RESEARCH ARTICLE

Soil salinity, pH, and indigenous bacterial community interactively influence the survival of *E. coli* O157:H7 revealed by multivariate statistics

Ziming Han [1,2,3] · Jincai Ma [1,2,4] · Ching-Hong Yang [5] · Abasiofiok Mark Ibekwe [6]

Abstract
Complexities of biotic-abiotic interactions in soils result in the lack of integrated understanding of environmental variables that restrict the survival of shiga toxin-producing *E. coli* O157:H7. Herein, we reanalyzed previously published data and highlighted the influence of soil abiotic factors on *E. coli* O157:H7 survivability and elucidated how these factors took effect indirectly through affecting indigenous bacterial community. Interaction network analysis indicated salinity and pH decreased the relative abundances of some bacterial taxa (e.g., *Acidobacteria_Gp4*, *Acidobacteria_Gp6*, and *Deltaproteobacteria*) which were positively correlated with the survival of *E. coli* O157:H7 in soils, and vice versa (e.g., *Gammaproteobacteria* and *Flavobacteria*) ($P < 0.05$). An array of multivariate statistical approaches including partial Mantel test, variation partition analysis (VPA), and structural equation model (SEM) further confirmed that biotic and abiotic factors interactively shaped the survival profile of *E. coli* O157:H7. This study revealed that some bacterial taxa were correlated with survival of *E. coli* O157:H7 directly, and salinity and pH could affect *E. coli* O157:H7 survival through changing these bacterial taxa. These findings suggest that salinity in soil might benefit the control of fecal pathogenic *E. coli* invasion, while soil acidification caused by anthropogenic influences could potentially increase the persistence of *E. coli* O157:H7 in agro-ecosystem.

✉ Jincai Ma
jincaima@jlu.edu.cn

1   Key Laboratory of Ground Water Resource and Environment, Ministry of Education, Jilin University, Changchun 130021, China

2   Jilin Provincial Key Laboratory of Water Resources and Environment, Jilin University, Changchun 130021, China

3   State Key Laboratory of Environmental Aquatic Chemistry, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

4   College of New Energy and Environment, Jilin University, Changchun 130021, China

5   Department of Biological Sciences, University of Wisconsin, Milwaukee, WI, USA

6   USDA-ARS U. S. Salinity Laboratory, Riverside, CA, USA

Introduction

Shiga toxin-producing *Escherichia coli* O157:H7 could be dangerous because of its low infective threshold (as few as 10-50 cells) and high pathogenicity (even death) (Ritchie et al. 2003), and thus, its transmission attracts great attention (Franz et al. 2018). Fresh fruit and vegetables served as vehicles for transmission of *E. coli* O157:H7 (Berger et al. 2010; Turner et al. 2019), because they could harbor pathogens on their surfaces or internalize pathogens during their growth (Kumar et al. 2017; Teplitski and de Moraes 2018). During the transmission process from manure to human, agricultural soils are a vital reservoir for the enteric pathogens because of the application of poorly composted manures and contaminated irrigation water (Berger et al. 2010; van Elsas et al. 2011). Infection probability of fresh produce depended on the survival time and physiological activity of *E. coli* cells in soils (Ongeng et al. 2015). Thus, a comprehensive understanding of the retention of *E. coli* O157:H7 in agricultural soils was critical to optimize agricultural management strategies and public health policies.

Study of the factors affecting *E. coli* O157:H7 survival was one of the core issues to understand the fate of pathogenic *E. coli* in the environment (Ongeng et al. 2015; van Elsas et al. 2011). Numerous studies have investigated the relationships between the survivals of *E. coli* O157:H7 in soils and environmental factors. From the abiotic factors perspective, nutrient availability generally had positive effects on *E. coli* O157:H7, due to the chemoheterotrophic nature of *E. coli*. It was found that dissolved organic carbon was one of the major determinants promoting the survival of *E. coli* O157:H7 in manure-amended soil (Franz et al. 2008). Furthermore, soil pH (Balamurugan et al. 2015; Xing et al. 2019), moisture (Ongeng et al. 2011), and texture (Cai et al. 2018) were also correlated to the *E. coli* O157:H7 survivability. From the biotic factors perspective, indigenous bacterial community constrained the survival and colonization of *E. coli* O157:H7 in soils (Mallon et al. 2015b; van Elsas et al. 2012). A possible reason for this phenomenon was that robust diversity increased resource competition and decreased invasion by niche preemption (Mallon et al. 2015a). In addition, agriculture management practices (Franz et al. 2008; Yao et al. 2013, 2015) and global climate change (Hellberg and Chu 2015) were also potential contributors to decline the concentration of *E. coli* in soils, which took effects mainly through changing soil abiotic and/or biotic variables. Undoubtedly, physicochemical properties and microbial community of soil were interactively correlated instead of working alone (Bahram et al. 2018), but how these complex interactions among biotic-abiotic factors of soils link to the persistence of invasive *E. coli* O157:H7 were not examined before. It was of importance to integrate the abiotic and biotic variables to obtain a full understanding of the survival mechanism of *E. coli* O157:H7 in soils.

Our previous studies have profiled the correlations between survival parameters of *E. coli* O157:H7 and soil physicochemical properties (electrical conductivity or salinity, pH, water soluble organic carbon (WSOC) and total soluble nitrogen (TN), etc.), and noticed salinity and pH showed deleterious effects on *E. coli* invasion (Ma et al. 2012a). We also have found that survival parameters were linearly correlated to relative abundance of major bacterial phyla, e.g., *Actinobacteria*, *Acidobacteria*, *Proteobacteria*, and *Bacteroidetes* (Ma et al. 2013). Obviously, these finding still could not answer the question that how the interactions between soil properties and bacterial community influence the survival of *E. coli* O157:H7 in soils.

In this study, we reanalyzed survival data of *E. coli* O157:H7 EDL933, high-throughput sequencing data of soil indigenous bacterial community and soil properties of 32 vegetable producing soils. These soils were collected from three major leafy green producing areas in the USA: Salinas Valley of northern California (SA), Imperial Valley of southern California (IM), and Yuma Valley of Arizona (AZ). An array

of statistical analytical tools including correlation analysis, random forest model, Mantel test, interaction network analysis, variation partition analysis (VPA), and structural equation model (SEM) was selected for data mining. The objectives of this study were to (1) disentangle how salinity and pH indirectly constrain *E. coli* O157:H7 survivability through shaping indigenous bacterial community and (2) quantify the individual and joint contributions of physicochemical properties and bacterial community to survival of *E. coli* O157:H7 in soils.

Materials and methods

Soil collection, characterization, DNA extraction, and sequencing

A total of 32 soil samples were collected from three major leafy green producing areas: Salinas Valley, northern California; Imperial Valley, southern California; and Yuma Valley, Arizona. The geographical location and environmental parameters were previously described by Ma et al. (2012b). Soil collection, pretreatment, and physicochemical properties characterization (including salinity (EC, dS m^{-1}), water soluble organic carbon (WSOC, mg kg^{-1}), total soluble nitrogen (TN, %), and pH) can be found in the study of Ma et al. (2012a). Soil DNA extraction, high-throughput sequencing, and sequence data processing can be found in the study of Ma et al. (2013). In brief, bacterial tag-encoded FLX amplicon pyrosequencing was carried out, and sequences that were < 200 bp after quality trimming were removed. Sequencing data were further analyzed using MOTHUR v1.9.1. The sequencing data could be downloaded in https://www.ars.usda.gov/ARSUSERFILES/20361500/PUBLIC%20DATA/SALINITY%20PRODUCE%20SEQUENCING%20DATA%202010-2015.ZIP.

