowledge of the species present. Obtaining a pure culture is a major step in any study involving traditional microbiology. However, less than 1 per cent of the information about bacterial diversity in a given environmental sample is provided by standard laboratory techniques (Coughlan et al., 2015). Significant advances have been made in culturing as-yet-uncultured microbes, although culture-independent techniques still provide a more promising way of accessing genetic information contained within the vast number of species in the environment. Metagenomics offers a molecular tool to analyse the DNA of microorganisms in a population as a whole. Analysis of total metagenomic DNA and sequencing can offer insight into metabolic activities and functional roles apart from species identity, giving better characterisation of the microbial life in a given environment (Coughlan et al., 2015).

The current advancements in DNA sequencing technologies have not only allowed for a finer characterisationcharacterisation of bacterial genomes but have also allowed for a deeper taxonomic identification of complex microbiomes, which, in terms of genomics, are the combined genetic material of the microorganisms inhabiting a particular environment, whether that environment be a body econiche (for example, the contents of a human intestine) or a food manufacturing facility econiche (for example, a floor drain) (Cao et al., 2017). The three primary sequencing techniques employed to date in the taxonomic identification and characterisation of food-related microbiomes are 16S rDNA sequencing, metagenomics, and metatranscriptomics (Cao et al., 2017). These sequencing techniques have identified DNA and RNA sequences using several NGS platforms. Traditionally, a food-related microbiome's taxonomic composition has been largely understood through 16S rDNA sequencing. Recently, species-level/strain-level characterisations provided by metagenomic methods have increased understanding of a microbiome (Cao et al., 2017). The functional characterisation of the intricate relationships between various microbial communities within a single microbiome has also benefited from metatranscriptomic techniques (Cao et al., 2017).

The development of next-generation sequencing (NGS) techniques could be advantageous over current methodologies for detecting foodborne infections. Theoretically, NGS permits the non-targeted detection of several spoiling agents and diseases as well as the characterisation of both culturable and non-culturable species (Lewis et al., 2020). Current approaches must be utilised to evaluate the make-up of the bacterial community on vegetables. When the objective is to identify the presence or viable abundance of particular bacterial populations, such as pathogens or indicator species, whose culture requirements are known, and for which specific selective and differential growth media exist, the use of culture-dependent methods is unquestionably justified. However, culture-independent molecular techniques make it possible to investigate the complete bacterial community, permitting a more thorough analysis of the existing microbiome. As they become more accessible, emerging technologies like next-generation sequencing will probably compete with old culture procedures in the routine assessment and monitoring of produce. These technologies can be used to detect populations that may be missed by standard culture approaches.

Targeted metagenomics involves targeting 16 S ribosomal RNA (16S rRNA) genes as taxonomical markers using PCR amplicon sequencing. 16S rRNA genes comprise of conserved and variable regions that provide identities of operational taxonomic units in a given sample. 16S rRNA method of characterising bacterial species is a genotypic technique that prevails without recourse to the culture-dependent phenotypic assessment (Clarridge, 2004). The 16S rRNA technique analyses for the 16S rRNA gene which is an approximately 1500 base pair gene encoding for a portion of the 30S ribosomal subunit. Next Generation Sequencing (NGS) is limited to short read sequences; hence, the development of primers targeting partial sequences in hypervariable regions, e.g. V1-V2, V1-V3, V3-V4, V4, etc. Chen reported that V3-V4 is more suitable for analysing gut microbiota (Kameoka et al., 2021).

Due to its cheaper cost, earlier stage of development, and standardised analytical workflows, amplicon sequencing, for instance of the 16S rRNA gene (16S)

for bacteria, is now the most extensively used technology for microbiome investigation (Cao et al., 2017). According to theory, these pipelines generate precise and repeatable results that enable comparisons between investigations (Thompson et al., 2017; Bolyen et al., 2019). The utilisation of 16S gene data can provide a trustworthy indicator of bacterial diversity in the field of food microbiology (Caporaso et al., 2011).

16S rRNA gene propensity for use as a molecular marker in phylogenetic analyses is in reliance on a number of reasons; (i) ubiquity in bacteria, almost invariably, usually existing as a multigene family, or operons; (ii) highly conserved function of the 16S rRNA gene over the evolutionary timeline; and (iii) a large enough sequence for comprehensive bioinformatics. 16S rRNA triumphs over culture-dependent techniques in its ability to provide rapid, accurate and determinate species-level identification of bacteria including isolates that do not fit in any recognised biochemical profiles.

METHODOLOGY

Study Areas Description and Produce Sample Collection

The study will be conducted at markets within the Juja area in Kiambu County, Kenya, and will adopt a cross-sectional design that cuts across a variety of fresh produce types and vendors in Juja open markets. The study area will be stratified into two sampling sites, and a purposive/ convenient sampling technique will be used to collect fresh produce items based on their tendency to be consumed raw or with minimal processing and location of display. The two open-air markets within close proximity to each other were sampled for fresh fruits and vegetables at the point of sale. The two open-air markets differ, by general observation, in terms of the level of infrastructure and the amount of foot traffic they receive. Juja market (latitude -1°6' 3.9168" N and longitude 37°0' 56.368" E) is the main market within Juja, while Gachororo market (latitude -1°5' 20.2452" N and longitude 37°1' 11.0748") s less popular. The main market usually teems with suppliers, vendors and customers by 6 AM. On the other hand, on average, the majority of the traders in Gachororo will have only opened shop by noon after purchasing their stock from the main market. This could in turn imply that produce items in Gachororo have experienced an extended storage and shelf period compared to foods sold in the Main market. Notwithstanding, there is a discrepancy in infrastructural development between the two markets. The main market has clear demarcations and a designated garbage disposal site, albeit right at the entrance, and still remains congested during peak hours. Some vendors in the main market practice wholesale and in part specialise in specific produce or a choice of related types of foods, e.g., collard greens and spinach. The stalls in Gachororo comprise elementary wooden shelters lining opposite sides of a 100-meter stretch of dirt road. Lacking any form of boundaries, the stalls are then adjacent to households on either side, and the area is compromised by a poor sewerage system. This is evidenced by the pools and streams of grey water gutting either side of the road and persisting even during dry seasons. Traders in Gachororo markets have a diverse selection of merchandise within the same display, some extending up to silverfish that constantly attract a plethora of flies. Suppliers of fresh produce come from within and outside Kiambu County.

Food items were inserted into sterile zip bags by the vendor. Fresh produce collected included: carrots, tomatoes, tamarillo fruits, spinach, bell peppers, cabbages, collard greens, zucchinis, mangoes and watermelons. For each produce type, three technical replicates were collected from three different vendors for the same item within a market setting. Samples were inserted into sterile zip bags by the vendor and analysed immediately after reaching the laboratory.

