Bacterial biomes and potential human pathogens in irrigation water and leafy greens from different production systems described using pyrosequencing

M Jongman, L Chidamba and L Korsten*

Department of Science and Technology/ National Research Foundation: Centre of Excellence Food Security and Department of Plant and Soil Sciences, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa.

***Corresponding author.** Mailing address: Department of Plant and Soil Sciences, University of Pretoria, South Africa. Phone: +27 12 420 3295. Fax: +27 12 420 4588. Electronic mail address: lise.korsten@up.ac.za.

Abstract

Aims: To investigate the influence of irrigation water microbial quality on leafy green vegetables produced in commercial and small-scale farms as well as homestead gardens using pyrosequencing.

Methods and results: Next generation sequencing analysis of the V1-V3 hypervariable region of bacterial 16S rDNA was used to compare bacterial diversity in irrigation water sources and on leafy vegetables. In all samples (12) analyzed, the phylum *Proteobacteria* (64.5%), class *Gammaproteobacteria* (56.6%) and genus *Aeromonas* (14.4%) were found to be dominant. Of the total *Escherichia* sequences detected in tested samples, lettuce (16.3%) from the one commercial farm harbored more sequences than cabbage from the small-scale farm (1.3%) or homestead gardens (1.9%). *Escherichia* sequences were detected in both irrigation water (4.6%) and on cabbage (1.3%) samples from the small-scale farm. The genus *Salmonella* was absent in borehole water but was detected in the holding dam water (<1%)

from commercial farm A. *Salmonella* sequences were present in river water (<1%) and on cabbages (1.9%) from the small-scale farm but were not detected on cabbage samples from the one commercial farm or the homestead gardens.

Conclusion: Water sources quality used for irrigation greatly influences the microbial dynamics of the irrigated crop.

Significance and Impact of the Study: Microbial biomes in irrigation water and on leafy greens were described with pyrosequencing and revealed insights into prevalence of potential and opportunistic pathogens across different production systems.

Keywords: Microbial biomes, Foodborne pathogens, Irrigation water, Fresh leafy greens, Next generation sequencing.

Introduction

The consumption of fresh leafy green vegetables has increased over the past few years, owing to well recognized health benefits (Ceuppens *et al*. 2014). However associated with this increase are food borne disease outbreaks, linked to the use of polluted irrigation water (Pachepsky *et al*. 2011; Allende and Monaghan 2015). Most of these reported epidemics have been attributed to pathogenic *Salmonella* and *Escherichia coli* (Nygard *et al*. 2008, Ceuppens *et al*. 2014). In particular, cabbage, lettuce and spinach have been associated with most of the foodborne disease outbreaks (Smith De Waal and Bhuiya 2007; Nygard *et al*. 2008; Ceuppens *et al*. 2014; Park and Kang 2015). According to Teplitski *et al*. (2009) these recent increases in human-related disease outbreaks is mainly due to adaptation and persistence of enteric pathogens on plant surfaces. The phyllosphere is therefore of fundamental importance to food safety and research on the microbial ecology of fresh produce contribute towards a better understanding of population dynamics and intervention strategies (Telias *et al*. 2011).

The presence of potential human pathogens in irrigation water and on the phyllosphere is a cause of concern (Pachepsky *et al*. 2011; Ceuppens *et al*. 2014). Consequently, guidelines based on good agricultural practices (GAP) aiming at improving food safety have recommended potable water for irrigation of vegetables consumed raw (Pachepsky *et al*. 2011). However, the scarcity of potable irrigation water is a persistent challenge in the agricultural sector. Hence, farmers may resort to use of water sources with compromised quality (Gemmell and Schmidt 2012). Water sources become contaminated when exposed to improperly treated municipal sewerage discharged into the main water ways or, untreated animal or human feces being washed into rivers during rain events. The link between microbial quality of irrigation water and safety of leafy green vegetables has been established in recent studies (Du Plessis *et al*. 2015).

Underestimation of microbial populations using culture dependent methods led to the use of next-generation sequencing technologies, including 454-pyrosequencing. This platform provides a more holistic account of microbial communities in diverse environments due to the amplified number of sequence reads obtained (Telias *et al*. 2011). Advantages of pyrosequencing include rapidity and high flexibility (Lim *et al*. 2010). Pyrosequencing technology has been used to explore soil microbial populations, exploring and quantifying fungal diversity in freshwater lake ecosystems (Monchy *et al*. 2011) and bacterial communities in spray water and on tomato surfaces (Telias *et al*. 2011).

Leafy green vegetables are often referred to as lettuce, spinach and cabbages commonly grown in commercial and small-scale farming systems or in homestead gardens. However, studies on the microbial quality of irrigation water and fresh produce have mostly been

focused on commercial production systems (Ceuppens *et al*. 2014; Du Plessis *et al*. 2015), while little attention has been paid to small-scale farming systems or homestead gardens (Speelman *et al*. 2008; Erickson *et al*. 2013). A previous study by Jongman and Korsten (2016a) explored the microbial quality of leafy greens from different production systems using culture based techniques. To the best of our knowledge, we are not aware of a study comparing the three different production systems using pyrosequencing as a novel tool to assess source tracking potential of waterborne pathogens and bacterial population dynamics. We there therefore used next Generation sequencing (NGS) and analysis of the V1-V3 hypervariable of the bacterial 16S rDNA to evaluate the microbiological quality and to identify potential foodborne and opportunistic pathogens in irrigation water sources and the associated leafy green vegetables from homestead gardens, small-scale and commercial farm production systems.

Material and Methods

Site description, sample collection and processing

A total of 12 representative samples from irrigation water sources (6) and leafy greens (6) (3 pooled samples each made up of 3 leaves per sample) from formal and informal vegetable production systems in South Africa (Table 1) were selected for pyrosequencing. Commercial farm A was Global G.A.P certified and grew cabbage, baby spinach and lettuce for a major retailer. The water source was borehole water which was first pumped into a temporary reservoir or holding dam before spray irrigation using a central pivot system. The other sites were not Global G.A.P certified and grew cabbages only. The sites included a commercial farm (B) irrigating the crop with river water applied as a spray, a small-scale farm using river water as water source applied through hose pipes and homestead gardens using river or

ground harvested rainwater (GHRW) using buckets. Water samples were collected and processed as previously described (Chidamba and Korsten 2015) with minor modifications. Water samples (750 ml) was concentrated through cellulose nitrate filters (0.45-mm pore size; Sartorius, Gottingen, Germany) while fresh baby spinach, cabbage and lettuce (25g) were suspended in sterile peptone buffered water (PBW) (225g) and stomached. The micro-floral wash was then centrifuged at 10,000 × g and the pellet stored at -80°C until DNA extraction was performed.