Survival of *E. coli* O157:H7 EDL933 in soils and survival data modeling

Survival experiments have been reported previously (Ma et al. 2012a, 2013). In brief, an *E. coli* O157:H7 strain EDL933 (ATCC 43895) wild type was tagged with nalidixic acid in addition to rifampicin resistance to facilitate the enumeration; its growth in LB (Luria−Bertani) broth and survival in soils were found to be identical to that of the non-tagged wild-type strain (Ma et al. 2011). Cells were added in soils to a final density of about 5×10^6 CFU per gram soil dry weight (gdw^{-1}) (Franz et al. 2008). All experiments were performed under room temperature (22 ± 1 °C), and moisture content was maintained constantly by adding more water to make up for evaporation according to non-inoculated control (Ma et al. 2013). The inoculated soils were sampled at days 0, 3, 6, 10, 14, 20, 27, 34, 40, and 48, and survivors of *E. coli* O157:H7

over time were counted (Ma et al. 2011). The measured data were fitted to the Weibull survival model (Albert and Mafart 2005) using GInaFiT version 1.5 (Geeraerd et al. 2005), and survival parameters including *ttd* (time needed to reach detection limit of 100 CFU gdw^{-1}), δ (time needed for first decimal reduction), and *p* (shape parameter) were calculated at the same time (Franz et al. 2008). More details on survival data modeling could be found in the Method part in Supplementary Information.

Statistical analysis

Pearson correlation analysis, permutational MANOVA (PERMANOVA) with Bray-Curtis dissimilarity of OTU table, and detrended corresponding analysis (DCA) with Bray-Curtis dissimilarity of OTU table were performed using vegan package (Dixon 2003) in R 3.3.3. The first coordinate of DCA (DCA1) of bacterial community explained 37.9% of the overall variation (Fig. S1d).

Mantel test could be used to calculate correlations between corresponding positions of two distance matrices derived from multivariate data with significance test by permutation (Mantel 1967). Partial Mantel test is an extension of Mantel test, where the influence of third matrix is controlled while the relationship of the first two is determined (Smouse et al. 1986). In other words, partial Mantel test attempts to correct for the effect of another series of variables stored in an additional matrix or matrices (Buttigieg and Ramette 2014). When considering a set of explanatory variables, partial Mantel test excluded the influence of other factors when calculating one factor's effect. Mantel test and partial Mantel test were performed with vegan package (Dixon 2003) in R 3.3.3.

Redundancy analysis (RDA) was used to extract and summarize the variation of response variables (i.e., *E. coli* O157:H7 survival parameters) that can be explained by a series of explanatory variables (i.e., soil properties and microbial community) (Buttigieg and Ramette 2014). Then, variation partition analysis (VPA) was performed to determine the partial effect of each explanatory matrix on the response data (Borcard et al. 1992), and its result was displayed in a Venn diagram. The shared partitions in VPA results meant the variation in response data could be explained by both explanatory matrices (Peres-Neto et al. 2006). A larger partition fraction means a more multicollinearity effect in the model. Thus, a larger shared partition indicates the complex interactions among explanatory variables. RDA and VPA were performed with vegan package (Dixon 2003) in R 3.3.3.

Random forest is an ensemble of unpruned regression or classification trees created using bootstrap samples of the training data and random feature selection in tree induction (Breiman 2001). Here, it was performed to quantitatively assess the important predictors to survival parameters (*ttd*, δ, and *p*) of *E. coli* O157:H7. The importance of each predictor was determined by assessing the decrease in prediction accuracy, i.e., increase in the mean square error (MSE) between observations and predictions, when the data for the predictor was randomly permuted (Breiman 2001). This decrease was averaged over all trees to produce the final measurement of importance. This accuracy of importance measure was calculated for each tree and averaged over the forest. These analyses were conducted using the randomForest package (Liaw and Wiener 2002), and the significance of predictor importance on the survival parameters was assessed by using the rfPermute package (Archer 2013) in R 3.3.3.

To visualize the associations among survival parameters (*ttd*, δ, and *p*) of *E. coli* O157:H7, soil physicochemical properties, and bacteria in the interaction network interface, a correlation matrix was constructed based on the possible pairwise Pearson's rank correlations. The distribution matrix of bacteria was standardized based on their relative abundance for network construction (Barberán et al. 2011). To reduce network complexity, all connections within bacterial groups were deleted. A valid co-occurrence was considered a statistically significant correlation between species with the Pearson's *P* value < 0.05. Network analysis was performed using the iGraph (Csardi and Nepusz 2006) and Hmisc (Harrell and Frank 2008) packages in R 3.3.3. The interaction network was visualized using Gephi 0.9.2 (Bastian et al. 2009).

Structural equation model (SEM) is an a priori method with the capacity to identify casual relationships between variables by fitting data to the models representing causal hypotheses (Byrne 2001; Eisenhauer et al. 2015). Here, SEM was used to gain a mechanistic understanding of how soil physicochemical properties and bacterial community influence survival of *E. coli* O157:H7, and direct and indirect standardized path coefficients (λ) of soil properties and bacterial community were calculated. Here, "direct" was defined as the link between an explanatory variable and a response variable, while "indirect" indicated an explanatory variable influenced another explanatory variable, and the latter directly influenced the response variable. The "total effect" was the combination of direct effect and indirect effect. This analysis was performed with lavaan package using the robust maximum likelihood estimation method (Rosseel 2012) in R 3.3.3. The χ^2 value and its associated *P* value, ratio of χ^2 and degrees of freedom (CMIN/DF), goodness-of-fit index (GFI), root mean square error of approximation (RMSEA), and Akaike information criterion (AIC) were used to judge the model fitness to the data. Concretely, these multiple goodness-of-fit criteria are as follows: $P > 0.05$, CMIN/DF < 2, GFI > 0.90, RMSEA < 0.08, and the AIC from the default model lower than that from saturated model and independence model (Hooper et al. 2008; Schermelleh-Engel et al. 2003).

Results

Survival of *E. coli* O157:H7 in soils and its influential factors

Based on Weibull model, the survival of *E. coli* O157:H7 was quantified by model parameters including *ttd* (time needed to reach detection limit of 100 CFU gdw^{-1}), δ (time needed for first decimal reduction), and *p* (shape parameter). *E. coli* O157:H7 survived longer in soils from Salinas Valley, California (SA, *ttd* = 29.55 ± 3.01 days, δ = 11.90 ± 2.80 days) than Imperial Valley, California (IM, *ttd* = 18.83 ± 3.53 days, δ = 6.14 ± 2.22 days) and Yuma Valley, Arizona (AZ, *ttd* = 21.15 ± 6.74 days, δ = 4.43 ± 1.37 days) ($P < 0.001$) (Fig. S1). Meanwhile, salinity and pH in SA were significantly lower than in soils from Imperial Valley and Yuma ($P < 0.05$); bacterial community in SA soil was also different compared with Imperial and Yuma soils ($P < 0.001$, PERMANOVA) (Figs. S1 and S2).

