Journal of Applied Microbiology

Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

Associations between *Escherichia coli* O157 shedding and the faecal microbiota of dairy cows

C. Stenkamp-Strahm, C. McConnel, S. Magzamen, Z. Abdo and S. Reynolds

College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA

Keywords
dairy, ecology, epidemiology, *Escherichia coli* O157, modelling.

Correspondence
Chloe Stenkamp-Strahm, CVMBS, Colorado State University, 1644 Campus Delivery, Fort Collins, CO 80523, USA.
E-mail: cstenkam@lamar.colostate.edu

2017/1714: received 29 August 2017, revised 1 December 2017 and accepted 15 December 2017

doi:10.1111/jam.13679

Abstract

Aims: Dairy cattle shed pathogenic *Escherichia coli* O157 (O157) in faeces, playing a role in human exposure. We aimed to measure faecal microbial communities in early lactation dairy cattle, and model outcomes with O157 shedding metrics.

Methods and Results: Daily faecal samples were collected from 40 cattle on two Colorado dairies for five consecutive days, and characterized for O157. 16S rRNA gene sequencing was used to measure sample-level microbial communities. Alpha-diversity metrics were associated with O157 outcomes via regression modelling, adjusting for confounders. Differential abundance of taxa were identified between O157(+) and O157(−) samples and between shedding days of individuals, using matched Wilcoxon rank-sum tests, zero-inflated Gaussian (ZIG) regression and negative binomial regression. After removing an outlier, multi-day and intermittently shedding cows had lower average richness compared to those that never shed. ZIG modelling revealed *Bacillus coagulans* to be more abundant in O157(−) samples, while *Moryella* were more abundant in O157(+) samples. Negative binomial models and Wilcoxon tests revealed no differentially abundant taxa between O157(+) *vs* O157(−) samples, or between shedding days of individuals.

Conclusions: Microbial diversity and some taxa may be influenced by or affect O157 shedding by dairy cattle.

Significance and Impact of the Study: If future work corroborates these findings, dairy cow microbial community changes may be used to guide on-farm strategies that mitigate O157 dissemination, protecting the human food chain.

Introduction

Pathogenic *Escherichia coli* O157 (O157) is a food safety pathogen that causes an estimated 2 801 000 cases of human disease each year, some of which lead to haemolytic uraemic syndrome, end-stage renal disease and death (Majowicz *et al.* 2014). Asymptomatically shedding ruminants are linked to a majority of these illness events (Nguyen and Sperandio 2012). Results from national studies suggest that O157 is ubiquitous in the environment of US dairy farms, implicating dairy cattle as principle O157 reservoirs (Wells *et al.* 1998; CEAH, USDA 2003). Despite current control measures (e.g. Food Safety and Inspection Service standard operating procedures for slaughter, laws for dairy pasteurization, United States Department of Agriculture codes for food preparation) O157 from dairies still regularly enters the human food supply (Croxen *et al.* 2013). As the human food chain is vast and complicated, controlling O157 at the dairy cattle source is a logical strategy to reduce the burden of human disease.

In the last decade, there has been much interest in characterizing the pattern in which cows shed O157 in their faeces (Shere *et al.* 1998; Robinson *et al.* 2004). Although cows in US dairies are exposed to O157 in the surrounding environment, only a portion of animals will presumably become colonized with the bacteria after ingestion, and intermittently shed detectable faecal

quantities (Cobbold *et al.* 2007). Cattle do not mount an immune response to O157, and lack formal clinical signs during carriage. This equates to an inability to discern shedding from nonshedding cattle without laboratory-based bacterial enrichment and characterization of faecal samples (Paton and Paton 1998; Wang *et al.* 2002). These difficulties and the intermittent and variable nature of shedding have required strategic design for risk factor studies when looking at naturally colonized herds. Regardless, previous evaluation of O157 risk factors has revealed that environment and host-level factors, including humidity, temperature, parity, lactation days, prior treatment, disease status and the specific operation, influence a given dairy cow's risk of having an O157 shedding event (Cobbaut *et al.* 2009; Menrath *et al.* 2010; Williams *et al.* 2015; Venegas-Vargas *et al.* 2016 Stenkamp-Strahm *et al.* 2017a). Given these findings, there remains a need to identify interventions and preventive strategies that may be implemented to reduce the number of shedding events in at-risk individuals.