Table 1 Description of sampling sites showing possible pollution sources, irrigation water and types of leafy greens grown at different agricultural settings.

Site	Size	Certification	Possible pollution source (s)	Irrigation water source (s)	Crop (s)
Commercial farm A	550	Global G.A.P. certified	Poultry & horse farms	Borehole, Holding dam	Baby spinach Cabbage Lettuce
Commercial farm B	200	None	Informal settlement	River	Cabbage
Small-scale farm	150M^2	None	Waste water treatment plant	River	Cabbage
Homestead gardens	1	None	Informal settlement	River Ground harvested rainwater	Cabbage

[a] Values are in hectares unless otherwise stated. Ha: hectares. M^2: Square meters. 1 Ha represents an average of 25 m2 at 40 households.

DNA extraction and pyrosequencing

Genomic DNA (gDNA) was extracted from the filter papers and the pelleted vegetable samples using ZM fungal/bacterial DNA miniprep™ kit (Zymo Research Corporation, USA) as per manufacture's specifications, and concentration determined with the Qubit 2.0 Fluorometer (Lifescience Technology, Johannesburg). Extracted gDNA was stored at -20 °C. DNA samples were sent to Inqaba Labs (Pretoria, South Africa) for Illumina MiSeq sequencing targeting the V1-V3 hypervariable region of the bacterial 16S rDNA using universal bacterial primer set 27 F (5'-AGAGTTTGATCMTGGCTCAG-3') (Weisburg *et al*. 1991) and 518R (5'-ATTACCGCGGCTGCTGG-3') (Muyzer *et al*. 1993).

Pyrosequencing data processing and analysis

Sequences that were of lower quality or shorter than 150bp in length were removed from the data sets using the RDP pyrosequencing pipeline initial processing (http://pyro.cme.msu.edu) (Cole *et al*. 2009). The RDP classifier was used to manually remove nonspecific or unexpected bacterial and archaeal reads (Cole *et al*. 2013). Operational taxonomic units (OTU) and rarefaction curves were generated at 3%, dissimilarity level using the RDP pyrosequencing pipeline (Hur and Chun 2004). The OTU table was normalized by rarefaction to an even sequencing depth in order to remove sample heterogeneity. The OUT table was uploaded into the Visualization and Analysis of Microbial Population Structures website (VAMPS) (http://vamps.mbl.edu) (Huse *et al*. 2014) from which relative abundance, and alpha diversity indices including Observed Species (Sobs), Chao1(Chao 1987), Shannon-Weaver (Shannon and Weaver 1963) and evenness were calculated. To analyse the unique detected OUT in both water and crop samples, we performed a Venn diagram analysis using an on-line tool (http://bioinfogp.cnb.csic.es/tools/venny/). Raw counts of observed species were used to analyses microbial diversity after normalizing by sequencing depth in MEGAN 5.10.5 (Huson 2016). Bray–Curtis distances were plotted as principal coordinate analysis (PCoA) (Bray and Curtis 1957). Agglomerative clustering analyses were performed with Bray–Curtis distances as input and the UPGMA clustering method specified.

Results

Characteristics of the sequenced data

Pyrosequencing, processing and analysis of the 12 samples generated a total of 66519 bacteria sequences which were assigned to 401 OTUs. The number of sequences in the

collected samples varied between 13721 and 1133 sequences, while the number of OTUs varied between 39 and 295 OTUs (Table 2). Analysis of a rarefied OTU table to an even depth of 1000 reads per sample revealed a higher number of OTUs in GHRW followed by river water (both from homestead gardens) and river water from the small-scale farm, while cabbage samples from commercial farm B had the lowest number of OTUs detected. The Chao1, species evenness estimator and Shannon diversity indices indicated high diversity and evenness for all samples with minimum values of 23.25 and 1.36. The most diverse bacterial populations were found in GHRW irrigation water samples from homestead gardens, while the lowest was observed in cabbage from commercial arm A. On farm holding dam samples had more diversity than the borehole samples at commercial farm A (Table 2).

Table 2 Alpha diversity parameters of bacterial communities from source water and leafy green vegetable samples from different production systems.

Sample source	Sample type	Sampling depth	Total OTUs	OTUs 1000[a]	CHAO	Shannon-Weaver Diversity Index
Commercial farm A	Borehole water	13721	41	16	44	1.4
	Cabbage	9129	50	24	42	1.11
	Holding dam water	1133	42	38	41	2.32
	Lettuce	5409	35	21	23.25	1.88
	Spinach	6671	50	30	55	1.67
Commercial farm B	Cabbage	1794	27	21	33.5	1.43
	Pond water	5148	67	44	66.5	2.54
Small Scale Farm	Cabbage	3562	39	32	47.13	1.36
	River water	2090	90	64	72.45	3.15
Homestead gardens	Cabbage	6701	45	22	43	1.8
	GHRW[b]	8496	295	114	150.1	3.48
	River water	2665	120	80	90	3.13

[a] OTUs determined at an even depth of 1000 sequences
[b] GHRW Ground harvested rain water

Phylum Level Diversity

At phylum level, 80% of the sequences were classified into 42 distinct bacterial phyla (Figure 1 and 2). Unclassified sequence reads were excluded from the data set. *Proteobacteria* were the most dominant in the data set (64.5%), with holding dam water from a commercial farm accounting for 28.4% of the total reads in the phyla. Cabbage sampled from homestead