Bacterial community, salinity, pH, WSOC, and TN were used to explain survival parameters (*ttd*, δ, and *p*) based on random forest model, and 59.52% variation of *ttd*, 62.11% variation of δ, and 37.82% variation of *p* could be explained by abovementioned variables (Table 1). In random forest model, the bigger mean square error (MSE) value means the greater influence of observations on predictions. For instance, the MSE of *ttd* increased 11.94% when removing the predictor of bacterial community ($P < 0.01$) and increased 12.99% when removing the predictor of salinity ($P < 0.01$) (Table 1). In addition, Mantel test also correlated environmental factors and survival parameters (Table 2). For instance, *ttd* was significantly correlated with bacterial community ($r = 0.391$, $P < 0.001$), salinity ($r = 0.444$, $P < 0.001$), and pH ($r = 0.220$, $P < 0.05$) by Mantel test. The results of random forest model and Mantel test confirmed the results of the correlation analysis (Table S1).

Interaction network among survival parameters, soil properties, and bacteria

To explore the specific Pearson correlations among survival parameters (*ttd*, δ, and *p*) of *E. coli* O157:H7, soil physicochemical properties, and major bacterial groups (at class level), interaction network was depicted (Fig. 1). An edge represented a significant ($P < 0.05$) correlation, and more details were shown in Table S2. Soil physicochemical properties were associated with few bacterial groups, and these groups were further associated with survival parameters (Fig. 1), which were indicated as green edges (i.e., survival parameters) and pink edges (i.e., physicochemical properties) linking blue nodes (i.e., bacterial groups) at the same times. A total of six major bacterial groups at class level whose relative abundance > 1%, including *Gammaproteobacteria*, *Solirubrobacterales*, *Acidobacteria_Gp6*, *Deltaproteobacteria*, *Acidobacteria_Gp4*, and *Flavobacteria*, had significant correlation with both soil physicochemical properties and survival parameters (Fig. 1, Table 3). Salinity and pH were negatively correlated with *Acidobacteria_Gp4*, *Acidobacteria_Gp6*, and *Deltaproteobacteria*, and these taxa were positively correlated with *ttd*. Salinity and pH were positively correlated with *Gammaproteobacteria* and *Flavobacteria*, and these taxa were negatively correlated with ttd ($P < 0.05$, Table 3). Similar trends were found in the influencing factors of δ and *p*. In addition, WSOC and TN were positively correlated with *Solirubrobacterales*, which was positively correlated with *ttd* ($P < 0.05$, Table 3). The interaction network showed that the bacterial community was associated with soil properties and survival parameters at the same time.

Correlation analysis between accumulated bacterial abundance at class level and survival parameters

Based on results of network analysis, plenty of bacterial taxa were found to be positively or negatively correlated with the

Table 1 The importance[a] (mean decrease accuracy in random forest model) of soil properties and bacterial community on survival parameters (*ttd*, δ, and *p*) of *E. coli* O157:H7

	Bacterial community[b]	Salinity[c]	pH	WSOC	TN	Variation explained
ttd	*11.936***	*12.986***	2.548	*12.881***	*7.697**	59.52%
δ	*13.191***	*14.772***	*5.944**	− 0.500	*8.582***	62.11%
p	*9.653**	*11.327***	4.444	2.741	3.922	37.82%

[a] The importance of each predictor was determined by assessing the decrease in prediction accuracy (that is, increase in the mean square error (MSE) between observations and predication) when the data for the predictor was randomly permuted. Models for the survival parameters were all significant at the 0.01 level with 500 trees

[b] The first coordinate of detrended correspondence analysis (DCA1 explained 37.9% of the variation) was used to represent the bacterial community composition

[c] Electrical conductivity (EC, mS/m) of soil extract was used to represent the soil salinity

WSOC water soluble organic carbon, *TN* total soluble nitrogen, *ttd* time needed to reach detection limit of 100 CFU gdw^{-1}, δ time needed for first decimal reduction, *p* shape parameter

Significant levels of each predictor were as follows: *$P < 0.05$ and **$P < 0.01$

Table 2 Mantel tests and partial Mantel tests showed contributions of environmental factors to the survival parameters (*ttd*, δ, and *p*) of *E. coli* O157:H7 were generally overestimated when interactions within environmental factors were neglected

	Methods	Bacterial community[a]	Salinity[b]	pH	WSOC	TN
ttd	Mantel	0.391***	0.444***	0.220*	0.338***	0.314**
	Partial Mantel	0.114*	0.377**	− 0.009	0.362***	0.118
δ	Mantel	0.605***	0.337**	0.438**	0.090	0.577***
	Partial Mantel	0.413**	0.233*	0.293*	0.083	0.472***
p	Mantel	0.153**	0.043	0.068	0.172*	0.168*
	Partial Mantel	0.150*	0.041	0.076	0.169*	0.152*

[a] The first coordinate of detrended correspondence analysis (DCA1) was used to represent the bacterial community composition

[b] Electrical conductivity (EC, mS/m) of soil extract was used to represent the soil salinity. *WSOC* water soluble organic carbon, *TN* total soluble nitrogen, *ttd* time needed to reach detection limit of 100 CFU gdw^{-1}, δ time needed for first decimal reduction, *p* shape parameter

Significant levels of each predictor were as follows: *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$

survival parameters of *E. coli*. It was also interesting to explore the links between survival parameters and cumulative relative abundance of class who were correlated with survival parameters. Concretely, the sums of relative abundance of classes who were positive with *ttd*, negative with *ttd*, positive with δ, and positive with *p* ($P < 0.05$) were calculated. Corresponding accumulated bacterial class abundance significantly correlated with survival parameters of *E. coli* O157 ($P < 0.05$, Fig. 2). In addition, similar phenomena were found between accumulated bacterial class abundance and soil physicochemical properties ($P < 0.05$, Fig. S3).

Redundancy analysis between environmental factors and survival profile

Redundancy analysis (RDA) and further variation partition analysis (VPA) were used to calculate the proportional contribution of environmental factors to *E. coli* O157 survival variation. Results of RDA and VPA indicated that selected environmental factors explained over 67% of the total variation of survival profile of *E. coli* O157 (Fig. 3). In VPA, all variables were divided into three groups including biotic factor (bacterial community), negative abiotic factors of survival (salinity and pH),

Table 3 Bacterial taxa (average relative abundance > 1% at class level) which correlated (Pearson's $P < 0.05$) with survival parameters (*ttd*, time needed to reach detection limit of 100 CFU gdw^{-1}; δ, time needed for first decimal reduction; *p*, shape parameter) of *Escherichia coli* O157:H7, and soil physicochemical properties at the same time

Taxa at class level and average relative abundance (%)		Pearson correlation between survival parameters and taxa			Pearson correlation between soil properties and taxa		
		Survival parameters	*r*	*P*	Soil properties	*r*	*P*
Gammaproteobacteria	9.690	*ttd*	− 0.387	0.031	Salinity	0.401	0.025
Solirubrobacterales	5.134	*ttd*	0.420	0.019	WSOC	0.552	0.001
Solirubrobacterales	5.134	*ttd*	0.420	0.019	TN	0.509	0.003
Acidobacteria_Gp6	2.967	*ttd*	0.365	0.044	Salinity	− 0.420	0.019
Acidobacteria_Gp6	2.967	δ	0.420	0.019	Salinity	− 0.420	0.019
Deltaproteobacteria	2.885	δ	0.416	0.020	Salinity	− 0.507	0.004
Deltaproteobacteria	2.885	*ttd*	0.538	0.002	WSOC	0.407	0.023
Deltaproteobacteria	2.885	δ	0.416	0.020	WSOC	0.407	0.023
Acidobacteria_Gp4	1.787	*ttd*	0.385	0.032	Salinity	− 0.516	0.003
Acidobacteria_Gp4	1.787	*p*	0.465	0.008	Salinity	− 0.516	0.003
Acidobacteria_Gp4	1.787	δ	0.600	< 0.001	Salinity	− 0.516	0.003
Acidobacteria_Gp4	1.787	*ttd*	0.385	0.032	pH	− 0.438	0.014
Acidobacteria_Gp4	1.787	*p*	0.465	0.008	pH	− 0.438	0.014
Acidobacteria_Gp4	1.787	δ	0.600	< 0.001	pH	− 0.438	0.014
Flavobacteria	1.022	*ttd*	− 0.377	0.036	Salinity	0.561	0.001