Recently, next-generation sequencing of the 16S rRNA gene has led to an increase in the number of microbial community surveys (MCS) measuring gastrointestinal (GI) microbes, and their influence on pathogen and disease presence (Cho and Blaser 2012). Most of these studies define changes in microbial community (beta diversity) among study environments via ordination methods, and tie these conclusions to environmental metadata. Studies of this nature in dairy cows have defined faecal communities and those within the rumen, focusing on community structure and its relation to animal growth, development and milk production (Jami and Mizrahi 2012; Jewell *et al.* 2015; Mao *et al.* 2015; Dill-McFarland *et al.* 2017). Animal-level characteristics identified as risk factors for O157 shedding have been shown through diverse mammalian MCS studies to be correlated with changes in the GI microbiota (Sekirov *et al.* 2008; Buffie and Pamer 2013; Carrothers *et al.* 2015; Chevalier *et al.* 2015). Although the GI microbiota in dairy cows has been studied previously, it has not been thoroughly evaluated relative to O157 colonization and the cow life-history features previously shown to impact shedding events.

Identifying specific taxa that are associated with shedding might discern whether the microbiota itself should be defined as a risk factor, or be considered as an O157 treatment target. In many MCS studies, including those of cows, relative changes in abundance of operational taxonomic units (OTUs) are measured after sequence counts are rarefied to the level of the lowest sample, commonly omitting large percentages of sequence data (McMurdie and Holmes 2014). With the progression of microbiome research has also come the development of analytic approaches that incorporate epidemiologic measures and metadata relating these communities to disease outcomes. It is also possible to measure sample-specific differential abundance of taxa without omitting sequence data, using robust normalization and modelling of data distributions (Thorsen *et al.* 2016). These different analytic approaches are advantageous when using MCS data to understand the associations between dairy cow microbial communities and the presence of O157. Given the considerations above, we postulated that the dairy cow GI microbiota would be associated with O157 colonization and shedding events. Specifically, we hypothesized that GI microbial diversity would be different between O157 shedding and nonshedding individuals, and that the presence of certain taxa would be associated with the presence of O157.

Materials and methods

Study population

Two commercial, freestall dairies (Farm 1, Farm 2) in northern Colorado and representing a combined population of 2350 lactating cattle participated in this pilot study. Both dairies consisted of predominantly Holstein Friesian cows fed total mixed rations (TMR) formulated by the same nutritionist and with nutrient contents similar to one another. Forage was predominantly corn silage and alfalfa hay with additional carbohydrates provided in the form of hominy, brewer's grain, distiller's grain and wheat middlings. Each TMR was supplemented with vitamin and trace minerals designed for high-producing cows and Diamond V original line XPC supplement (Cedar Rapids, IA). Early lactation cows were targeted for the study, as they are a cohort at higher risk for O157 shedding, and this enabled our detection of both shedding and nonshedding individuals (Mechie *et al.* 1997; Venegas-Vargas *et al.* 2016).

Our sampling design was employed to assess the change in microbial communities between all O157-positive and -negative samples, and between shedding events and shedding patterns of individual cows. On sample day 1, all cows within the first 21 days postpartum on each dairy ($n = 74$) were sampled by obtaining >10 g faeces via rectal palpation. Samples were kept on ice prior to laboratory characterization. On-farm record systems were used to gather animal life-history features: lactation number (parity), days in milk (DIM), disease during current lactation and disease treatments during current lactation (Dairy Comp 305™, Valley Agricultural Software, Tulare, CA; DHI-Plus, DHI Computing Service Inc., Provo, UT). Diseases recorded included retained placenta, mastitis, metritis, fever of unknown origin, pneumonia, enteric

disease, dystocia (including severity), ketosis and lameness. Recorded treatments included penicillin, oxytetracycline, ceftiofur, flunixin meglumine, drench (oral electrolytes) and propylene glycol.

Laboratory enrichment and latex agglutination procedures described below were used to identify bacterial isolates containing the *rfb* (O157) gene in day 1 faecal samples. Based on those preliminary results, 10 cows that shed O157 on day 1, and 10 cows that did not, were selected per farm and tested in the same fashion for five consecutive days (n = 40 cows, n = 200 samples).