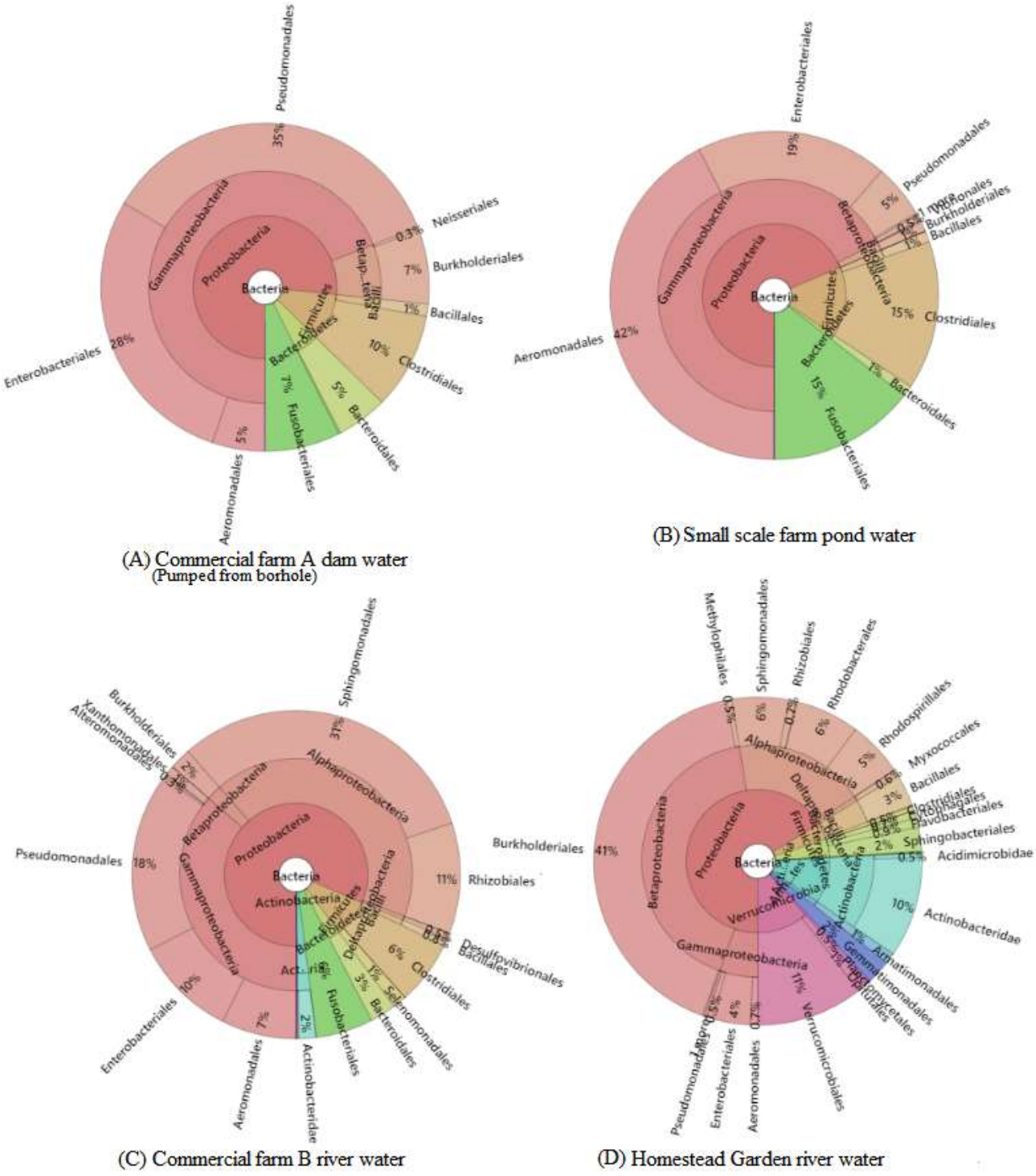

Figure 1 Graphical representation of the relative abundance of bacterial diversity from phylum to order level of irrigation water samples from various production systems visualized using Krona visualization tool.

gardens had the most diversity (*n*=30 phyla) while the least was observed in cabbage from a small-scale farm (*n*=7 phyla). The phylum *Firmicutes* (12.2%) was the second most

dominance and it was detected in all water and vegetable samples. The phylum was detected at 33.9% in cabbage from homestead gardens but at <1% in cabbage from commercial farm B. *Planctomycetes* were detected across all tested samples and were mostly dominant in GHRW (42.7%). Other phyla with interesting distribution include *Firmucutes* (63.6%) and *Proteobacteria* (32.2%) in cabbage from small-scale farm, while, *Gemmatimonadetes* was only detected in river (1.2%) and GHRW (<1%) (Figure 1).

Class Level Diversity

A total of 61 taxonomic classes were obtained at class level and 10% of the sequences were unclassified and consequently omitted (Table 3). Four classes accounted for over 76% of identified taxonomic groupings: *Gammaproteobacteria* (73.4%), *Bacilli* (7.5%), *Clostridia* (4.4%) and *Alphaproteobacteria* (4.1%). These four classes dominated the data set in all the tested irrigation water sources and leafy green vegetables. Cabbage samples from commercial farm B was dominated by sequences belonging to *Gammaproteobacteria* (99.6%); and was the highest observed of all other samples. Overall, *Gammaproteobacteria* were identified in both irrigation water and on leafy greens from commercial (76.9 and 36.1%), small-scale (84.4 and 82.1%) farms, and homestead gardens (84 and 71.2%). *Gammaproteobacteria* was detected on baby spinach (6.8%) and lettuce (4.3%) from commercial farm A (Figure 2). Although *Flaviobacteria* accounted for 2% of the total sequences, 19.7% of the reads were from spinach samples from commercial farm A where the taxonomic grouping represented the highest detected classifications within the tested samples. In all cabbage samples, commercial farms had more sequence reads belonging to *Bacilli* (65%) than small-scale (8.4%) and the homestead gardens (<1%). Other classes detected include *Betaproteobacteria* (3.2%), *Sphingobacteriia* and *Planctomycetia* (<1%) (Figure 2 and 3).

Table 3 Average relative taxonomic abundance profile at genus level of classification for bacterial sequences in irrigation water and on leafy green vegetable from various production systems.