Indirect and direct effects of soil properties and bacteria on survival parameters

To obtain deeper understanding of the "joint effect" of biotic and abiotic factors revealed by VPA, partial Mantel test and structural equation model (SEM) were performed. The correlation coefficient (r) of partial Mantel test was generally smaller than that of Mantel test (Table 2), indicating the actual effect of one environmental variable was corrected when controlling the effects of other variables. For instance, the correlation coefficient between ttd and salinity was 0.444 ($P < 0.001$) by Mantel test and 0.377 ($P < 0.01$) by partial Mantel test, indicating partial influence of salinity reflected by other environmental factors (i.e., bacterial community). The correlation coefficient between ttd and pH was 0.220 ($P < 0.05$) by Mantel test and -0.009 ($P > 0.05$) by partial Mantel test, indicating almost all influence of pH reflected by other environmental factors (i.e., bacterial community). In short, these tests suggest the survival of $E.$ $coli$ O157:H7 was interactively influenced by biotic and abiotic factors.

Based on structural equation model (SEM), the interactions among environmental variables were quantitatively depicted as the indirect and direct effects of soil properties and bacterial community on survival parameters (Fig. 4). Salinity, pH, and TN had indirect links to ttd, δ, and p through bacterial community. Concretely, salinity and pH showed negative effects on survival parameters, while TN showed positive effects (Fig. 4). The deleterious effect of salinity on ttd was the combination of direct effect and indirect effect through bacterial community, while the deleterious effect of pH on ttd was only obtained as indirect pathway. A similar trend was found in the deleterious effect of salinity and pH on δ and p. The SEM results clearly showed the indirect and direct effects of soil properties and bacteria on survival parameters of $E.$ $coli$ O157:H7. Notably, the R^2 value of p was smaller than that of ttd and δ, indicating p was shaped by more unmeasured variables.

Discussion

Based on multivariate statistical methods, this study systematically disentangled the direct and indirect effects of soil biotic and abiotic factors on the survival profile of $E.$ $coli$ O157:H7, and highlighted salinity and pH constrained the survival of $E.$ $coli$ O157:H7 through changing indigenous bacterial community. To the best of our knowledge, this study may be the first to reveal the deleterious influence of high soil salinity and pH on $E.$ $coli$ O157:H7 through biotic-abiotic interaction perspective. The longer survival of $E.$ $coli$ O157:H7 in Salinas Valley, California, in comparison with others might be shaped by plenty of factors, and the lower salinity, lower pH, and distinct indigenous bacterial
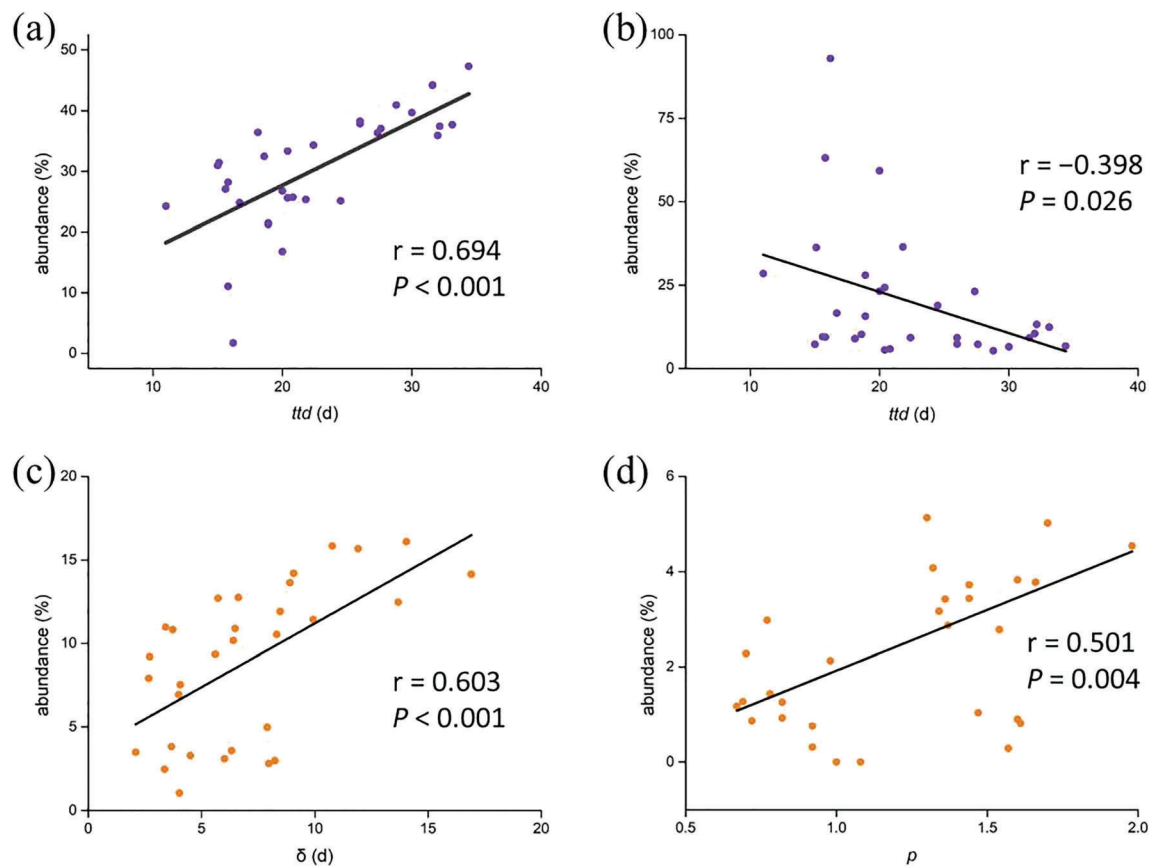

Fig. 1 Network analysis depicting interaction patterns among survival parameters (ttd, δ, and p) of *Escherichia coli* O157:H7, soil physicochemical properties, and major bacterial groups (at class level). The size of each node was proportional to the number of connections. An edge represented a significant (Pearson's $P < 0.05$) correlation (all connections within bacterial groups were deleted), and edges were weighted according to the correlation coefficient. The nodes were colored according to the variable types: survival parameters, soil properties, and bacterial classes were colored as green, pink, and blue, respectively. Only bacterial classes who average relative abundance > 1% were labeled. Electrical conductivity (EC, mS/m) of soil extract was used to represent the soil salinity. WSOC, water soluble organic carbon; TN, total soluble nitrogen; Gamma, Gammaproteobacteria; Delta, Deltaproteobacteria; Gp4, Acidobacteria_Gp4; Gp6, Acidobacteria_Gp6; ttd, time needed to reach detection limit of 100 CFU gdw^{-1}; δ, time needed for first decimal reduction; p, shape parameter

and positive abiotic factors of survival (WSOC and TN). Notably, bacterial community alone and negative abiotic factors alone contributed only 3.32%, 7.92% to overall variation of survival profile, respectively, while the joint contribution of bacterial community and negative abiotic factors to overall variation of survival profile was 30.49% (Fig. 3). Similarly, the joint effect of bacterial community and positive abiotic factors was greater than the effect of bacterial community alone (5.39% > 3.32%). Overall, the joint effect of all biotic and abiotic variables contributed 49.15% (30.49% + 13.27% + 5.39%) to the variation of survival profile, while biotic variables and abiotic variables alone contributed 3.32% and 15.20% (7.92% + 1.14% + 6.14%) to the variation of survival profile, respectively. The VPA results indicated the effect of bacterial community on the survival of $E.$ $coli$ O157:H7 was mainly reflected in the joint effects with abiotic variables. In addition, the unmeasured soil variables (e.g., unmeasured physicochemical properties, bacteriophage) might contribute to the unexplained part (32.32%) in VPA.