Bacterial DNA Extraction, Amplification and Sequencing

Sterile cotton swabs soaked in phosphate-buffered saline (PBS) were used to harvest microbiota on the surfaces of fresh produce (Sare et al., 2020). DNA extraction was done using the Isolate II Genomic DNA Kit from Meridian Bioscience Company according to the manufacturer's guidelines. PCR amplification of the 16SrRNA gene V4–V7 variable regions was carried out on the extracted DNA using primers; 515F (GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) as previously described by Ghilamicael et al. (2018). The amplification was done in 30 cycles using 5X

MyTaq Bioline® Reaction Buffer 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 extension step at 72°C performed for 5 min and stored at 4°C. The quality of the PCR products was assessed using a 2 per cent agarose gel to determine the success of the amplification and the relative intensity of the bands. The PCR products were then pooled together according to the sampling market and produce type and shipped for sequencing. Sequencing was performed at Molecular Research DNA (www.mrdalab.com, Shallowater, TX, USA) on a MiSeq 2x300bp following the manufacturer's guidelines.

Microbiome Diversity Analysis

QIIME 2 version 2021.4 was used to analyse the sequences for microbiome diversity (Bolyen et al., 2019; Muriuki et al., 2021). Sequences arrived from the sequencing facility in CASAVA 1.8 format, having undergone some level of quality assessment by Illumina Miseq software. This quality was assessed by generating quality plots based on the Phred score. Another quality filtering step was performed using DADA2, which is a denoising tool provided by QIIME 2 software. The sequences were filtered to a read length of 240 bp before merging the forward and reverse reads, and lastly, discarding chimaeras. To account for the uneven sequencing depth across samples, the data were normalised by rarefying. Rarefaction is a method for sample normalisation via sub-sampling without replacement that reduces samples to a uniform sampling depth of 127,000, i.e., maintaining the number of features from which a comparison of diversity/abundance can be made. A phylogenetic tree was generated profiling genetic diversity metrics, including Faith's Phylogenetic Diversity and Unifrac. PERMANOVA test was used to determine the significance of the beta-diversity by analysing the within-group distances from each group and how different they are from the between-group distance. Taxonomic classification was performed using consensus BLAST against the SILVA 132 reference sequence database. Interactive bar plots were generated to display relative abundancies of organisms at different taxonomic levels. Lastly, machine learning classifiers available in the QIIME 2 pipeline were utilised to see if microbiome composition can predict sample characteristics. Sequences were submitted to the NCBI data repository under the accession codes SRR15245134 — SRR15245150

FINDINGS AND DISCUSSION

Presence of Microbiological Contaminants in Fresh Foods

Microbiological assessment for *E. coli* resulted in rose red colonies on MacConkey agar, while *Salmonella* sp. were detected by cloudy colonies on MacConkey agar plates. Pure isolates of *Salmonella* sp. were identified as *Salmonella paratyphi* using biochemical tests. Fecal coliform *E. coli* was detected in all fresh produce samples except for onion samples while *Salmonella paratyphi* was present in all samples except apples **(Figure 3)**. Cilantro, spinach, chilli and tamarillo fruit samples had the same levels of contamination with both organisms **(Figure 3)**.

Figure 1. Presumptive *E. coli* colonies on MacConkey agar

Figure 2. Presumptive *Salmonella* sp. colonies on MacConkey agar

Table 1. Results for Biochemical Analysis on Presumptive Pure Isolates

Presumptive Isolate for	*Escherichia coli*	*Salmonella paratyphi*
Gram stain	-ve	-ve
Catalase test	+ve	+ve
Oxidase test	-ve	-ve
Citrate test	-ve	+ve
Motility test	+ve	+ve
Indole test	-ve	-ve
Urease test	-ve	+ve
Methyl red test	+ve	+ve
Triple Sugar Iron	AG/A	AG/NC

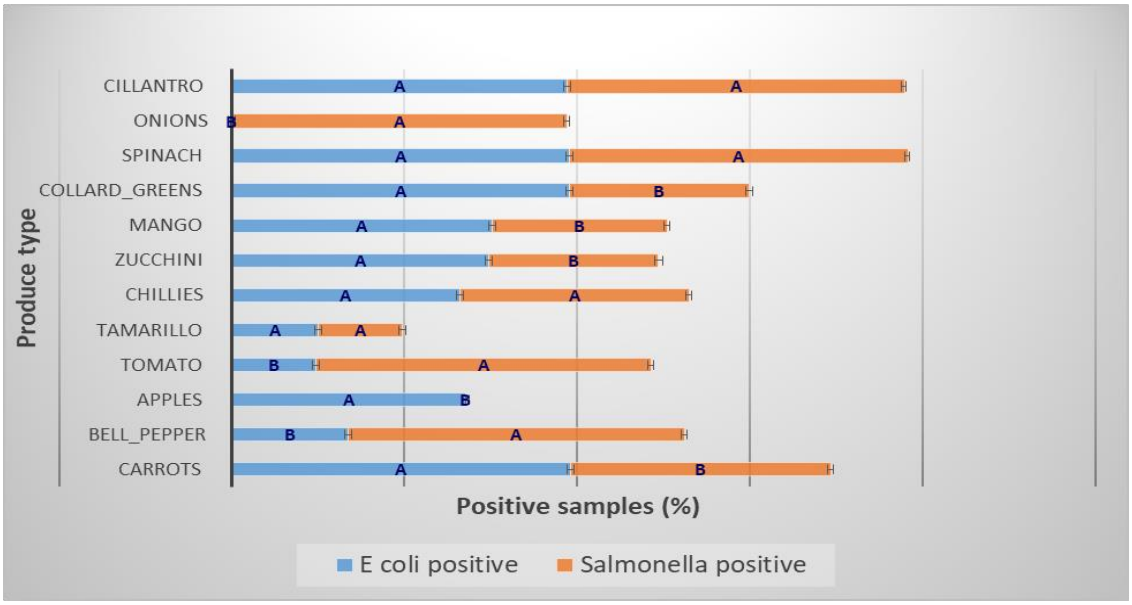


Figure 3. Analysis of Means of Samples Positive for *E. coli* and *Salmonella paratyphi*

Cilantro, spinach, chilli and tamarillo fruits recorded homogeneous means for samples positive for both *E. coli* and *Salmonella paratyphi*. *E. coli* was absent in all onion samples.