	Commercial farm A					Commercial farm B		Small-scale farm		Homestead gardens				
	Water source		Crop							Water Source		Crop		
Bacterial genera	Borehole	Holding dam	Cabbage	Spinach	Lettuce	Pond water	Cabbage	River water	Cabbage	GHRW	River	Cabbage	Total seqs	%
Aeromonas	93,60	6,72	0,27	14,69	0	47,60	-	12,80	-	0,36	0,62	0,13	6912	38,74
Exiguobacterium	0,03	0,92	0,67	0	0,08	0,21	0,32	-	1,04	0,98	2,10	90,09	2084	11,68
Serratia	0,03	0,31	0,38	4,15	52,47	-	-	-	-	0	0,13	1,02	714	4
Cetobacterium	-	12,88	0,19	8,07	-	18,69	-	6,01	-	0,06	0,13	-	596	3,34
Pantoea	-	-	33,11	24,67	0,08	0,04	-	-	6,65	0,06	0,13	-	588	3,30
Clostridium XI	-	13,80	4,91	0,12	1,70	13,35	-	2,79	5,41	2,07	0,13	-	515	2,89
Pseudomonas	0,02	41,10	0,10	8,54	0,08	0,64	0,32	0,73	19,33	1,04	-	1,73	380	2,13
Plesiomonas	3,77	0,92	0,10	-	-	4,49	-	2,93	-	0,23	-	-	353	1,98
Escherichia/Shigella	0,21	-	-	1,90	16,28	0,68	-	4,55	1,25	-	0,13	1,91	326	1,83
Anaerosporobacter	-	-	15,98	18,86	-	-	-	-	-	-	-	-	325	1,82
Acinetobacter	0,15	0,31	0,19	0,47	-	4,83	22,90	9,53	-	3,05	0,13	-	320	1,79
Flavobacterium	-	-	-	-	-	-	-	-	-	18,01	0,52	-	317	1,78
Enterococcus	0,02	0,31	-	-	1,30	-	-	0,15	55,09	-	0,13	0,27	291	1,63
Sphingomonas	0,03	-	0,10	-	-	-	0,65	27,42	-	4,78	1,44	0,13	289	1,62
Turicibacter	0,03	-	-	-	18,87	-	-	-	-	1,15	-	-	255	1,43
Limnohabitans	0,02	-	-	-	-	-	-	-	-	2,19	27,13	-	246	1,38
GpIIa	-	-	-	-	-	-	-	-	-	-	30,14	-	230	1,29
Clostridium sensu stricto	0,02	-	7,99	1,19	4,78	0,08	1,29	0,44	-	2,93	-	-	213	1,19
Leclercia	-	-	16,07	0,12	-	-	-	-	2,08	-	-	-	178	1
Klebsiella	0,02	-	-	-	-	0,51	48,39	0,44	0,62	-	-	-	169	0,95
Morganella	1,15	1,84	0,10	0,12	-	0,76	-	0,88	-	-	0,13	0,40	109	0,61
Raoultella	0,07	0,31	-	-	-	0,93	22,90	-	1,04	-	0,13	-	104	0,58
Novosphingobium	-	-	-	-	-	-	-	-	-	5,47	0,39	-	98	0,55
Porphyrobacter	-	-	-	-	-	-	-	-	-	5,12	0,52	-	93	0,52
Enterobacter	-	0,31	7,31	0,24	-	-	-	-	1,87	-	-	0,04	89	0,50
Pluralibacter	0,03	-	0,19	0,83	0,89	0,13	-	0,29	-	-	0,13	2,49	84	0,47
Bacillus	-	-	1,06	-	2,75	0,72	1,94	0,29	-	0,40	0,13	0,13	81	0,45
Curvibacter	-	-	-	-	-	-	-	-	-	4,26	0,26	-	76	0,43
Luteolibacter	-	-	-	-	-	-	-	-	-	0,17	8,65	-	69	0,39
Parcubacteria	-	-	-	-	-	-	-	-	-	3,80	0,26	-	68	0,38
Yokenella	-	0,31	5,97	0,12	-	-	0,32	-	-	-	-	-	65	0,36
Bradyrhizobium	-	-	-	-	-	-	-	5,57	-	0,58	0,26	-	50	0,28
Others (<0.3%)	0,41	18,71	5,29	13,88	0,73	4,28	0,97	25,22	5,61	43,15	26,08	1,64	1555	8,72
Total seqs	1039	843	2250	1738	763	682	5816	481	1235	310	2359	326	17842	

GHRW: Ground harvested rainwater, seqs: sequences.

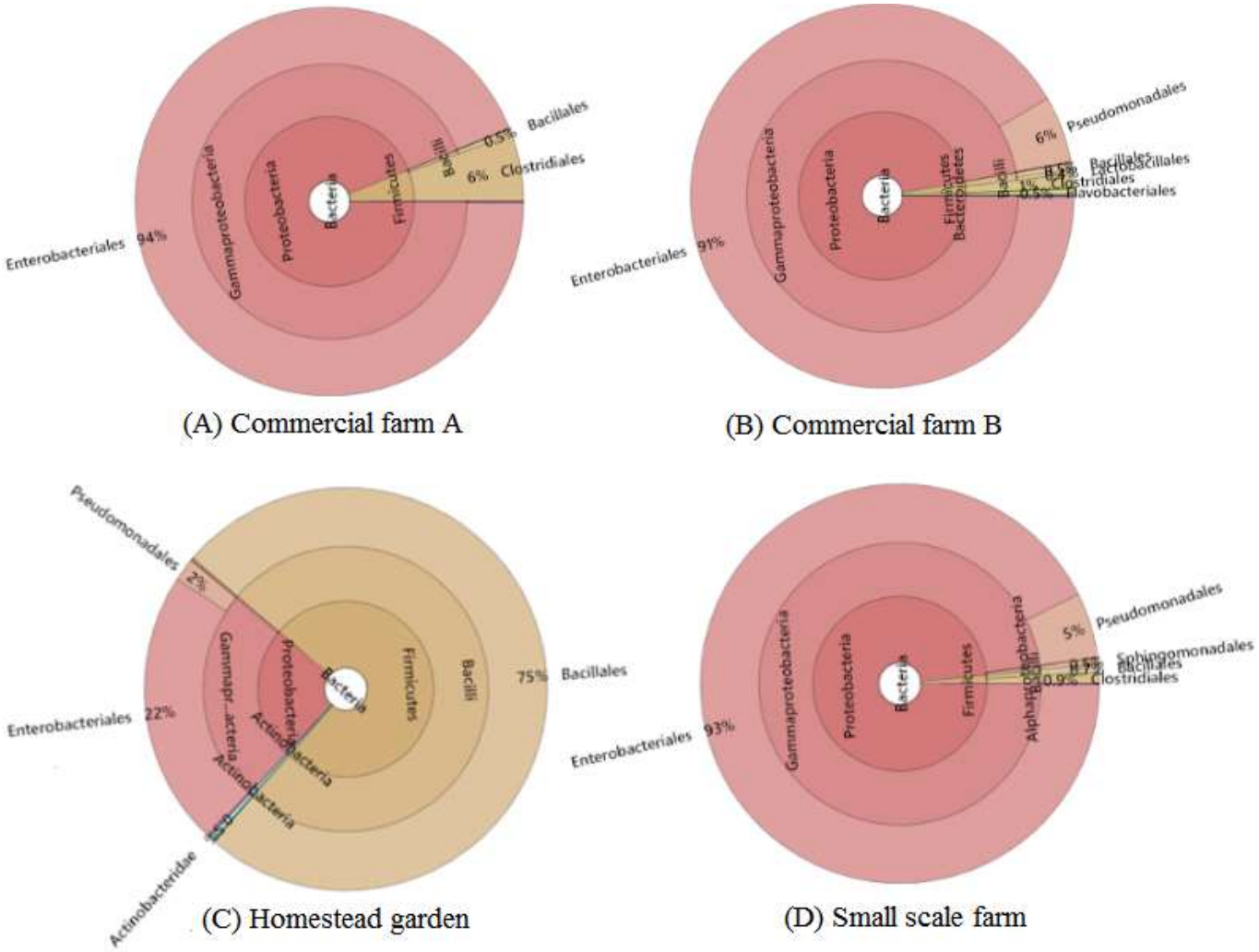

Figure 2 Graphical representation of the relative abundance of bacterial diversity from phylum to order level of cabbage samples from various production systems visualized using Krona visualization tool.