Laboratory characterization of O157

O157 isolation was performed via selective enrichment and detection 'gold standard' procedures with slight modification, as described previously (Stenkamp-Strahm *et al.* 2017a,b). Briefly, samples were mixed 1 : 10 in buffered peptone water (BPW) for both enrichment and initial direct plating. One hundred microlitre was spread plated on sorbitol MacConkey agar with BCIG (Oxoid Diagnostic Reagents, Basingstoke, Hampshire, UK) containing 1·25 mg potassium tellurite and 0·025 mg cefixime (CT-SMAC-BCIG; HiMedia Laboratories, Mumbai, India). These direct plates were incubated at 37°C for 24 h (March and Ratnam 1986). As pathogenic O157 has been known to adapt a sorbitol fermenting phenotype within 24 h, 'suspect' O157 colonies seen on plates throughout experiments were deemed as those with straw, grey, pink–grey or too small/difficult to characterize colony coloration (Schmidt *et al.* 1999; Ayaz *et al.* 2014).

Direct plates containing ≥100 suspect colonies after incubation were chosen for latex agglutination. Three to 15 colonies per plate were tested for O157 by agglutination using an *E. coli* O157 latex kit, following manufacturer's instructions (Oxoid Diagnostic Reagents). Positive colonies were enriched in BPW for 6 h and stored at −80°C in 10% sterile glycerol. For PCR experiments, 10 μl of thawed isolates were centrifuged at 5000 ***g*** for 5 min and re-suspended in 30 μl molecular grade water. A volume of 5 μl re-suspended culture template was placed into Qiagen Multiplex PCR Plus Kit reactions, according to the manufacturer's instructions (Qiagen, Venlo, the Netherlands). In brief, each 25-μl PCR reaction consisted of 12·5 μl master mix, 2·5 μl primer mix containing 0·2 μmol l^{-1} each primer, 5 μl molecular grade water and 5 μl culture template. The thermal cycling conditions consisted of an initial incubation at 95°C for 5 min to activate the polymerase, followed by 40 cycles of amplification with denaturation at 95°C for 30 s, annealing at 57°C for 1 min and 30 s and extension at 72°C for 30 s, ending with a final extension at 68°C for 10 min. Thermocycling was performed using a Bio-Rad S1000 Thermal Cycler (Bio-Rad, Sydney, Australia). PCR products were analysed by agarose gel electrophoresis using a 2% agarose gel (Lonza Group Ltd., Basel, Switzerland).

The faecal dilution remaining after direct plating was enriched for 6 h at 37°C. Enriched samples not confirmed as O157 positive through direct plating were subjected to immunomagnetic separation (IMS) using Dynabeads anti-*E. coli* O157 and a BeadRetriever System (Life Technologies, Oslo, Norway). IMS samples were subsequently plated onto CT-SMAC-BCIG and incubated for 24 h at 37°C. Suspect colonies were confirmed by latex agglutination and PCR targeting O157 *rfb* (O-antigen) gene (Wang *et al.* 2002). All *rfb*-positive preliminary isolates were subsequently PCR tested for *stx1*, *stx2* (shiga toxin genes) and *eaeA* (a variant of the *eae* intimin gene) using the same PCR protocol outlined above (primers; Paton and Paton 1998). O157 were not enumerated after IMS detection and PCR confirmation.

Samples were deemed to be enterohaemorrhagic *E. coli* (EHEC) O157 positive when isolates contained *rfb*, *eaeA* and any *stx* genes. Samples were deemed to be atypical enteropathogenic *E. coli* (aEPEC) O157 positive when isolates contained *rfb* and *eaeA* genes. For subsequent statistical analysis, all pathogenic strains (EHEC and aEPEC) were considered together and are hereafter referred to as 'O157'.

Library preparation and 16S sequencing

DNA library preparation and 16S rRNA sequencing were performed at the Argonne National Laboratory and followed Earth Microbiome Project suggested protocols (Gilbert *et al.* 2014; www.earthmicrobiome.org, accessed August 2015). Briefly, genomic DNA was extracted from faecal samples stored at −80°C using the PowerSoil DNA Isolation Kit (MoBio/Qiagen, Carlsbad, CA). To support pooling of all collected samples during a paired-end 2 × 150-base pair Illumina sequencing run, the amplification primer set contained nine extra bases in the adapter region of the forward primer and a 12-base Golay barcode sequence in the reverse amplification primer (Caporaso *et al.* 2012). For amplification of the V4 hypervariable region of the 16S rRNA gene, 515F (5′-GTGYCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACNVGGGTWTCTAAT-3′) primers with the defined barcodes and Illumina flow cell adapter sequences were used for amplification. Each 25-µl PCR reaction contained 12 µl of certified DNA-free water (MoBio/Qiagen), 10 µl of 5-Prime HotMasterMix (1x, Quanta Biosciences, Beverly, MA), 1 µl of forward primer (5 μmol l^{-1} concentration, 200 pmol l^{-1} final), 1 µl Golay Barcode Tagged Reverse Primer (5 µmol l^{-1} concentration,