Fig. 2 Correlation analysis between accumulated bacterial class abundance and survival parameters. Accumulated abundance indicated that the sum of classes` relative abundance who were positive with ttd (time needed to reach detection limit of 100 CFU gdw^{-1}) (**a**), negative with ttd (**b**), positive with δ (time needed for first decimal reduction) (**c**), and positive with p (shape parameter) (**d**) based on correlation analysis (Pearson's $P < 0.05$)

community in Salinas Valley, California, might make sense (Fig. S1). In fact, although the three sampling sites of this study were all major fresh produce growing areas in the

USA, some outbreaks of this pathogen caused by fresh produce have been traced back to Salinas Valley, California, instead of Imperial Valley, California, or Yuma Valley,

Fig. 3 Variation partition analysis (VPA) based on redundancy analysis (RDA) differentiating the effects of salinity, pH, water soluble organic carbon (WSOC), total soluble nitrogen (TN), and bacterial community on survival parameters (ttd, δ, and p) of *Escherichia coli* O157:H7 in soils. Electrical conductivity (EC, mS/m) of soil extract was used to represent the soil salinity. WSOC, water soluble organic carbon; TN, total soluble nitrogen
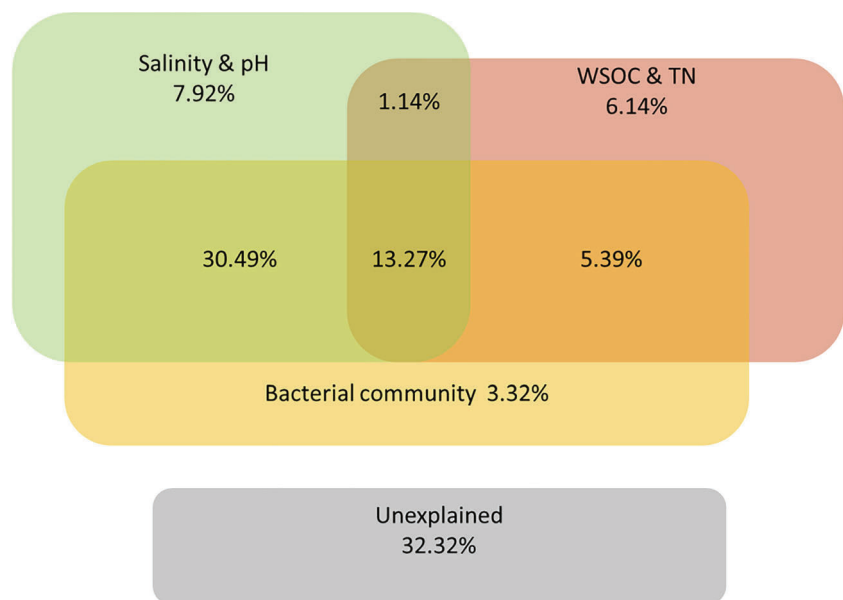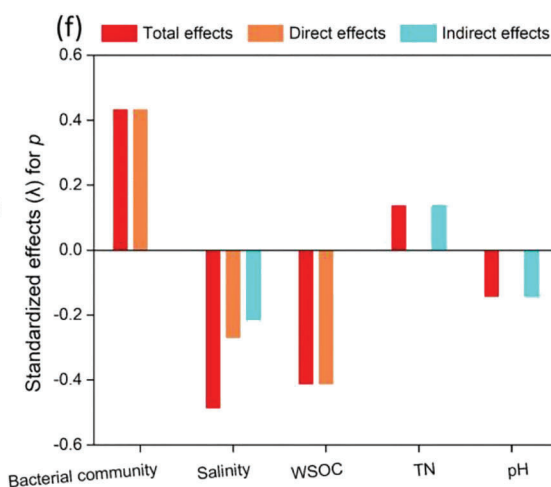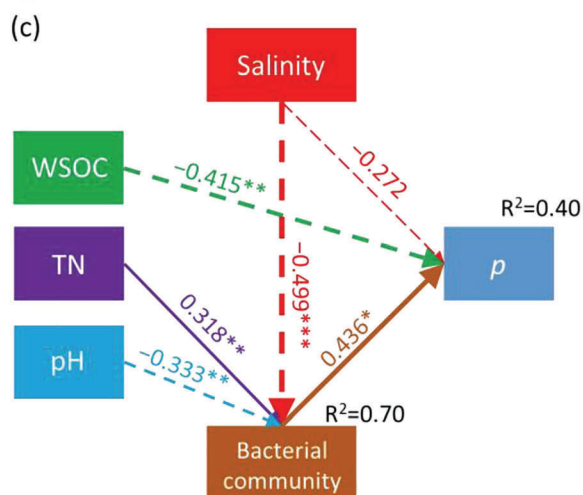
χ²=3.194, *P*=0.363, CMIN/DF=1.065, GFI=0.967, RMSEA=0.046

(a)
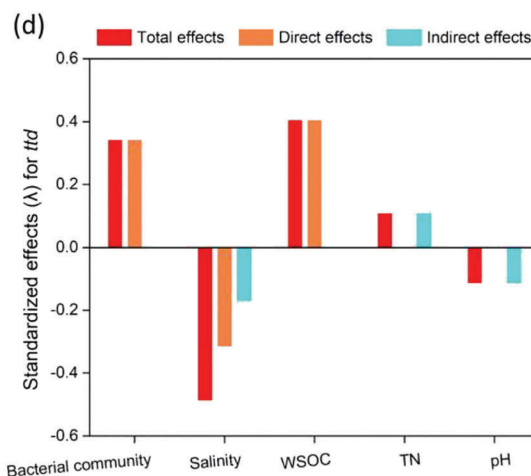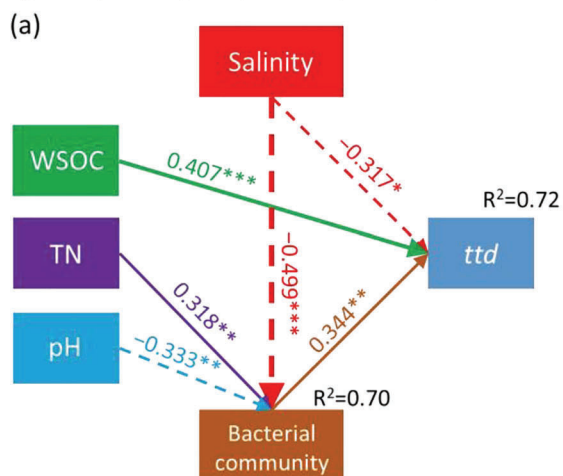
χ²=2.495, *P*=0.476, CMIN/DF=0.832, GFI=0.974, RMSEA<0.001