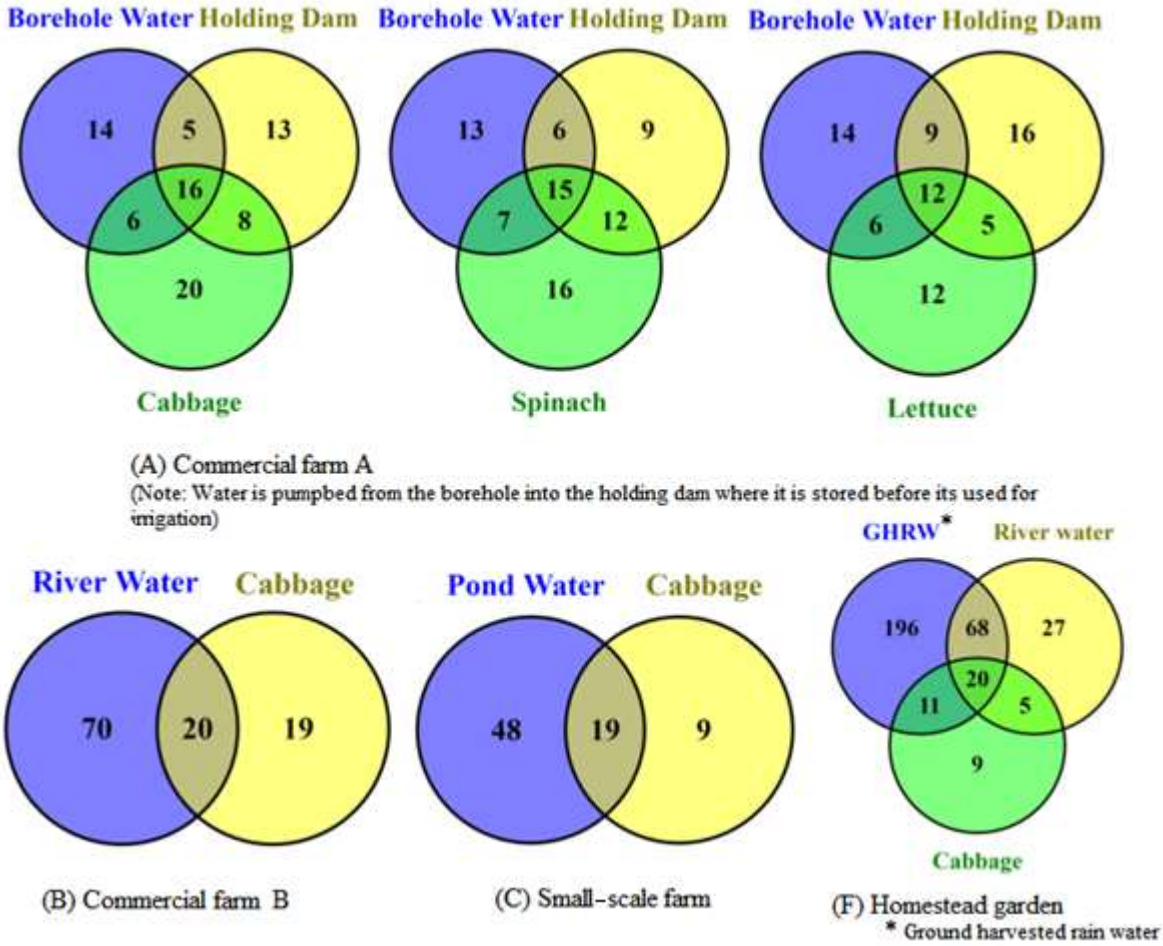

Figure 3 Graphical representation comparing shared and unique operational taxonomic units (OTUs) (depicted by numbers) in various irrigation water sources and leafy green vegetables from various production systems visualized using Krona visualization tool.

Genus Level Diversity

At genus level, variations amongst the microbial populations of various irrigation water sources and leafy green vegetables were detected (Table 3). Sequences belonging to *Aeromonas* (38.7%) and *Exiguobacterium* (11.7%) dominated the data set. The genus *Aeromonas* was detected in Borehole (93.6%) and pond (47.6%) water samples from commercial farms A and B, respectively. The genus was also detected in river water (12.8%) but not on cabbage samples at the small-scale farm. Other genera detected at ≤2% in irrigation water and on leafy greens across all sites include *Bacillus*, *Enterococcus*,

Enterobacter, *Klebsiella* and *Pseudomonas* (Table 3). *Escherichia*/*Shigella* spp. (<2%) were detected in irrigation water sources and leafy green vegetables (Table 3). Of the total *Escherichia* sequences detected in tested samples, lettuce (16.3%) from commercial farm A harbored more than cabbage from small-scale (1.3%) and homestead gardens (1.9%). *Escherichia* sequences were detected in both irrigation water (4.6%) and on cabbage (1.3%) samples from small-scale but were detected only in source water (Pond, <1%) and not on cabbage at commercial farm B (Table 3). The detection of sequences belonging to the genus *Escherichia* is consistent with our previous results where isolates were recovered using conventional culture methods (Jongman and Korsten, 2016a). *Shigella* sequences in the various irrigation water and leafy greens were detected at relatively the same levels as *Escherichia*.

Some sequences detected below 1% have been grouped together (Table 3). These include *Salmonella*. The genus *Salmonella* was absent in borehole water but was detected in holding dam water (<1%) from commercial farm A. *Salmonella* sequences were present in river water (<1%) and cabbage (1.9%) from the small-scale farm. Sequences of the genus *Salmonella* were not detected on cabbage samples from commercial farm B and homestead gardens. *Salmonella* was also present in spinach and cabbage samples (<2%) from commercial farm A.

In contrast, *Salmonella* isolates were not recovered using traditional culture methods in our previous study (Jongman and Korsten, 2016a). A comparison of shared OTUs in commercial farm A showed significant differences with the water after storage in the holding dam (Figure 3). More OTUs were shared between the dam water and crops then between crops and

borehole source water. Comparison between GHRW and river water used in a homestead garden showed GHRW to have a wide diversity compared to river water and this high diversity appear to influence the observed microbial diversity on irrigated cabbage. In all samples unique OUTs were observed both in the irrigation water and on crop samples (Figure 3).