200 pmol l^{-1} final) and 1 μl of genomic DNA. The PCR conditions were as follows: 94°C for 3 min, 35 cycles of 94°C for 45 s, 50°C for 60 s, 72°C for 90 s and a final single extension of 10 min at 72°C. Following PCR, amplicons were quantified using a Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Waltham, MA). Based on different quantification values, volumes of each sample were pooled to achieve equal representation. Pools were cleaned using the UltraClean® PCR Clean-Up Kit (MoBio/Qiagen) and quantified *de novo*. Pool molarity was determined, and the pool was then diluted to 2 nmol l^{-1}, before denaturing with NaOH. The sample was then repeat diluted to a final concentration of 2 pmol l^{-1}. A 30% PhiX spike was added, prior to loading on an Illumina HiSeq sequencer. Samples were sequenced using a 300-cycle V2 reagent cartridge (Illumina, San Diego, CA).

Sequence preprocessing and OTU selection

Sequence read quality was analysed via FastQC. Raw fastq files were demultiplexed with a maximum barcode error of 0 using default methods in QIIME (Caporaso *et al.* 2010b). Reads not assigned via barcode were removed. QIIME was then used to preprocess reads to OTUs. During this approach, reads were clustered using USEARCH, and chimeras were removed using the UCHIME algorithm (ver. 9.0, Edgar 2010). Reads were aligned to the Greengenes core alignment with 97% sequence identity using PyNAST (DeSantis *et al.* 2006; Caporaso *et al.* 2010a). Taxonomy was assigned using the Ribosomal Database Project 2.2 classifier and the Greengenes 13.8 taxonomy reference (Wang *et al.* 2007; McDonald *et al.* 2012). A maximum likelihood approximated tree was built using FastTree 2.1.9 (Price *et al.* 2010).

During this preprocessing, an open reference approach was utilized with the packages listed above (Rideout *et al.* 2014). The open reference algorithm allowed for an initial clustering and closed reference sequence alignment, followed by *de novo* clustering of reads that failed to align. During the second clustering, representative sequences were created using the centroid of those that failed to align, and these were used during a closed reference picking process. OTUs that were only represented by a single read were discarded.

Nonmetric multi-dimensional scaling with a Bray–Curtis distance was used to ordinate OTUs from O157-positive and -negative samples in space. An unweighted unifrac distance was used to perform a principal coordinate analysis of OTUs from O157-positive and -negative samples. The relative abundance of taxa within each sample was measured using the R package Phyloseq (McMurdie and Holmes 2013).

Statistical analysis

O157 categories were used to classify the presence of O157 at either the sample or the cow level. These categories are listed in Table 1 and include for the sample level: pathotype, day prior *vs* day of and day of *vs* day after. 'Pathotype' classified samples as containing aEPEC/EHEC, or no pathogenic O157. For 'day prior *vs* day of' and 'day of *vs* day after' samples were paired to compare either the day prior to shedding or the day after shedding to the day of the shedding event in that individual. At the cow level, categories included pattern and ever *vs* never. 'Pattern' was a category that defined cows as multi-day (≥2 days shed), intermittent (1 day shed) or never shedders (0 days shed). 'Ever *vs* never' classified cows as having shed or not shed pathogenic O157 at least once during the study period. For 'pattern' and 'ever *vs* never', each cow was classified the same across days.

Using these samples and cow defined categories, two different statistical approaches were used to analyse data from microbial count tables: epidemiologic modelling of microbial alpha diversity and measurement of taxa

Table 1 Definition of O157 categories used to classify cows and samples for subsequent statistical analyses

Category name	Measure	Levels of variable	Analysis
Pathotype	Sample	No O157 O157	Epidemiologic modelling, differential abundance testing
Ever *vs* never	Cow	Never shed O157 Shed O157 at least one time	Epidemiologic modelling
Pattern	Cow	Never shed O157 Shed O157 one time (intermittent) Shed O157 ≥2 times (multi-day)	Epidemiologic modelling
Day prior *vs* day of (matched)	Sample	Day prior to a shedding day Shedding day	Differential abundance testing
Day of *vs* day after (matched)	Sample	Shedding day Day after a shedding day	Differential abundance testing

differential abundance. These approaches are described, in turn, below. As taxa within microbial communities are defined at the sample level, for cow-level categories (pattern, ever *vs* never) only the epidemiologic modelling of diversity measures was used. For matched samples (day of *vs* day after, day prior *vs* day of), only measuring differential abundance of taxa was used.