(b)
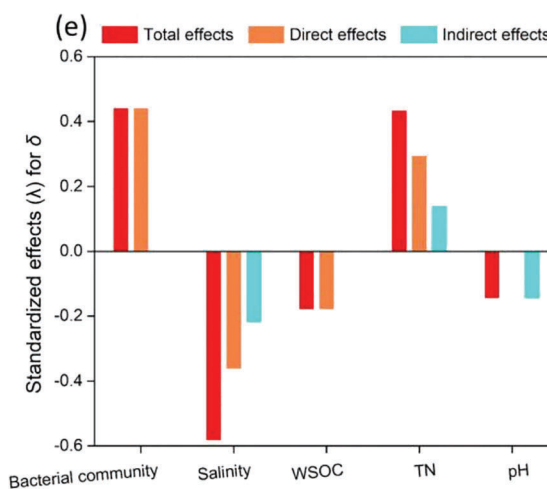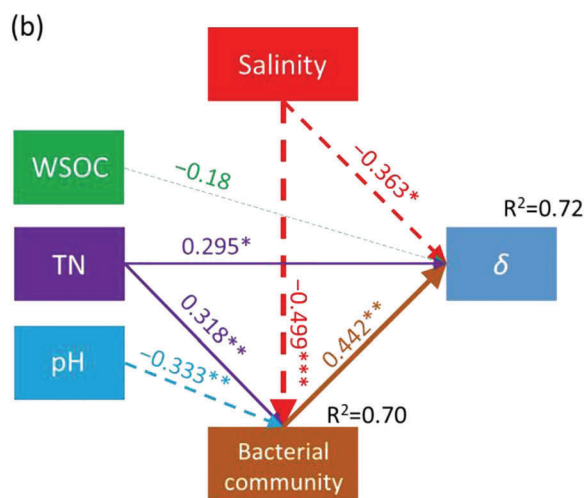
χ²=1.598, *P*=0.660, CMIN/DF=0.533, GFI=0.983, RMSEA<0.001

(c)

◀ **Fig. 4** Structural equation model (SEM) quantified the indirect and direct effects of soil physicochemical properties and bacterial community on survival parameters (*ttd*, time needed to reach detection limit of 100 CFU gdw^{-1}; *δ*, time needed for first decimal reduction; *p*, shape parameter) of *Escherichia coli* O157:H7 (**a**, **b**, and **c**). The width of the arrows indicated the strength of the standardized path coefficient (*λ*). The solid lines indicated positive path coefficients while dashed lines indicated negative path coefficients. R^2 values represented the proportion of the variance explained for each endogenous variable. The total effects were the sum of direct and indirect effects (**d**, **e**, and **f**). The first coordinate of detrended correspondence analysis (DCA1) was used to represent the bacterial community composition. CMIN/DF, ratio of χ^2 and degrees of freedom; GFI, goodness-of-fit index; RMSEA, root mean square error of approximation. Electrical conductivity (EC, mS/m) of soil extract was used to represent the soil salinity. WSOC, water soluble organic carbon; TN, total soluble nitrogen. Significant levels of each path were as follows: *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$

Arizona, (Cooley et al. 2007). It was meaningful to disentangle the adverse effects of high salinity and pH on *E. coli* O157:H7 for further agriculture management strategies.

Previous studies have noticed the sublethal effects of salt on *E. coli* (Anderson et al. 1979), and the adverse effects of salinity on *E. coli* cells might include a general osmotic effect and specific ion toxicity, causing decrease of enzymatic activities in metabolic process (Shabala et al. 2009). The *rpoS* gene was critical for *E. coli* to acquire resistance to multiple-stress, e.g., salinity, and the regulation of *rpoS* gene expression was useful for the understanding of *E. coli* O157:H7 survival mechanisms (Munro et al. 1995).

In addition, our data showed *Acidobacteria* and *Deltaproteobacteria* positively correlated with the survival of *E. coli* O157:H7, while *Gammaproteobacteria* and *Flavobacteria* were negatively correlated with the survival of *E. coli* O157:H7 (Fig. 1, Table 3). *Acidobacteria* and *Deltaproteobacteria* could promote carbon turnover by depleting the low molecular weight of carbon sources for *E. coli* to use easily (Eichorst et al. 2011; Westphal et al. 2011), which might be the possible reason of their positive correlations with *ttd*. As a member of *Gammaproteobacteria*, *E. coli* O157:H7 shared similar ecological niche with indigenous *Gammaproteobacteria*, and abundant *Gammaproteobacteria* in soils might restrain the colonization of invasive *E. coli*. As a major class of *Bacteroidetes*, *Flavobacteria* showed suppressive effect on the survival of *E. coli* O157:H7 in sand-based dairy livestock bedding (Westphal et al. 2011), which was in line with our data. Interestingly, salinity increased the abundance of *Gammaproteobacteria* and *Flavobacteria* while decreased the abundance of *Acidobacteria* and *Deltaproteobacteria*. In other words, salinity increased the bacteria negatively correlated to *E. coli* O157:H7, while decreased bacteria beneficial for *E. coli* O157:H7. Such indirect influence was shown in specific correlations among soil properties, bacterial taxa, and survival parameters, and was also proved by multivariate statistical methods (Table 2, Figs. 3 and 4).

Because of the exposure of *E. coli* O157:H7 to acidic pH in the gastrointestinal tract of animals, this pathogen seems to have developed molecular mechanisms to resist acid stress (Lin et al. 1996). Plenty of studies have also noticed the shorter survival time in alkaline pH than in acidic pH (Balamurugan et al. 2015; Franz et al. 2005; Wang et al. 2014; Xing et al. 2019; Yao et al. 2015). As mentioned previously, *Acidobacteria* was helpful for *E. coli* O157:H7 to survive in soils (Eichorst et al. 2011). Notably, our data revealed higher pH decreased the abundance of *Acidobacteria_Gp4*, which was positive correlated with survival time of *E. coli* (Fig. 1, Table 3). Thus, the adverse effect of high pH on *E. coli* seems to be the combination of direct stress and decrement of beneficial bacterial taxa for this pathogen, e.g., *Acidobacteria*. In addition, WSOC and TN increased the abundance of *Solirubrobacterales* and *Deltaproteobacteria*, which were beneficial for *E. coli* O157:H7 (Fig. 1, Table 3).

In the current study, multiple statistical methods were used, and each method had its goal. Firstly, correlation analysis, Mantel test, and random forest were used to explore the impact of environmental factors on *E. coli* O157 survival in soil. Secondly, network analysis could provide detailed information on the correlation between bacterial taxa and survival parameters of *E. coli* O157. Thirdly, both biotic and abiotic variables were considered at the same time in redundancy analysis (RDA) and further variation partition analysis (VPA), and their proportional contributions to variation of *E. coli* O157 survival were explored; VPA also indicated the importance of interactions among biotic and abiotic factors shaping *E. coli* O157 survival. Finally, partial Mantel test and structural equation model (SEM) were used to disentangle abovementioned "interaction," and direct and indirect influence of environmental factors on *E. coli* O157 survival were obtained. These multivariate statistical approaches have been applied in research on microbial ecology, but some of them might be used for *E. coli* survival data for the first time. Liang et al. (2015) explored the relationship between temperature and microbial community structure controlling for plant and soil variables by partial Mantel test. Zhu et al. (2017) used VPA to differentiate the effects of environmental factors on ARG structures in Chinese estuary samples. The current study attempted to extend the use of those methods on data mining for *E. coli* survival dynamic as indicated in a recent study (Huang et al. 2020).