Cluster analysis

The phylogenetic diversity comparison of the different samples is shown in the UPGMA cluster dendrogram and PCoA (Figure 4). Marked clustering differences can be observed with crop samples falling into three distinct groups i) Commercial farm A (cabbage and spinach) and Small-scale farm cabbage; ii) homestead gardens cabbage and iii) Commercial farm B cabbage. Although borehole water was pumped in the holding dam, the two were grouped separately. Interestingly the holding dam water samples were grouped together with the irrigated cabbage and spinach, although the same could not be said for spinach. A similar observation was made with cabbage and GHRW from homestead gardens which clustered together. River water samples from Commercial (A and B) and small-scale farms also clustered together away from the irrigated produce and holding dam water as well as GHWR.

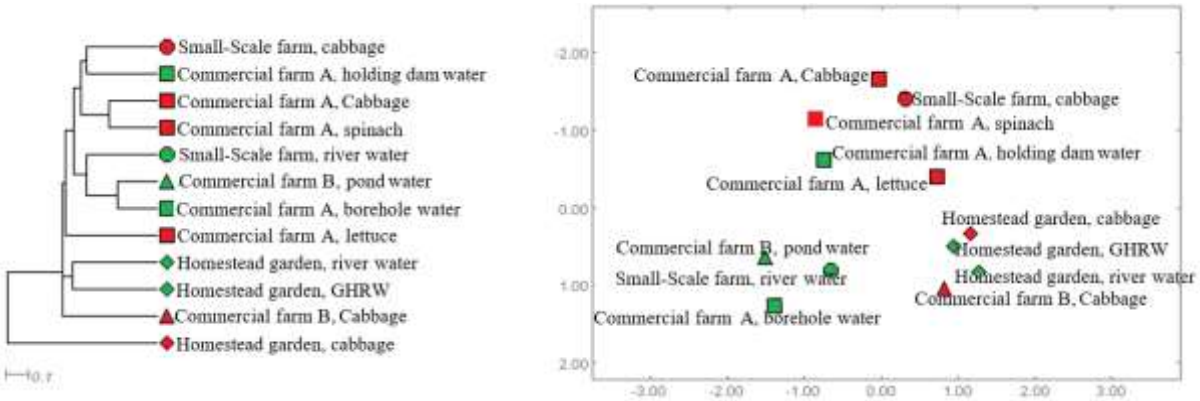

Figure 4 Principal coordinate analysis of UPGMA visualisations of similarities amongst irrigation water sources and leafy greens from different production systems. Plots were made using Bray Curtis ecological index.

Discussion

Studies previously conducted by the authors have shown irrigation water microbial quality to greatly influence observed microflora on associated leafy greens (Jongman and Korsten, 2016a). However, traditional culture methods were used in the study and have the disadvantage that they focus on individual or closely related species, rather that the whole microbiome (Jongman and Korsten, 2016b). Pyrosequencing evaluation of the V1–V3 hypervariable regions of the 16S RNA to investigate microbial communities in both water and irrigated produce revealed interesting findings. These included variations in microbial quality of irrigation water from different environmental settings and associated leafy greens from different production systems. To our knowledge this is the first study to use NGS to investigate various irrigation water sources (river, pond, dam, borehole and GHRW) and the irrigated crops i.e. leafy green vegetables (baby spinach, cabbage and lettuce) across different production systems (commercial and small-scale farms, and homestead gardens).

The observed dominance of *Proteobacteria* (64.5%) *Firmicutes* (12.2%) and *Bacteroidetes* (4.9%) in all samples, with the exception of cabbage from homestead garden, is similar to reports by Telias *et al*. (2011), although the phyla *Actinobacteria* were more prevalent than *Bacteriodetes* in their findings. Microbial communities in cabbage samples from homestead garden were dominated by the phylum firmicutes (75%) although the phylum was observed at less than 18% in all other samples. Interestingly, the irrigation water used at homestead gardens had comparatively higher diversity, however, *Proteobacteria* was the dominant phylum not *Firmicutes*. The observed variation in the dominance of firmicutes on cabbage samples may suggest the influence of environmental or inhabiting microbial communities in promoting its dominance. Generally *Proteobacteria* have been reported to be the main component of the leaf endophytic community on field grown leafy green vegetables (Dees *et*

al. 2015). Our study reports the bacterial communities in holding dam water (31%), cabbage (19%), spinach (11%) and lettuce (9.9%) to be dominated by *Gammaproteobacteria*, results which have also been noted by previous studies (Leff and Fierer, 2013).

Taxonomic sequences belonging to genera of known foodborne pathogens such as *Escherichia* and *Shigella*, *Salmonella* were detected in irrigation water sources and leafy green vegetables using pyrosequencing. The detection of *Salmonella* in the holding dam but not in borehole source water can be explained by the observation made by Telias *et al*. (2011) who stated that the sustenance of enteric pathogens is less likely in groundwater sources owing to soil filtering mechanisms, however poorly managed open water sources were vulnerable to pollution. In their study on tomatoes Telias *et al*. (2011) detected high levels of *Enterobacteriaceae*, contrary to our findings and they could not confirm the presence of *Salmonella* sequences. Our study found that *Enterobacteriaceae* sequences accounted for 34.6% of all taxonomic classifications, similar to levels reported by Leff and Fierer (2013). Of all *Salmonella* sequences detected, 51% were on cabbage samples from the small-scale farm. The absence of *Salmonella* in the source water (river) and the irrigated cabbages from homestead gardens suggests the product to be less risky than from the farming systems using source water containing *Salmonella* sequences.

The detection of taxonomic signatures belonging to *Escherichia*, in irrigation water and leafy greens is a cause of concern since these crops are mostly consumed raw. Several strains belonging to this genus, including *E. coli* O157: H7 (Park and Kang 2015), have been reported to be the leading cause of foodborne illnesses associated with fresh and ready-to-eat vegetables such as cabbage and lettuce (Smith De Waal and Bhuiya 2007). Although *Listeria*

is one of the major concerns in food borne illnesses, no signatures of the pathogen were detected across all samples, similar to previous studies in source water (Dobrowsky *et al*. 2014), spinach and lettuce (Jackson *et al*. 2013). However, findings are similar to those by Jackson *et al*. (2013), Ahmed *et al*. (2014) and Dobrowsky *et al*. (2014) in revealing the presence of *Salmonella*, *Shigella*, *Enterococcus*, *Bacillus*, *Pseudomonas* and *Aeromonas* in water. Despite benefits of recent advances in rapid detection methods, presence of nucleic acids belonging to pathogenic microorganism(s) does not imply the presence of live organism(s) (Ceuppens *et al*. 2014). However, Muller and Ruppel (2014) showed that 0.1-50% of the total bacterial community detected with culture-independent molecular methods were cultivable. Therefore, although it is true that the presence of nucleic sequences belonging to pathogens does not constitute a food safety risk (Ceuppens *et al*. 2014), findings of Muller and Ruppel (2014) suggest a potential hazard.