Bacterial Species Diversity Varied by Market

DNA harvested from the microbiota of fresh produce was first subjected to PCR amplification targeting the v4 region of the 16S rRNA gene using primers 515F and 806R. The amplification yielded amplicons of approximately 300 bp. A total of 3,008,879 reads were recovered from the 17 pooled samples in which 2468 distinct OTUs were detected and the distribution of reads per sample was uneven, ranging from 252609 in carrots from Gachororo to 63429 in zucchinis from Juja market (**Table 2**). Data was normalised using rarefaction curves that plateaued at a sampling depth of 127, 000 resulting in the loss of one sample prior to downstream alpha and beta diversity analysis (**Figure 5**).

Table 2. Total Number of OTUs per Sample

Produce type	**Market of origin**	**Number of OTUs**
Carrots	Gachororo	252609
Collard greens	Gachororo	217370
Chili peppers	Juja market	211442
Tamarillo fruits	Juja market	203097

Water melons	Juja market	202522
Cabbages	Gachororo	201752
Bell peppers	Gachororo	187814
Carrots	Juja market	183054
Tomatoes	Juja market	182074
Collard greens	Juja market	173258
Cilantros	Juja market	172232
Mangoes	Juja market	171672
Tomatoes	Gachororo	159929
Bell peppers	Juja market	156507
Mangoes	Gachororo	146215
Zucchinis	Gachororo	127503
Zucchinis	Juja market	63429

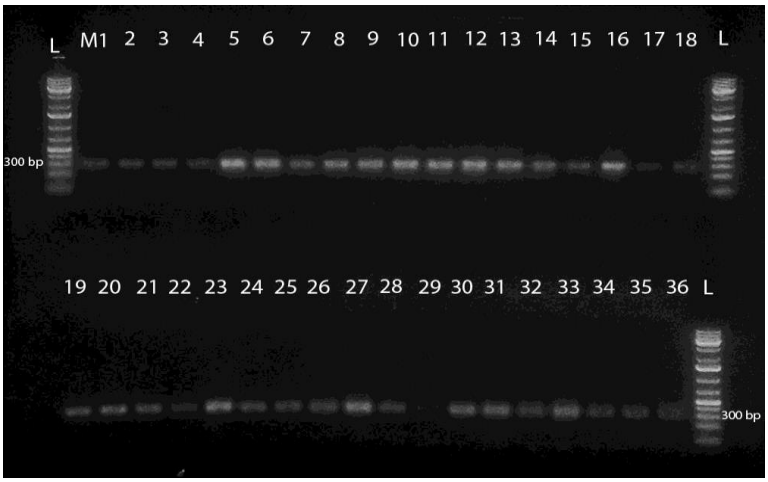


Figure 4. PCR Amplification of DNA from Fresh Produce Microbiota

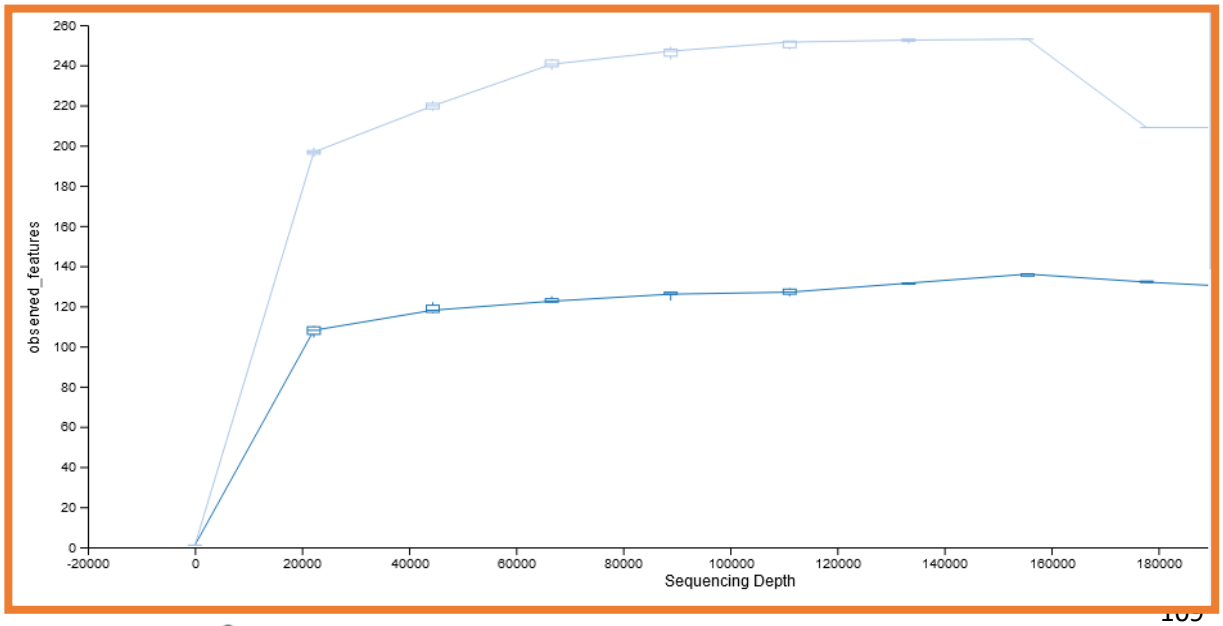


Figure 5. Data Normalisation: Normalisation via Rarefaction Curves

Juja market fresh food samples had significantly higher numbers of OTUs detected compared to Gachororo market fresh produce samples (p = 0.05) **(Figure 6)**. Weighted Unifrac distance measures revealed striking variations among the microbiota of the fresh foods **(Figure 7)**. The first axis, accounting for 50.50 per cent of the variation in beta-diversity, separates the samples into two groups corresponding to the market of origin. However, when comparing the Faith's Phylogeny **(Figure 8)** and Pielou's evenness **(Figure 9)** alpha-diversity indices of the two markets, the difference observed was not significant. The alpha-diversity results were conflated by Venn diagrams, which revealed 38 per cent of total phyla were shared across both markets, and the Gachororo market had only one phylum unique to it **(Figure 10)**.