As a biological pollutant, invasive *E. coli* O157:H7 undoubtedly interacts with biotic variables in soils, as evidenced that indigenous microbial composition (Yao et al. 2014), biodiversity (van Elsas et al. 2012), niche breadth (Mallon et al. 2015a), and fitness difference (Li et al. 2019) showed close linkages to the survival of *E. coli* O157:H7. In this study, biotic and abiotic factors of soils were integrated using multivariate statistical methods to explore the effects of the

interactions of complex environmental factors on *E. coli* O157:H7. The VPA results indicated the influence of bacterial community was interaction with soil properties (49.15%) rather than individual contribution (3.32%) (Fig. 3), and SEM results showed salinity restricted the survival of *E. coli* O157:H7 directly and indirectly, and pH restriction was mainly indirect (Fig. 4). The influence of bacterial community on the survival of *E. coli* O157:H7 might be influenced by the soil physicochemical properties as indicated in a recent study.

Disentangling the survival of *E. coli* O157:H7 was useful to establish its control strategies. Recently, plasma-activated water, a promising alternative of traditional sanitizer, was used to decontaminate fresh produce polluted by pathogens (Ma et al. 2015). Clay was used as an additive in manure compost to shorten the survival of pathogenic bacteria (Awasthi et al. 2019). Our results indicated that appropriate salinity in soil might be beneficial for the control of *E. coli* O157:H7, while soil acidification caused by anthropogenic influences could potentially increase the persistence of *E. coli* O157:H7 in pre-harvest environment. The interactions among biotic and abiotic factors controlling the *E. coli* O157:H7 survival were highlighted, and more verification experiments and in situ experiments in future must be done to further explore the interaction mechanisms between biotic and abiotic factors in survival studies, as well as practical *E. coli* O157:H7 control protocols.

Conclusions

This research revealed that salinity and pH decreased the survival time of *E. coli* O157:H7 in soils through changing indigenous bacterial community to a more adverse state for the survival. The interactions between soil properties and bacterial community explained more variation of *E. coli* O157:H7 survival than soil properties or bacterial community alone, indicating the integration of biotic and abiotic factors in understanding the survival profile of invasive pathogenic *E. coli*. In practice, appropriate soil salinity and controlled soil acidification might be meaningful approaches to restrict *E. coli* O157:H7 in agro-ecosystem.

Authors' contributions Ziming Han: conceptualization, methodology, formal analysis, and writing—original draft preparation; Jincai Ma: methodology, conceptualization, supervision, funding acquisition, and writing—reviewing and editing; Abasiofiok Mark Ibekwe: project administration, funding acquisition, resources, and writing—reviewing and editing; Ching-Hong Yang: methodology and writing—reviewing and editing. All authors read and approved the final manuscript.

Data availability The sequencing data could be downloaded in https://www.ars.usda.gov/ARSUSERFILES/20361500/PUBLIC%20DATA/SALINITY%20PRODUCE%20SEQUENCING%20DATA%202010-2015.ZIP.

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Albert I, Mafart P (2005) A modified Weibull model for bacterial inactivation. Int J Food Microbiol 100:197–211

Anderson IC, Rhodes M, Kator H (1979) Sublethal stress in Escherichia coli: a function of salinity. Appl Environ Microbiol 38:1147–1152

Archer E (2013) Estimate permutation p-values for importance metrics. R package version 1.5.2

Awasthi MK, Chen H, Duan Y, Liu T, Awasthi SK, Wang Q, Pandey A, Zhang Z (2019) An assessment of the persistence of pathogenic bacteria removal in chicken manure compost employing clay as additive via meta-genomic analysis. J Hazard Mater 366:184–191

Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, Huerta-Cepas J, Medema MH, Maltz MR, Mundra S, Olsson PA, Pent M, Põlme S, Sunagawa S, Ryberg M, Tedersoo L, Bork P (2018) Structure and function of the global topsoil microbiome. Nature. 560:233–237

Balamurugan S, Ahmed R, Gao A (2015) Survival of Shiga toxin-producing Escherichia coli in broth as influenced by pH, water activity and temperature. Lett Appl Microbiol 60:341–346

Barberán A, Bates ST, Casamayor EO, Fierer N (2011) Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J 6:343

Bastian M, Heymann S, Jacomy M (2009) Gephi: An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media

Berger CN, Sodha SV, Shaw RK, Griffin PM, Pink D, Hand P, Frankel G (2010) Fresh fruit and vegetables as vehicles for the transmission of human pathogens. Environ Microbiol 12:2385–2397

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73:1045–1055

Breiman L (2001) Random forests. Mach Learn 41:15–32

Buttigieg PL, Ramette A (2014) A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiol Ecol 90:543–550

Byrne BM (2001) Structural equation modeling with AMOS: basic concepts, applications, and programming. Lawrence Erlbaum Associates Publishers, Mahwah

Cai P, Liu X, Ji D, Yang S, Walker SL, Wu Y, Gao C, Huang Q (2018) Impact of soil clay minerals on growth, biofilm formation, and virulence gene expression of Escherichia coli O157:H7. Environ Pollut 243:953–960

Cooley M, Carychao D, Crawford-Miksza L, Jay MT, Myers C, Rose C, Keys C, Farrar J, Mandrell RE (2007) Incidence and tracking of Escherichia coli O157:H7 in a major produce production region in California. PLoS One 2:e1159

Csardi G, Nepusz T (2006) The Igraph software package for complex network research. Int J Complex Syst 1965

Dixon P (2003) VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930

Eichorst SA, Kuske CR, Schmidt TM (2011) Influence of plant polymers on the distribution and cultivation of bacteria in the phylum acidobacteria. Appl Environ Microbiol 77:586–596

Eisenhauer N, Bowker MA, Grace JB, Powell JR (2015) From patterns to causal understanding: Structural equation modeling (SEM) in soil ecology. Pedobiologia 58:65–72

Franz E, van Diepeningen AD, de Vos OJ, van Bruggen AHC (2005) Effects of cattle feeding regimen and soil management type on the fate of Escherichia coli O157 : H7 and Salmonella enterica serovar typhimurium in manure, manure-amended soil, and lettuce. Appl Environ Microbiol 71:6165–6174

Franz E, Semenov AV, Termorshuizen AJ, de Vos OJ, Bokhorst JG, van Bruggen AHC (2008) Manure-amended soil characteristics affecting the survival of E-coli O157 : H7 in 36 Dutch soils. Environ Microbiol 10:313–327

Franz E, Rotariu O, Lopes BS, MacRae M, Bono JL, Laing C, Gannon V, Soderlund R, van Hoek AHAM, Friesema I, French NP, George T, Biggs PJ, Jaros P, Rivas M, Chinen I, Campos J, Jernberg C, Gobius K, Mellor GE, Chandry PS, Perez-Reche F, Forbes KJ, Strachan NJC (2018) Phylogeographic analysis reveals multiple international transmission events have driven the global emergence of Escherichia coli O157:H7. Clin Infect Dis