Epidemiologic modelling of microbial diversity

Relative abundance of OTUs aggregated to different taxa levels were quantified using Phyloseq. Alpha-diversity metrics were quantified using Phyloseq and Vegan, including observed richness (R), Shannon's index (S) and Pielou's measure of species evenness ((S)/Log(R)) (McMurdie and Holmes 2013; Oksanen *et al.* 2017). Previous studies have shown that rarefying 16S rRNA count data prior to analysis is necessary to control for sequencing depth, but can negatively influence results (McMurdie and Holmes 2014). For analyses, data were normalized using cumulative sum scaling (CSS) in metagenomeSeq to correct for this bias (Paulson *et al.* 2013).

Random effects, logistic and multinomial modelling were completed to measure the association between O157 categories and diversity measures. In brief, for the category 'pathotype', random effects regression was used with sample-level diversity measures, and a random effect for cow. For the category 'ever *vs* never', a logistic regression model was used with diversity measures averaged across days for each individual cow. For the 'pattern' analysis, multinomial models were used with diversity measures averaged across days for each individual cow. Separate animal-level covariates that modify the risk of shedding in dairy cows have been reported previously (Menrath *et al.* 2010; Venegas-Vargas *et al.* 2016; Stenkamp-Strahm *et al.* 2017a). As these variables may cause, confound or mediate the association between alpha diversity and O157 shedding status, a directed acyclic graph was used to evaluate the role of these factors in the association between O157 categories and diversity (Fig. S1). The distribution of the variables parity, days in milk, disease status, farm and treatment were measured across O157 categories. These variables were assessed individually for associations with diversity, and those that met the screening criteria ($P \leq 0{\cdot}2$) were included in the O157 models of diversity. Odds ratios and confidence intervals were calculated from model coefficients and standard errors using the interquartile range (IQR) of diversity values.

Differential abundance testing

Using metagenomeSeq, taxa were aggregated to the family, genus and species levels (Paulson *et al.* 2013). Aggregated tables were filtered to reflect taxa present in at least 25–30% of samples, and differential abundance at each taxonomic level was measured in turn. The ideal methodology used to test for differential abundance between nonpaired 16S rRNA count tables is not agreed upon in the literature (McMurdie and Holmes 2014; Thorsen *et al.* 2016; Weiss *et al.* 2017); therefore, two modelling approaches were used to assess changes in abundance for the 'pathotype' metric. In metagenomesSeq, zero-inflated Gaussian (ZIG) regression was used to evaluate changes between CSS normalized count tables (Paulson *et al.* 2013). In DESeq2, negative binomial regression was used to evaluate the same changes in count tables internally normalized by calculating geometric means and median count ratios (Love *et al.* 2014). For matched samples ('day prior *vs* day of' and 'day of *vs* day after'), aggregated count tables were normalized with CSS and compared using a Wilcoxon rank-sum test on table differences. After adjusting for multiple testing using the Benjamini and Hochberg (1995) correction, a *P*-value cutoff of $<0{\cdot}1$ was used to detect differences. This *P*-value was selected due to the preliminary nature of the project, and relatively low power.

All procedures were approved by Colorado State University's Institutional Animal Care and Use Committee. All statistical analyses were performed in R ver. 3.3.2 or later (R Core Team 2016).

Results

O157 isolates and initial abundance data

The study population consisted of 40 cows (20 per farm) that were sampled for five consecutive days. Of these individuals, 14 (35%) shed EHEC, 4 (10%) shed aEPEC and 31 (78%) shed *rfb* isolates without virulence genes. Samples taken on day 1 of the study were preliminarily assessed for *rfb* (38% shed rfb+ strains, 62% did not), prior to continued sampling of O157-positive and -negative cows on days 2–5. PCR characterization of isolates revealed many fewer cows harbouring O157 with virulence genes (EHEC and aEPEC) than those identified as *rfb* positive on day 1. Of the 200 study samples, 20 (10%) contained EHEC isolates, 4 (2%) contained aEPEC isolates and 67 (34%) contained *rfb* isolates that did not have virulence genes.

During 16S library preparation of faecal sample DNA, four samples either did not amplify well during PCR or did not sequence adequately, and were omitted from further analysis. Of these samples, three were from Farm 1 and one was from Farm 2. None of the omitted samples were identified as having aEPEC or EHEC. After sequencing and prior to OTU analysis in QIIME, all Illumina DNA sequence reads were seen to have an average quality PHRED score $\geq$30 via FASTQC. The data set revealed a

total count of 12 225 598 reads, and 159 354 unique OTUs. Mean sequence read count per sample was 39 059.