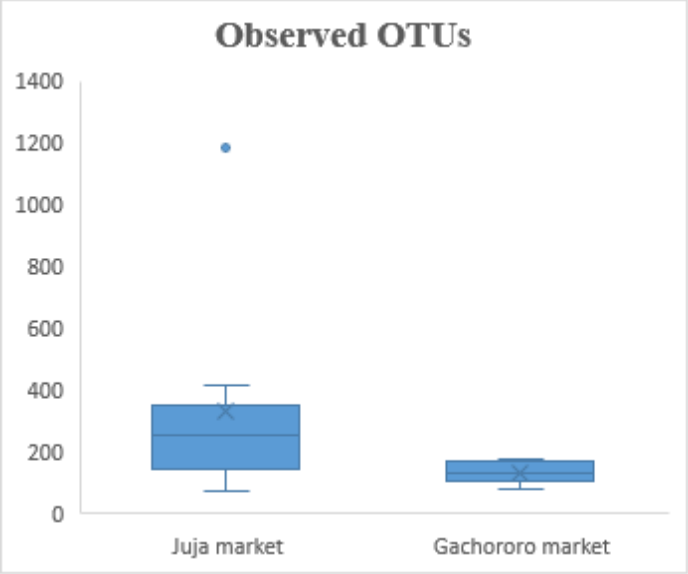


Figure 6. Observed OTUs box plot
Juja market samples had a significantly higher mean of detected OTUs compared to Gachororo market at a p-value = 0.05

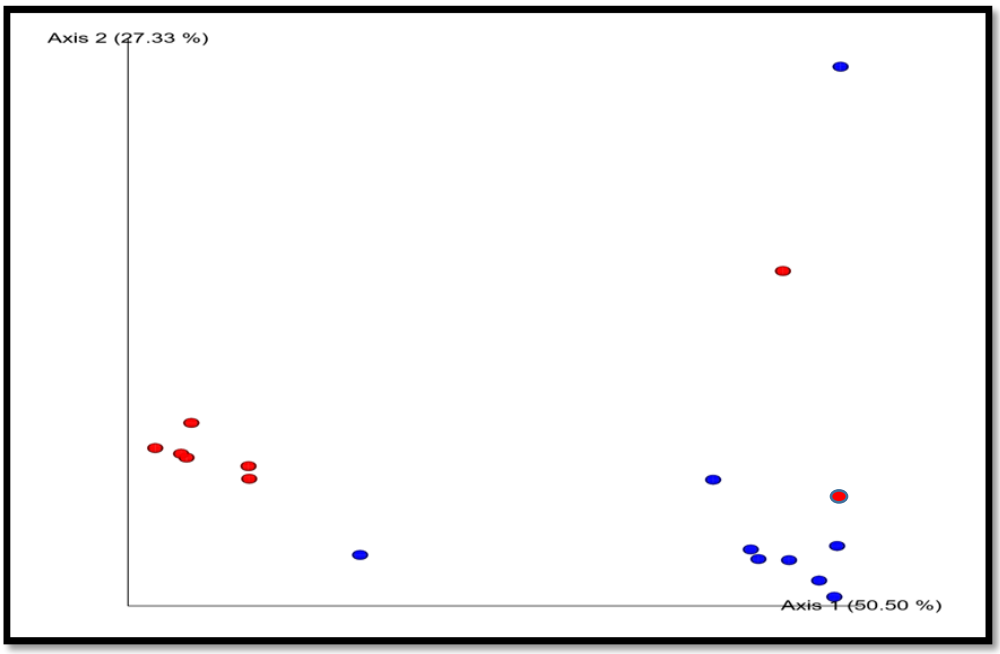


Figure 7. Weighted Unifrac PCoA plot
Data points are separated into distinct clusters due to the evolutionary branch lengths and abundances of the species/ OTUs registered in either market. •refer to Juja market's data points and Gachororo's data points.

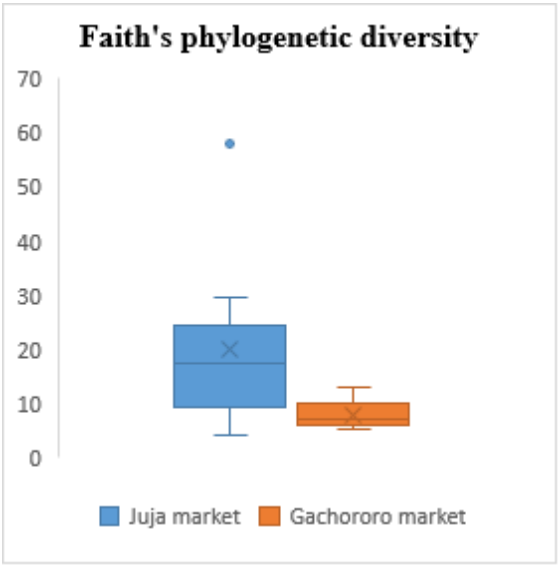


Figure 8. Faiths Phylogenetic Diversity: Box Plots
Faith's phylogeny alpha diversity metric indicates that the Juja market has more bacterial diversity; however, it was not significantly different from Gachororo (p-value = 0.06).

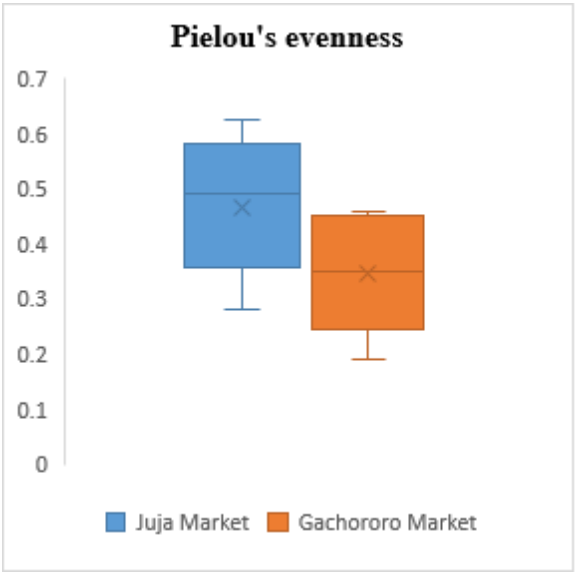


Figure 9. Pielou's Evenness Box plots
Juja market samples have a more complete evenness, that is, equality in the number of individuals in each species present, compared to Gachororo although is not significantly different (p-value=0.06).

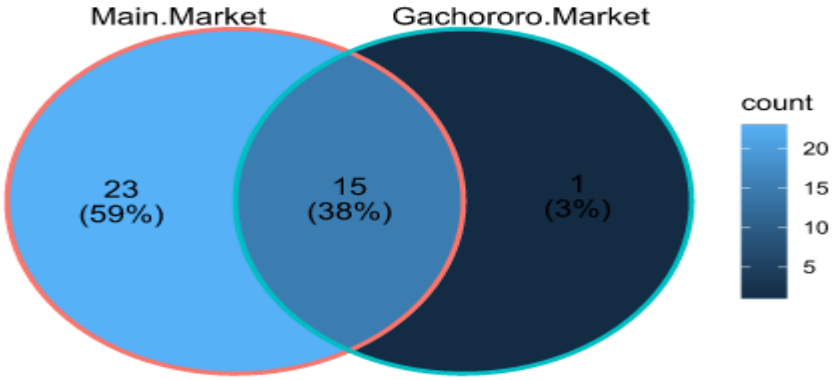


Figure 10. Shared Phyla between the two Markets.