Geeraerd AH, Valdramidis VP, Van Impe JF (2005) GInaFiT, a freeware tool to assess non-log-linear microbial survivor curves. Int J Food Microbiol 102:95–105

Harrell J, Frank E (2008) Hmisc: Harrell miscellaneous. R package version 3.5-2

Hellberg RS, Chu E (2015) Effects of climate change on the persistence and dispersal of foodborne bacterial pathogens in the outdoor environment: a review. Crit Rev Microbiol:1–25

Hooper D, Coughlan J, Mullen MR (2008) Structural equation modelling: guidelines for determining model fit. Electron J Bus Res Methods 6:53–60

Huang G, Liao J, Han Z, Li J, Zhu L, Lyu G, Lu L, Xie Y, Ma J (2020) Interaction between fungal communities, soil properties, and the survival of invading E. coli O157:H7 in soils. Int J Envriron Re Public Health 17:3516

Kumar GD, Ravi S, Micallef SA, Brown EW, Macarisin D (2017) Aeolian contamination of fruits by enteric pathogens: an unexplored paradigm. Curr Opin Food Sci 19:138–144

Li SP, Tan J, Yang X, Ma C, Jiang L (2019) Niche and fitness differences determine invasion success and impact in laboratory bacterial communities. ISME J 13:402–412

Liang Y, Jiang Y, Wang F, Wen C, Deng Y, Xue K, Qin Y, Yang Y, Wu L, Zhou J, Sun B (2015) Long-term soil transplant simulating climate change with latitude significantly alters microbial temporal turnover. ISME J 9:2561–2572

Liaw A, Wiener M (2002) Classification and Regression by RandomForest. R News:18–22

Lin JS, Smith MP, Chapin KC, Baik HS, Bennett GN, Foster JW (1996) Mechanisms of acid resistance in enterohemorrhagic Escherichia coli. Appl Environ Microbiol 62:3094–3100

Ma J, Ibekwe AM, Yi X, Wang H, Yamazaki A, Crowley DE, Yang C-H (2011) Persistence of Escherichia coli O157:H7 and Its Mutants in Soils. PLoS One 6

Ma J, Ibekwe AM, Crowley DE, Yang C-H (2012a) Persistence of Escherichia coli O157:H7 in major leafy green producing soils. Environ Sci Technol 46:12154–12161

Ma J, Ibekwe AM, Leddy M, Yang C-H, Crowley DE (2012b) Assimilable organic carbon (AOC) in soil water extracts using Vibrio harveyi BB721 and its implication for microbial biomass. PLoS One 7

Ma J, Ibekwe AM, Yang C-H, Crowley DE (2013) Influence of bacterial communities based on 454-pyrosequencing on the survival of Escherichia coli O157:H7 in soils. FEMS Microbiol Ecol 84:542–554

Ma R, Wang G, Tian Y, Wang K, Zhang J, Fang J (2015) Non-thermal plasma-activated water inactivation of food-borne pathogen on fresh produce. J Hazard Mater 300:643–651

Mallon CA, Poly F, Le Roux X, Marring I, van Elsas JD, Salles JF (2015a) Resource pulses can alleviate the biodiversity-invasion relationship in soil microbial communities. Ecology 96:915–926

Mallon CA, van Elsas JD, Salles JF (2015b) Microbial invasions: the process, patterns, and mechanisms. Trends Microbiol 23:719–729

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Munro PM, Flatau GN, Clement RL, Gauthier MJ (1995) Influence of the Rpos (Katf) sigma-factor on maintenance of viability and culturability of Escherichia-Coli and Salmonella-Typhimurium in seawater. Appl Environ Microbiol 61:1853–1858

Ongeng D, Muyanja C, Geeraerd AH, Springael D, Ryckeboer J (2011) Survival of Escherichia coli O157:H7 and Salmonella enterica serovar Typhimurium in manure and manure-amended soil under tropical climatic conditions in Sub-Saharan Africa. J Appl Microbiol 110:1007–1022

Ongeng D, Geeraerd AH, Springael D, Ryckeboer J, Muyanja C, Mauriello G (2015) Fate of Escherichia coli O157:H7 and Salmonella enterica in the manure-amended soil-plant ecosystem of fresh vegetable crops: a review. Crit Rev Microbiol 41:273–294

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87:2614–2625

Ritchie JM, Campbell GR, Shepherd J, Beaton Y, Jones D, Killham K, Artz RRE (2003) A stable bioluminescent construct of Escherichia coli O157 : H7 for hazard assessments of long-term survival in the environment. Appl Environ Microbiol 69:3359–3367

Rosseel Y (2012) lavaan : an R package for structural equation modeling. J Stat Softw 48

Schermelleh-Engel K, Moosbrugger H, Müller H (2003) Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods Psychol Res 8:23–74

Shabala L, Bowman J, Brown J, Ross T, McMeekin T, Shabala S (2009) Ion transport and osmotic adjustment in Escherichia coli in response to ionic and non-ionic osmotica. Environ Microbiol 11:137–148

Smouse PE, Long JC, Sokal RR (1986) Multiple-regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 35:627–632

Teplitski M, de Moraes M (2018) Of mice and men....and plants: comparative genomics of the dual lifestyles of enteric pathogens. Trends Microbiol

Turner K, Moua CN, Hajmeer M, Barnes A, Needham M (2019) Overview of leafy greens-related food safety incidents with a California link: 1996 to 2016. J Food Prot 82:405–414

van Elsas JD, Semenov AV, Costa R, Trevors JT (2011) Survival of Escherichia coli in the environment: fundamental and public health aspects. ISME J 5:173–183

van Elsas JD, Chiurazzi M, Mallon CA, Elhottova D, Kristufek V, Salles JF (2012) Microbial diversity determines the invasion of soil by a bacterial pathogen. Proc Natl Acad Sci U S A 109:1159–1164

Wang HZZ, Wei G, Yao ZYY, Lou J, Xiao KCC, Wu LSS, Wu JJJ, Xu JMM (2014) Response of Escherichia coli O157:H7 survival to pH of cultivated soils. J Soils Sediments 14:1841–1849

Westphal A, Williams ML, Baysal-Gurel F, LeJeune JT, Gardener BBM (2011) General suppression of Escherichia coli O157:H7 in sand-

based dairy livestock bedding. Appl Environ Microbiol 77:2113–2121

Xing J, Wang H, Brookes PC, Salles JF, Xu J (2019) Soil pH and microbial diversity constrain the survival of E-coli in soil. Soil Biol Biochem 128:139–149

Yao Z, Wei G, Wang H, Wu L, Wu J, Xu J (2013) Survival of Escherichia coli O157:H7 in soils from vegetable fields with different cultivation patterns. Appl Environ Microbiol 79:1755–1756

Yao Z, Wang H, Wu L, Wu J, Brookes PC, Xu J (2014) Interaction between the microbial community and invading Escherichia coli O157:h7 in soils from vegetable fields. Appl Environ Microbiol 80:70–76

Yao Z, Yang L, Wang H, Wu J, Xu J (2015) Fate of Escherichia coli O157: H7 in agricultural soils amended with different organic fertilizers. J Hazard Mater 296:30–36

Zhu Y-G, Zhao Y, Li B, Huang C-L, Zhang S-Y, Yu S, Chen Y-S, Zhang T, Gillings MR, Su J-Q (2017) Continental-scale pollution of estuaries with antibiotic resistance genes. Nat Microbiol 2:16270