Bacterial communities and member taxa were strikingly similar among cow samples, and did not cluster in ordination space based on O157 presence or absence (Fig. 1). Measures related to animal life-history characteristics and other shedding categories also did not show clustering of samples in space. Relative abundances of taxa at the family and phylum levels further revealed that communities with and without O157 were similar (Fig. 1a,b). Differences, where noted, were only slight. O157-positive and -negative bacterial communities contained 62 and 56% Firmicutes, 35 and 31% Bacteroidetes and 2·2 and 2·4% Tenericutes respectively. Actinobacteria was slightly higher in non-O157 samples (2·5 *vs* 1·9% in O157-positive samples) as

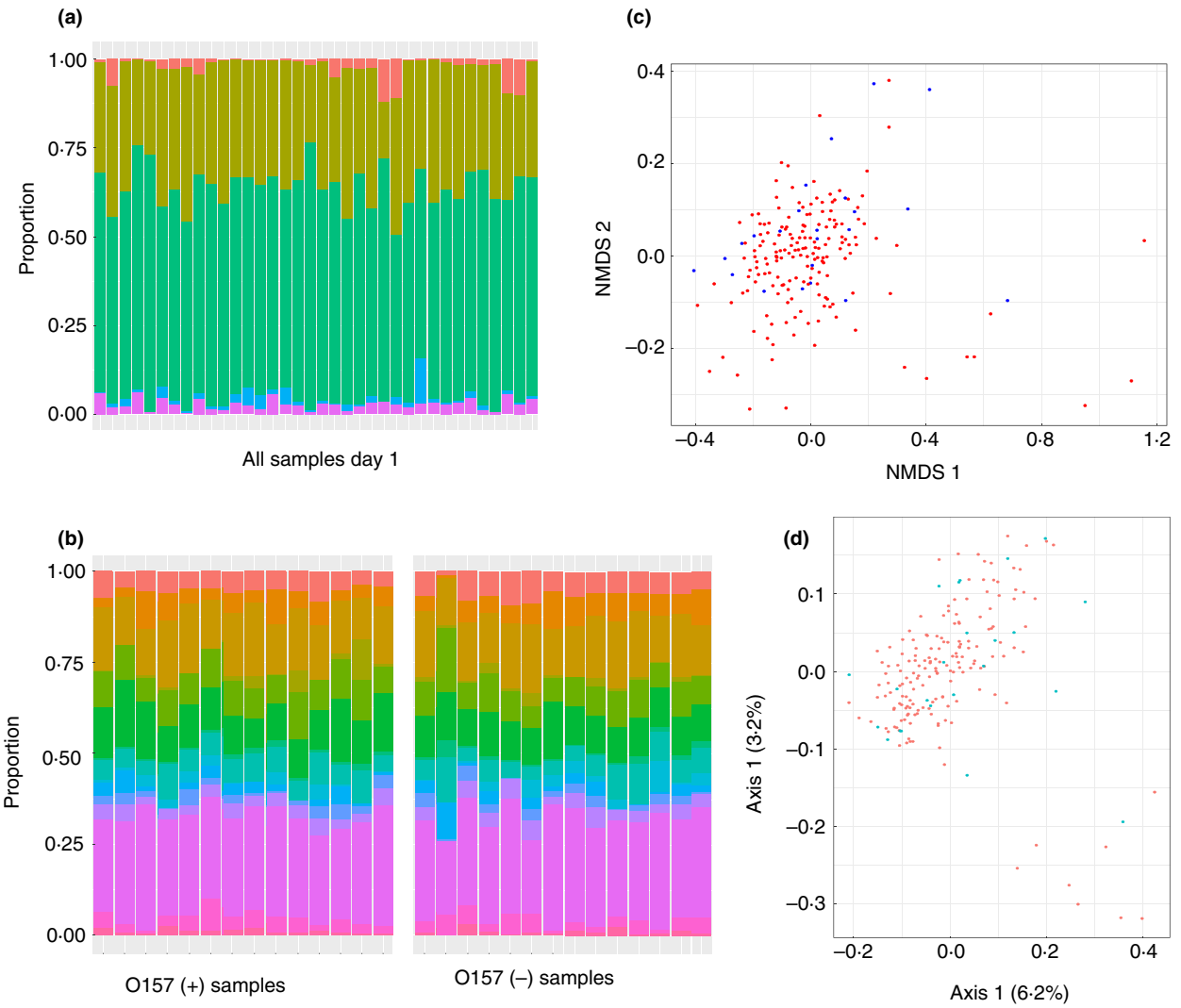

Figure 1 Microbial communities of dairy cow faecal samples were mostly uniform. Proportions of microbial community members at the phylum level among all cow samples taken on the initial day of the study (a) showed low variability. Proportions of community members at the family level, in a subset of O157-positive and -negative samples taken throughout the study (b; one sample per cow), additionally showed that taxa were not dependent on O157 pathogen status. Looking at communities via ordination with NMDS and Bray–Curtis. (a) Phylum: (■) Actinobacteria; (■) Bacteroidetes; (■) Firmicutes; (■) Spirochaetes; (■) Tenericutes. (b) Family: (■) Paraprevotellaceae; (■) Bacteroidaceae; (■) Bacteroidales-unclassified; (■) Bifidobacteriaceae; (■) Clostridiaceae; (■) Clostridiales-unclassified; (■) Erysipelotrichaceae; (■) Lachnospiraceae; (■) p-2534-18B5; (■) Peptostreptococcaceae; (■) RF39-unclassified; (■) Rikenellaceae; (■) Ruminococcaceae; (■) S24-7; (■) Spirochaetaceae (c; status of O157, O157 positive (●), O157 negative (●)) or PCoA with unweighted Unifrac (d; status of O157, O157 positive (●), O157 negative (●)) revealed a lack of clustering in samples, and a spatial resolution that was not related to the presence or absence of O157. [Colour figure can be viewed at wileyonlinelibrary.com]

were Spirochaetes (1·6 *vs* 1·0% in O157-positive samples). Proteobacterial percentages were 4·2% in non-O157 samples and 4·6% in O157 containing samples.