After assigning taxonomy, the predominant taxa dictating the diversity topography between the two groups were revealed. Firmicutes was the dominant phylum in Gachororo market samples, with proportional abundances at an average of 75.4 per cent, while Juja market samples were dominated by Proteobacteria, averaging at 80.7 per cent (**Figure 11)**. Collard greens obtained from Juja market differed from other fresh produce samples, whereby Cyanobacteria (55.6 %) were the most abundant phylum (**Figure 11)**. Similarly, water melons from Juja market differed from other produce samples, where Firmicutes (53.5 %) were predominant, followed by Proteobacteria (45.7 %) (**Figure 11)**. In the case of Gachororo produce samples, tomatoes recorded two highly abundant phyla, Proteobacteria (44.6%) and Cyanobacteria (42.1 %), and Firmicutes reduced to 11.6 per cent (**Figure 11)**.

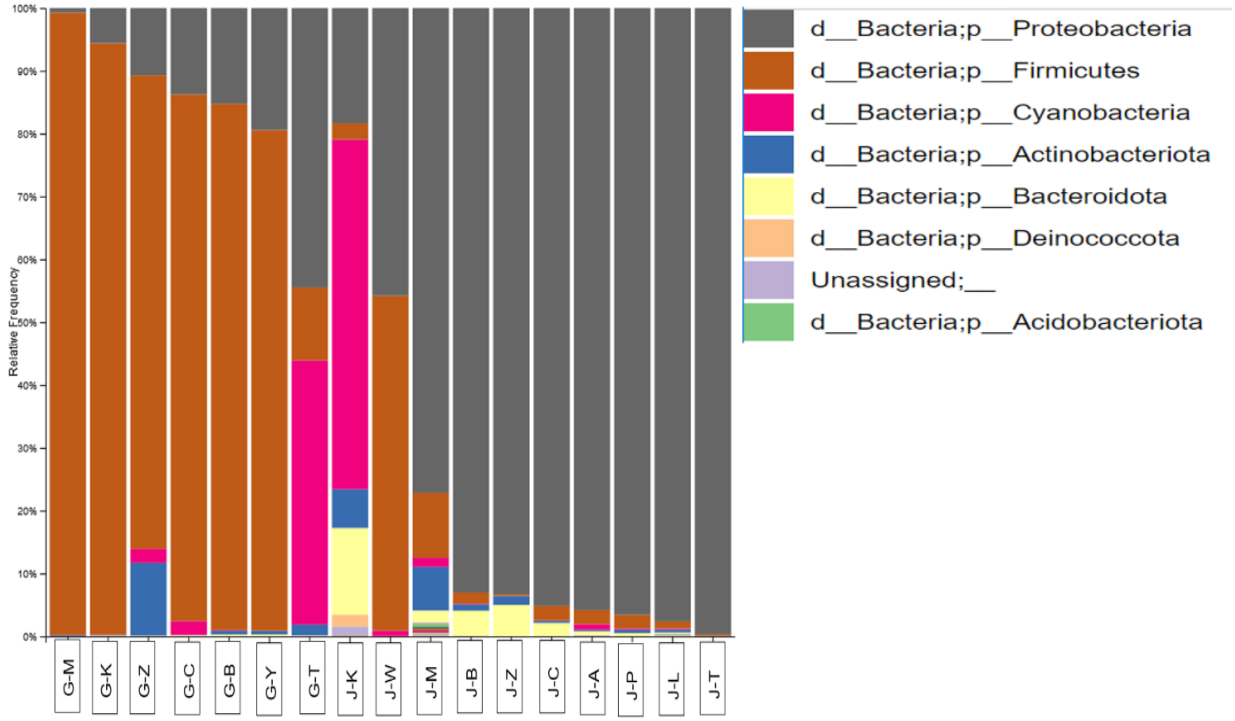
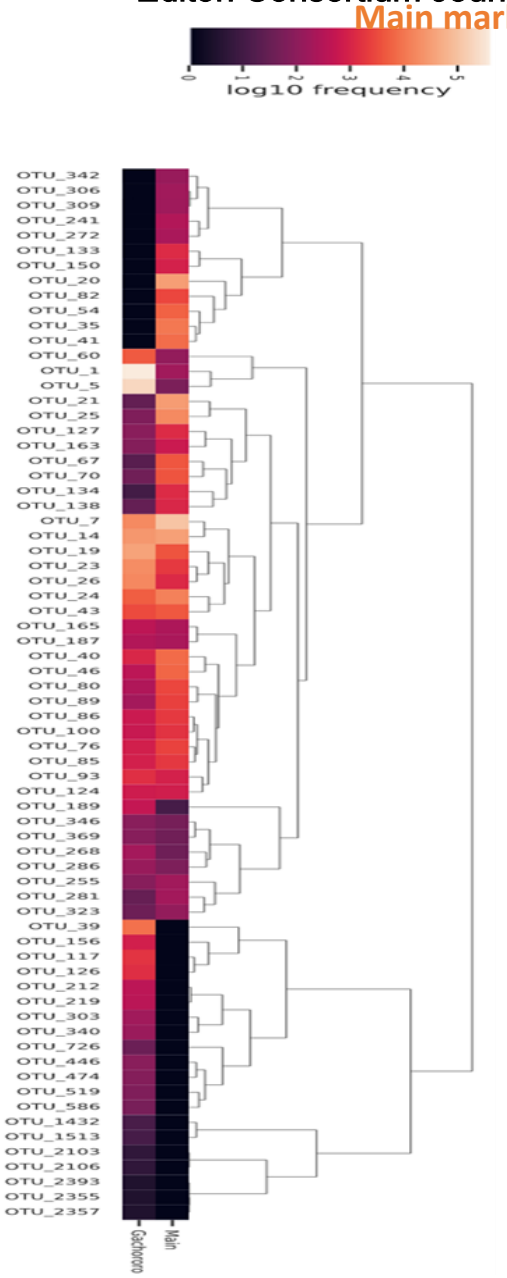


Figure 11. Taxonomic Bar Chart of Bacterial Community at Phylum Level

Distribution of the top seven phyla across produce types in both markets. **Legend:** Gachororo (G-): Mango (G-M), Collard greens (G-K), Zucchini (G-Z), Carrots (G-C), Bell peppers (G-B), Cabbage (G-Y), Tomatoes (G-T). Juja market (J-): Collard greens (J-K), Water melon (J-W), Mangoes (J-M), Bell peppers (J-B), Zucchinis (J-Z), Carrots (J-C), Tamarillo fruits (J-A), Chilli peppers (J-P), Cilantro (J-L), Tomatoes (J-L)