Phyllosphere bacteria are capable of intricate associations with human pathogens thus potentially affecting fresh produce safety (Lopez-Velasco *et al*. 2013). Some *Pantoea* spp. have significant roles in promoting plant growth in the phyllosphere (Leff and Fierer 2013). Similar to the results by Leff and Fierer (2013), our study found that sequences belonging to the genus *Pantoea* had a high relative abundance in samples that also had a high relative prevalence of *Enterobacteriaceae*.

Sequences of the genus *Pseudomonas* dominated bacterial populations in holding dam water (41%) and cabbage (19%) samples from commercial and small-scale systems, respectively. This finding was similar to other studies by Jackson *et al*. (2013) and Leonard *et al*. (2015) on spinach, although they detected the genera at much higher levels. Phylotypes belonging to

the genus *Pseudomonas* were dominant on spinach compared to other leafy greens, similar to trends reported by Jackson *et al*. (2013). Sequences of the genus *Legionella* were detected only in GHRW and river water samples from homestead gardens. This concurs with results of a previous study (Navarro-Noya *et al.* 2013). High temperatures, combined with presence of biofilm pioneers such as *Pseudomonas* (detected in all tested samples except river water from homestead gardens in this study) and *Klebsiella* (in all water sources at commercial farms) may lead to proliferation of *Legionella* thus posing a potential human health risk owing to the Legionnaires disease (Wingender and Flemming 2011; Navarro-Noya *et al.* 2013).

Phylotypes belonging to the genus *Arcobacter* were detected in river water from homestead gardens. This result is similar to that of Hausdorf *et al*. (2013). However, Hausdorf *et al*. (2013) also detected *Arcobacter* in spinach samples, contrasting with our results. The infectious dose of pathogenic *Arcobacter* strains is unknown. However, if the infectious dose of *Acrobacter* is similar to their phylogenetic close relative, *Campylobacter jejuni* (500-800 cells), and contaminated irrigation water may be a potential health hazard especially for leafy green vegetables consumed raw (Hausdorf *et al*. 2013). The detection of phyllotypes belonging to genera of known human pathogens is therefore a potential health hazard, irrespective of the production system used for leafy green vegetables.

The findings in this study show the influence of water quality and crop production systems. This is made apparent by the differences in groupings between borehole source water and the holding dam where it was initially pumped from. Moreover microbial communities on the irrigated crops were grouped away from the borehole water, but were grouped together with holding dam water. The observation of water quality variation between borehole and holding

dam water suggest caution on water handling processes as they have the potential to introduce contamination in otherwise clean water (Telias *et al*. 2011). This is important considering that holding dam water is exposed to birds and other small animals which can result in fecal contamination. Moreover biofilm formation can allow the proliferation of some bacteria such as *Legionella*. The piping used for irrigation has also been implicated in irrigation water contamination from flacking biofilms in the pipes (Van der Merwe *et al*. 2013). While irrigation water may greatly influence microbial communities on irrigated produce, it appears that there may be other processes influencing observed microbial communities. This is of particular importance in agricultural systems where animal manure is used as fertilizer, or close proximity to faecal sources (i.e. cattle kraal) where dust can be blown onto vegetable surfaces. This was a scoping study to assess the potential of novel techniques to assess the bacterial footprint in different production systems. A more in depth study should follow that determine level of variation within systems.

Water sources quality used for irrigation greatly influences the microbial dynamics of the irrigated crop. Although borehole water may be used as a premium water source, storing this water in a holding dam can affect the quality and may result in the introduction of potential Pathogens such as *E. coli* and *Salmonella*. Using novel methods i.e. pyrosequencing provides a more in depth insight in population dynamics and can contribute to source tracking studies. The study of water and crop microbial quality targeting whole communities can provide greater insight into factors influencing observed variations, compared to the study of single viable species only.

Acknowledgements

This study was part of a solicited research project (K5/1875/4: Evaluation of the risks associated with the use of rainwater harvested from rooftops, for domestic use and homestead food gardens; and groundwater for domestic use and livestock watering) funded by the Water Research Commission and co-funded by Department of Agriculture, Forestry, and Fisheries, Republic of South Africa. This study was also funded by the DST-NRF Center of Excellence, South Africa.

Conflict of interest

The authors have no conflict of interest to declare. We confirm that this is an original research paper that has not been submitted anywhere else for publication.

References

Ahmed, W., Brandes, H., Gyawali, P., Sidhu J.P.S. and Toze S. (2014) Opportunistic pathogens in roof-captured rainwater samples, determined using quantitative PCR. *Water Res* 5, 361-369.

Allende, A. and Monaghan, J. (2015) Irrigation water quality for leafy crops: A perspective of risks and potential solutions. *Int J Environ Res Public Health* 12, 7457-7477.

Bray, J.R. and Curtis, J.T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol monographs* 27, 325-349.

Ceuppens, S., Hessel, C.T., Quadros Rodrigues, R., Bartz, S., Tondo, E.C. and Uyttendaele, M. (2014) Microbiological quality and safety assessment of lettuce production in Brazil. *Int J Food Microbiol* 181, 67-76.

Chao A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43, 783-791.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R. and Tiedje, J.M. (2013) Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic acids Res* 1-10.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M. and Tiedje, J.M. (2009) The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic acids research 37*, 141-145.

Dees, M.W., Lysøe, E., Nordskog, B. and Brurberg, M.B. (2015) Bacterial communities associated with surfaces of leafy greens: Shift in composition and decrease in richness over time. *Appl Environ Microbiol* 81, 1530-1539.

Dobrowsky, P. H., De Kwaadsteniet, M., Cloete, T. E. and Khan W. (2014). Distribution of indigenous bacterial pathogens and potential pathogens associated with roof-harvested rainwater. *Appl Environ Microbiol* 80, 2307-2316.