When looking at community composition at the family level, several taxa were seen to be slightly higher in abundance in non-O157 samples compared to O157 containing samples: Paraprevotellaceae (5·4 *vs* 4·7% in O157 positive), Bacteroidaceae (4·7 *vs* 3·6% in O157 positive), Bifidobacteraceae (2·4 *vs* 1·7% in O157 positive) and Spirochaetaceae (1·5 *vs* 1·0% in O157 positive). Meanwhile, Lachnospiraceae (6·0% in non-O157 *vs* 7·1% in O157 positive), Erysipelotrichaceae (0·9% in non-O157 *vs* 1·4% in O157 positive) and Peptostreptococcaceae (2·2% in non-O157 *vs* 2·8% in O157 positive) were slightly higher in O157-positive samples.

Epidemiologic modelling of microbial diversity

Five different variables described previously to influence a dairy cow's risk of shedding pathogenic O157 (farm, DIM, parity, disease status and treatment status) may confound or mediate the association between microbial diversity and O157. Cows in the current study experienced ketosis ($n = 5$), metritis ($n = 4$), mastitis ($n = 2$), retained placenta ($n = 2$), lameness ($n = 2$), enteric disease ($n = 2$), fever of unknown origin ($n = 1$) and pneumonia ($n = 1$). The variable 'disease' was collapsed into a binary yes/no category due to sparseness of most disease types (no disease: $n = 21$; disease: $n = 19$). Treatments given to cows in the current study included penicillin ($n = 3$), oxytetracycline ($n = 5$), ceftiofur ($n = 7$), flunixin meglumine ($n = 2$), drench ($n = 3$) and propylene glycol ($n = 4$). Due to sparseness of treatment types, the variable 'treatment' was collapsed into a categorical variable with three levels. These levels represented cows that were never treated ($n = 23$), cows that were ever treated with any antibiotic (penicillin, oxytetracycline, ceftiofur; $n = 11$) and cows that were not treated with an antibiotic but were treated with a nonantibiotic agent (flunixin meglumine, propylene glycol, drench; $n = 6$). Enrolled cattle were parity 1 ($n = 19$), parity 2 ($n = 7$), parity 3 ($n = 8$), parity 4 ($n = 3$), parity 5 ($n = 1$) and parity 6 ($n = 2$). The variable parity was also collapsed into three categories due to sparseness of cells and biological reasoning (parity 1, $n = 19$; parity 2, $n = 7$; parity ≥ 3, $n = 14$).