Main market defining features

OTU ID	Taxonomic Classification
OTU_342	Paenibacillus (genus)
OTU_306	Paenibacillus (genus)
OTU_309	Noviherbaspirillum (genus)
OTU_241	Shinella (genus)
OTU_272	Pseudorhodoferax (genus)
OTU_133	**Acidovorax (genus)**
OTU_150	**Bacillus (genus)**
OTU_20	**Enterobacteriaceae (family)**
OTU_82	**Sphingobacterium (genus)**
OTU_54	**Klebsiella (genus)**
OTU_35	**Pantoea (genus)**
OTU_41	**Pseudomonas (genus)**

Gachororo market defining features

OUT ID	Taxonomic Classification
OTU_39	**Bacillus (genus)**
OTU_156	**Staphylococcus (genus)**
OTU_117	**Bacillus (genus)**
OTU_126	**Bacillus (genus)**
OTU_212	**Hymenobacter (genus)**
OTU_219	**Acinetobacter (genus)**
OTU_303	Rhodobacteraceae (family)
OTU_340	Aneurinibacillus thermoaerophilus
OTU_726	Nosocomiicoccus (genus)
OTU_446	Desulfovibrio (genus)
OTU_474	Bacillus (genus)
OTU_519	Brevibacterium (genus)
OTU_586	Blastocatellaceae (family)

Figure 12. Heatmap Highlighting the Top 70 Features that can be used to Predict which Market a Sample came from

Furthermore, clustering using the Random Forest algorithm revealed OTUs that were prevalent in one market with frequencies above 300 and completely absent in the other **(Figure 12)**. For example, the Main market samples had high means for *Klebsiella* spp. (321), *Pantonea* spp. (592), *Pseudomonas* spp. (434), all of which were totally missing from Gachororo samples **(Figure 12)**. On the other hand, *Bacillus* spp. (average = 492) dominated fresh foods from Gachororo and could not be detected in any of the Juja market samples **(Figure 12)**.

Discussion

According to Berg et al. (2014), vegetables have extremely varied microbiomes that act as reservoirs for opportunistic infections. In Kenya, recent cholera outbreaks that resulted in 76 deaths and 3967 illnesses in the first eleven months of 2017 are most

likely caused by tainted food (Cholera - Kenya, 2021). The onset of foodborne outbreaks is significantly influenced by the presence of enteropathogenic bacteria in fresh produce. The faecal indicator organism E. coli was found in abundance in cilantro, spinach, collard greens, mangoes, zucchinis, chillies, tamarillo fruits, tomatoes, apples, bell peppers, and carrots in this investigation. Similar results were observed by Harris et al., who hypothesised that regional market factors influence produce contamination. Food safety indicator organisms are extensively employed to assess poor sanitation since their presence serves as a marker for the potential appearance of ecologically comparable pathogens. Escherichia coli's detection in food has been utilised to signal a higher possibility that pathogens like Salmonella and E. coli O157:H7 were also present in food since it was discovered to be common in faeces (Halkman & Halkman, 2014). An ideal food safety indication would be absent from foods that are devoid of the target pathogen and present whenever the pathogen is present. In our investigation, Salmonella paratyphi was found in a variety of foods, including cilantro, onions, spinach, collard greens, mangoes, zucchini, chillies, tamarillo fruits, tomatoes, bell peppers, and carrots. Salmonella is a rod-shaped, motile, gram-negative, non-spore-forming bacterium that belongs to the tribe Salmonellae and the family Enterobacteriaceae. Nature contains a lot of Salmonella. It may thrive in conditions like pond water sediment and can colonise the intestines of animals (Nutrition & Services, 2012). The faecal-oral route is how it is transmitted (Nutrition & Services, 2012). Only human hosts can harbour S. Typhi and S. Paratyphi, and untreated sewage is typically the source of these organisms' contamination in drinking and/or irrigation water (Nutrition & Services, 2012). Salmonella has been linked to fresh produce, including low-moisture items like raw nuts and spices. Onions, for instance, have been connected to the most current salmonellosis outbreak (Salmonella Outbreak Linked to Onions, 2021).

Microbiological testing revealed faecal contamination, demonstrating an ecological link between pathogenic Salmonella and other restricted human gut commensals. However, the use of culture-based techniques is constrained because they only identify culturable microbes, which frequently make up a small portion of the microbiome, take a long time (often days to weeks) to produce a positive result, and require a lot of money and labour when applied on a large scale. Additionally, this type of conventional strategy typically only permits the identification of one species at a time, necessitating the use of various approaches according to the range of species being targeted (Lewis et al., 2020).

For outbreak detection and treatment, quick identification and subtyping of foodborne pathogens, along with targeted epidemiologic inquiry, are essential (Alegbeleye & Sant'Ana, 2020). Due to its untargeted nature and capacity to research non-culturable (and/or difficult to culture) organisms while producing genomic data about the microbiota, the Illumina sequencing technology was utilised in this study to trace microbial contaminants on fresh produce (Lewis et al., 2020). Numerous bacteria have been linked to fresh produce thus far, but only a small number of studies have used NGS to detect contamination with human diseases (Lewis et al., 2020). More pathogens that are on the WHO's list of global priority diseases were found in this investigation using the NGS approach (Asokan et al., 2019a). In samples of produce from the Juja market, Enterobacteriaceae were most frequently found at the critical tier level. Bla-TEM predominated among the majority of the ARGs examined in the produce resistome for the Enterobacteriaceae (Figure 13). Members of the Enterobacteriaceae family produce large amounts of Bla-TEM, one of the genes encoding for extended-spectrum beta-lactamases, which provide resistance to host bacteria by changing the target site of penicillin and cephalosporin antibiotics (Bajpai et al., 2017). This incident was not definitive, though, and more research employing whole genome sequencing (WGS) is necessary.

Enterobacter sp., while being regarded as a critical tier pathogen (Asokan et al., 2019a; Tacconelli et al., 2017) dominated produce samples from Juja market at very high proportions; mangoes (5.7%), water melons (14.9%), carrots (7.4%), zucchinis (1.6%), bell peppers (9.7%), tomatoes (58.9%), cilantro (5.6%), chili peppers (13.4%), tamarillo fruits (0.8%) (**Figure 14**). Another new opportunistic pathogen, Acinetobacter baumannii, was discovered in carrots and zucchini from Juja market after being found in

bell peppers from both marketplaces (Asokan et al., 2019b; Howard et al., 2012). Another pathogen of the critical tier, Serratia sp., was found in the carrots, tomatoes, bell peppers, and cabbage at the Gachororo market, as well as in the watermelons, bell peppers, cilantro, chilli peppers, and bell peppers at the Juja market. The opportunistic pathogen Providencia sp. was only found in mangoes and collard greens from Juja market. Only mangoes from the Juja market were found to contain Morganella morganii and Proteus sp.