Du Plessis, E., Duvenage, F. and Korsten, L. (2015) Determining the potential link between irrigation water quality and the microbiological quality of onions by phenotypic and genotypic characterization of *Escherichia coli* isolates. *J Food Prot* 78, 643-651.

Erickson, M.C., Liao, J., Payton, A.S., Webb, C.C., Ma, L., Zhang, G., Flitcroft, I., Doyle, M. P. and Beuchat, L.R. (2013) Fate of *Escherichia coli* O157:H7 and *Salmonella* in soil and lettuce roots as affected by potential home gardening practices. *J Sci Food Agric* 93, 3841-3849.

Gemmell, M.E. and Schmidt, S. (2012) Microbiological assessment of river water used for the irrigation of fresh produce in a sub-urban community in Sobantu, South Africa. *Food Res Int* 47, 300-305.

Hausdorf, L., Neumann, M., Bergmann, I., Sobiella, K., Mundt, K., Fröhling, A., Schlüter, O. and Klocke, M. (2013) Occurrence and genetic diversity of *Arcobacter* spp. in a spinach-processing plant and evaluation of two *Arcobacter*-specific quantitative PCR assays. *Syst Appl Microbiol* 36, 235-243.

Huse, S.M., Mark Welch, D.B., Voorhis A., Shipunova, A., Morrison HG, Eren, A.M., and Sogin, M.L. (2014) VAMPS: A website for visualization and analysis of microbial population structures. *BMC Bioinformatics* 15, 41.

Huson, D.H. (2016) User manual for MEGAN V5.11.3. *http://ab.inf.uni-tuebingen.de/data/software/megan5/download/manual.pdf*. Accessed on 21.02.2016.

Hur, I. and Chun, J. (2004) A method for comparing multiple bacterial community structures from 16S rDNA clone library sequences. *J Microbiol* 42, 9-13.

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

Jongman, M. and Korsten, L. (2016a) Assessment of irrigation water quality and microbiological safety of leafy greens in different production systems. *J Food Saf* doi:10.1111/jfs.12324.

Jongman, M. and Korsten, L. (2016b) Genetic diversity and antibiotic resistance of *Escherichia coli* isolates from different leafy green production systems. *J Food Prot* 79, 1846–1853.

Leff, J.W. and Fierer, N. (2013) Bacterial communities associated with the surfaces of fresh fruits and vegetables. *PLoS One* 8, e59310.

Leonard, S.R., Mammel, M.K., Lacher, D.W. and Elkins, C.A. (2015) Application of metagenomic sequencing to food safety: Detection of shiga toxin-producing *Escherichia coli* on fresh bagged spinach. *Appl Environ Microbiol* 81, 8483-8491.

Lim, Y.M., Kim, B.K., Kim, C., Jung, H.S., Kim, B.S., Lee, J.H. and Chun, J. (2010). Assessment of soil fungal communities using pyrosequencing. *J Microbiol* 48, 284-289.

Lopez-Velasco, G., Carder, P.A., Welbaum, G.E. and Ponder, M.A. (2013) Diversity of the spinach (*Spinacia oleracea*) spermosphere and phyllosphere bacterial communities. *FEMS Microbiol Lett* 346, 146-154.

Monchy, S., Sanciu, G., Jobard, M., Rasconi, S., Gerphagnon, M., Chabé, M., Cian, A., Meloni, D., Niquil, N., Christaki, U., Viscogliosi, E. and Sime-Ngando, T. (2011). Exploring and quantifying fungal diversity in freshwater lake ecosystems using rDNA cloning/sequencing and SSU tag pyrosequencing. *Environ Microbiol* 13, 1433-1453.

Muller, T. and Ruppel, S. (2014) Progress in cultivation-independent phyllosphere microbiology: Mini review. *FEMS Microbiol Ecol* 87, 2–17.

Muyzer, G., De Waal, E.C. and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl environ microbiol* 59, 695-700.

Navarro-Noya, Y.E., Suárez-Arriaga, M.C., Rojas-Valdes, A., Montoya-Ciriaco, N.M., Gómez-Acata, S., Fernández-Luqueño, F. and Dendooven, L. (2013) Pyrosequencing analysis of the bacterial community in drinking water wells. *Microbiol Ecol* 66, 19-29.

Nygard, K., Lassen, J., Vold, L., Andersson, Y., Fisher, I., Lofdahl, S., Threlfall, J., Luzzi, I., Peters, T., Hampton, M., Torpdahl, M., Kapperud, G. and Aavitsland. P. (2008) Outbreak

of *Salmonella* Thompson infections linked to imported rucola lettuce. *Foodborne Path Dis* 5, 165–173.

Pachepsky, Y., Shelton, D.R., McLain, J.E.T., Patel, J. and Mandrell, R.E. (2011) Irrigation waters as a source of pathogenic microorganisms in produce: A review. *Adv In Agronomy* 113, 74-105.

Park, S. and Kang, D. (2015) Combination treatment of chlorine dioxide gas and aerosolized sanitizer for inactivating foodborne pathogens on spinach leaves and tomatoes. *Int J Food Microbiol* 207, 103-108.

Shannon, C.E. and Weaver, W. (1963) The mathematical theory of communication. University of Illinois Press, Urbana, IL.

Smith DeWaal, C. and Bhuiya, F. (2007) Outbreak alert 2007.Center for Science in the Public Interest. Ninth Edition. Accessed on 06. 06. 2014. http://cspinet.org/new/pdf/outbreakalertreport09.pdf.

Speelman, S., D'Haese, M., Buysse, J. and D'Haese, L. (2008) A measure for the efficiency of water use and its determinants, a case study of small-scale irrigation schemes in North-West Province, South Africa. *Agric Syst* 98, 31–39.

Telias, A., White, J.R., Pahl, D.M., Ottesen, A.R. and Walsh, C.S. (2011) Bacterial community diversity and variation in spray water sources and the tomato fruit surface. *BMC Microbiol* 11, 81.

Teplitski, M., Barak, J.D and Schneider, K.R. (2009) Human enteric pathogens in produce: Un- answered ecological questions with direct implications for food safety. *Curr Opin Biotechnol* 20, 166-171.

Van der Merwe, V., Duvenage, S. and Korsten, L. (2013) Comparison of biofilm formation and water quality when water from different sources was stored in large commercial water storage tanks. *J Water Health* 11, 30-40.

Weisburg, W.G., Barns, S.M., Pelletier, D.A. and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J bacteriol* 173, 697-703.

Wingender, J. and Flemming, H.C. (2011) Biofilms in drinking water and their role as reservoir for pathogens. *Int J Hyg Environ Health* 214, 417.