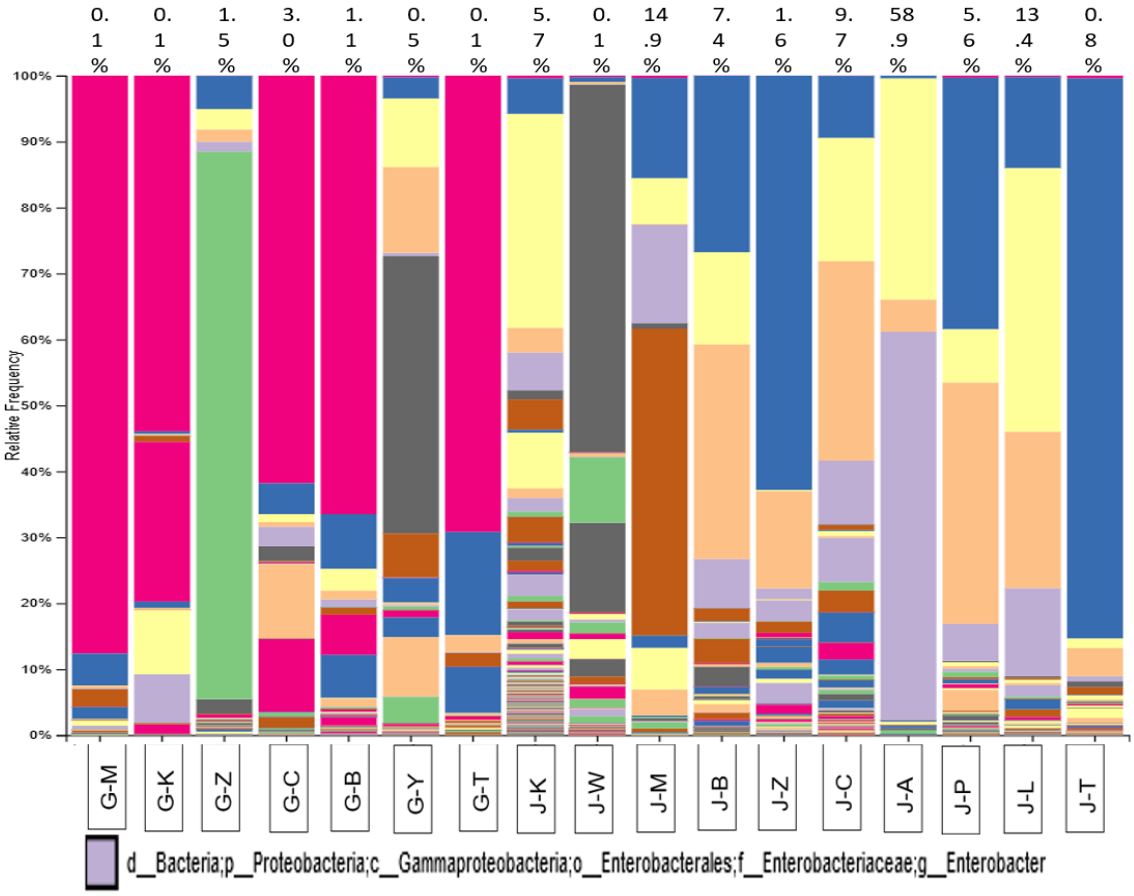


Figure 14. Taxonomic Bar Chart of Bacterial Community at Genus Level

Distribution of the top ***Enterobacter*** **sp.** across produce types in both markets.
Legend: Gachororo (G-): Mango (G-M), Collard greens (G-K), Zucchini (G-Z), Carrots (G-C), Bell peppers (G-B), Cabbage (G-Y), Tomatoes (G-T). Juja market (J-): Collard greens (J-K), Water melon (J-W), Mangoes (J-M), Bell peppers (J-B), Zucchinis (J-Z), Carrots (J-C), Tamarillo fruits (J-A), Chilli peppers (J-P), Cilantro (J-L), Tomatoes (J-L).

The phylum Proteobacteria dominated the fresh produce samples from the Juja market in the current study. This is consistent with earlier studies (Muriuki et al., 2021) that evaluated the microbial populations on fresh fruits and vegetables. However, Firmicutes predominated the fresh food sample populations from the Gachororo market. The various microbiome diversities seen between the two markets can be attributed to this dynamic change in dominant taxa. Two separate clusters of the two markets could be seen on the first axis of the PCoA charts presented using the Weighted Unifrac beta-diversity metric **(Figure 7).** However, according to the alpha diversity analysis, neither the mean phylogenetic diversity nor the species abundances of the two markets' particular microbial compositions were

significantly different from one another. Juja market samples had a mean branch length of 18.09, while Gachororo market recorded an average of 8.07, according to an analysis of Faith's phylogenetic diversity, which accounts for phylogenetic differences between species by summing the branch lengths of a phylogenetic tree connecting all species **(Figure 8).** Pielou's evenness, which examines the counts of individuals from each of the several species within a market, was the other alpha-diversity statistic that was evaluated. Gachororo samples were more evenly distributed than Juja market samples, although the difference between the two was not statistically significant ($p = 0.06$) **(Figure 9).** On the other hand, Juja market samples had an average of 329 species, whereas Gachororo market samples had 126 species ($p = 0.05$), indicating a substantial difference in the number of species within each market **(Figure 6).** According to Faith's pd and Pielou's alpha-diversity measurements, there has been a loss of species from Juja to Gachororo markets.

NGS had certain limitations in its ability to characterise microbial diversity. The primary one is the difficulty in separating sequences at even the most fundamental taxonomic levels. Salmonella paratyphi and E. coli might not have been included in the taxonomic profile because of this. The Enterobacteriaceae family, which includes both of the aforementioned species, was one of the groups responsible for the variety in the Juja market samples. According to Stecher et al. (2010), a taxon family's prominent abundance may indicate a favourable habitat for any human infections that may belong to it. Furthermore, according to statistics on the distribution of ARGs, the cluster of at least three Juja market product items in **(Figure 13)** suggests that closely related bacteria are more likely to acquire ARGs.

The research examining taxonomic indicators of cause-specific mortality risk in the human gut microbiome suggests a closer examination of the taxa dictating diversity in either market, which is arguably not the least important. According to Berg et al. (2014), the principal phyla that colonise the human gut and plants, respectively, are Proteobacteria, Firmicutes, and Actinobacteria. Proteobacteria also include a number of well-known human pathogens that have a part in both extraintestinal and intestinal disorders (Rizzatti et al., 2017). Salosensaari et al. noted a high correlation between an increase in the number of Enterobacteriaceae in the human gut and deaths from gastrointestinal and respiratory conditions. Given that eating is the most likely method of introducing novel microorganisms into the human GIT, this information seems to have significant ramifications. But to learn more about the microorganisms that settle in the gut, researchers need to conduct yet another study looking at the microbiomes of people who consume these meals. However, this suggests that evaluating the baseline microbiome is crucial in order to provide a large lead in the effort to achieve food safety.

CONCLUSION AND RECOMMENDATIONS

Conclusion: Microbiomes of fresh fruits and vegetables in open-air markets harbour human pathogenic bacteria. Salmonella paratyphi was identified using conventional techniques, and its presence on fresh produce was indicated by the faecal indicator organism Escherichia coli. Human pathogenic bacteria prioritised by the WHO, including Enterobacter sp., Acinetobacter baumannii, Serratia sp., Providencia sp., Morganella morganii and Proteus sp., were detected in fresh produce using NGS techniques. These pathogens were discovered in fresh produce at the point of sale. Furthermore, ARGs providing bacterial resistance against antibiotics were also detected. It was also observed that bacterial diversity varied by market.

Recommendation: The complexity of the sample to be investigated and the level of bacterial taxonomic detail required determine the best sequencing strategy for studying various food matrices. The microbial composition of a food sample could be broadly outlined by an initial 16S rDNA sequencing-based profile. However, this method is unable to give species- or strain-level identification due to its lack of resolution. Additionally, it won't offer an evaluation of these organisms' functional capacity within the sample. Metagenomics and metatranscriptomics would therefore be helpful for detailed species-level, strain-level, or functional characterisation of the many members of the microbiome.

REFERENCES

As., A., Robin, T., & Tennant, A. (2015). Biochemical and molecular characterisation of pyrene and anthracene metabolising bacteria isolated from oil contaminated water and soil in Malaysia. *Journal of Applied & Environmental Microbiology*, *3*, 25–30. https://doi.org/10.12691/jaem-3-1-5

Ackers, M. L., B. E. Mahon, E. Leahy, B. Goode, T. Damrow, P. S. Hayes, W. F. Bibb, D. H. Rice, T. J. Barrett, L. Hutwagner, P. M. Griffin, & L. Slutsker. (1998). An outbreak of Escherichia coli O157:H7 infections associated with leaf lettuce consumption. *J. Infect. Dis.*177:1588–1593

Alegbeleye, O. O., & Sant'Ana, A. S. (2020). Pathogen subtyping tools for risk assessment and management of produce-borne outbreaks. *Current Opinion in Food Science*, *32*, 83–89. https://doi.org/10.1016/J.COFS.2020.02.007

Asokan, G. V., Ramadhan, T., Ahmed, E., & Sanad, H. (2019). WHO global priority pathogens list: A bibliometric analysis of medline-pubmed for knowledge mobilisation to infection prevention and control practices in Bahrain. *Oman Medical Journal*, *34*(3), 184–193. https://doi.org/10.5001/omj.2019.37

Asokan, G. V., Ramadhan, T., Ahmed, E., & Sanad, H. (2019). WHO global priority pathogens list: A bibliometric analysis of medline-pubmed for knowledge mobilisation to infection prevention and control practices in Bahrain. *Oman Medical Journal*, *34*(3), 184. https://doi.org/10.5001/OMJ.2019.37

Bajpai, T., Pandey, M., Varma, M., & Bhatambare, G. S. (2017). Prevalence of TEM, SHV, and CTX-M Beta-Lactamase genes in the urinary isolates of a tertiary care hospital. *Avicenna Journal of Medicine*, 7(1), 12. https://doi.org/10.4103/2231-0770.197508

Barrera, M. J., Blenkinsop, R., & Warriner, K. (2012). The effect of different processing parameters on the efficacy of commercial post-harvest washing of minimally processed spinach and shredded lettuce. *Food Control*, *25*(2), 745-751.

Bengtsson-Palme, J., Kristiansson, E., & Larsson, D. J. (2018). Environmental factors influencing the development and spread of antibiotic resistance. *FEMS microbiology reviews*, *42*(1), fux053

Bengtsson-Palme, J., & Larsson, D. J. (2016). Concentrations of antibiotics predicted to select for resistant bacteria: proposed limits for environmental regulation. *Environment International*, 86, 140-149.

Berg, G., Erlacher, A., Smalla, K., & Krause, R. (2014). Vegetable microbiomes: Is there a connection among opportunistic infections, human health and our "gut feeling"? *Microbial Biotechnology*, *7*(6), 487–495. https://doi.org/10.1111/1751-7915.12159

Beuchat, L. R. (1999). Survival of enterohemorrhagic Escherichia coli O157:H7 in bovine feces applied to lettuce and the effectiveness of chlorinated water as a disinfectant. *J. Food Prot*, *62*, 845-849.

Bhullar, K., Waglechner, N., Pawlowski, A., Koteva, K., Banks, E. D., Johnston, M. D., ... & Wright, G. D. (2012). Antibiotic resistance is prevalent in an isolated cave microbiome. *PloS one*, *7*(4), e34953.

Bielen, A., Šimatović, A., Kosić-Vukšić, J., Senta, I., Ahel, M., Babić, S., & Udiković-Kolić, N. (2017). Negative environmental impacts of antibiotic-contaminated effluents from pharmaceutical industries. *Water Research*, *126*, 79-87.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Byrne, L., Fisher, I., Peters, T., Mather, A., Thomson, N., Rosner, B., ... & on behalf of the International Outbreak Control Team, C. (2014). A multi-country outbreak of Salmonella Newport gastroenteritis in Europe associated with watermelon from Brazil, confirmed by whole genome sequencing: October 2011 to January 2012. *Eurosurveillance*, *19*(31).

Cao, Y., Fanning, S., Proos, S., Jordan, K., & Srikumar, S. (2017). A review on the applications of next-generation sequencing technologies as applied to food-related microbiome studies. *Frontiers in*

microbiology, *8*, 1829.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, *108*(Supplement), 4516-4522.

Chalmers, R. M., H. Aird, and F. J. Bolton. (2000). Waterborne Escherichia coli O157. *J. Appl. Microbiol.88*, 124S–132S.

Chen, Z., Hui, P. C., Hui, M., Yeoh, Y. K., Wong, P. Y., Chan, M. C. W., Wong, M. C. S., Ng, S. C., Chan, F. K. L., & Chan, P. K. S. (2019). Impact of Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling. MSystems, 4(1). https://doi.org/10.1128/MSYSTEMS.00271-18/SUPPL_FILE/MSYSTEMS.00271-18-S0001.PDF

Cholera – Kenya. (n.d.). Retrieved on September 5, 2022, from https://www.who.int/emergencies/disease-outbreak-news/item/11-december-2017-cholera-kenya-en