All of the variables known to influence shedding of O157 were grouped by O157 categories, described in Table 2. Of these variables, farm, parity and treatment differed by O157 'pathotype' status (Table 2; $P \leq 0{\cdot}08$). Treatment and parity had variable distributions across 'pattern' categories ($P = 0{\cdot}03$ and $P = 0{\cdot}33$ respectively) and treatment varied between categories of 'ever *vs* never' ($P = 0{\cdot}13$).

Richness, evenness and Shannon's diversity were computed for all samples using normalized 16S rRNA read counts. The IQR of these values by O157 categories are presented in Fig. 2. In multinomial models, cows that were classified as multi-day shedders were seen to have lower average richness (Fig. 3; OR = 0·51; 95% CI: 0·41–0·64) compared to cows that never shed during the study period. Odds ratios and confidence intervals were nonsignificant for other diversity models comparing O157 categories.

Variables that may confound or mediate the association between O157 categories and diversity measures (Fig. S1, Table 2) were assessed for individual associations with richness, Shannon's and evenness values (Table S1). Variables with $P < 0{\cdot}2$ model results were considered further in influencing the regression outcomes between O157 category and diversity seen in Fig. 3. These variables included parity, farm and treatment, and were the same variables seen to be different when distributed by O157 shedding metrics (Table 2).

These three variables were included in models of O157 and alpha diversity. Each variable was first assessed by itself with an O157 metric, and then in turn with other variables. For most models, the nonsignificant associations seen in the original crude values were not altered with adjustment (Table S2). Cows with a multi-day shedding pattern still had lower average richness compared to never shedding cows, regardless of adjusting with any or all variables. However, when the variable parity was added to pattern models with any combination of farm, treatment or no other confounders, cows with an intermittent shedding pattern were seen to have significantly higher average richness compared to never shedding cows (Table S2; parity added alone: OR = 1·23; 95% CI: 1·07–1·40). Due to the discrepant nature of these conclusions, raw values of richness for cows of all parity levels were evaluated. A single intermittently shedding, parity 1 cow had much higher richness values than nearly all other cows in the study, including on the day she shed O157. The average richness value for this cow on days 1–5 was 7644, compared to average values that fell in the range of 3000–5000 for other cows. Given the relatively small sample size in this pilot project, it was hypothesized that this individual was driving the increase in average richness seen between intermittently and never shedding cows, when controlling for parity. A sensitivity analysis was performed removing this outlier (Fig. 4). When re-running models of richness and O157 pattern adjusting for confounders, intermittently shedding cows were seen to have significantly lower average richness (OR = 0·568; 95% CI: 0·49–0·66) compared to never shedding cows. Similar to results when the single high-richness cow was included, multi-day shedding cows still had significantly lower average richness (OR = 0·415; 95% CI: 0·33–0·52) compared to never shedding cows.

Table 2 Herd and cow-level descriptors by O157 pathotype and shedding status

	Pathotype, *n* = sample (196)		
	No O157	aEPEC/EHEC	*P*-value
Farm*			
1	83 (84%)	16 (16%)	
2	90 (93%)	7 (7%)	0·08
Days in milk (median (IQR))§	9 (7·0)	8 (7·5)	0·752
Parity†			
1	83 (88%)	11 (12%)	
2	27 (77%)	8 (23%)	
≥3	63 (94%)	4 (6%)	0·047
Disease*			
Absent	92 (90%)	10 (10%)	
Present	81 (86%)	13 (14%)	0·51
Treatment†			
None	100 (90%)	11 (10%)	
Antibiotic	53 (96%)	2 (4%)	
Other	20 (66%)	10 (33%)	0·0004
	Ever/never, *n* = cow (40)		
	Never shed	Shed ≥1 time	
Farm*			
1	11 (55%)	9 (45%)	
2	13 (65%)	7 (35%)	0·75
Days in milk (SD)‡	8·04 (4·54)	7·06 (4·54)	0·51
Parity†			
1	12 (63%)	7 (37%)	
2	3 (37%)	5 (63%)	
≥3	9 (69%)	4 (31%)	0·4
Disease*			
Absent	12 (57%)	9 (43%)	
Present	12 (63%)	7 (37%)	0·95
Treatment†			
None	13 (57%)	10 (43%)	
Antibiotic	9 (82%)	2 (18%)	
Other	2 (33%)	4 (66%)	0·13
	Pattern, *n* = cow (40)		
	Never shed	Intermittent	Multi-day
Farm†			
1	11 (58%)	5 (26%)	4 (21%)
2	13 (65%)	7 (35%)	0 (0%)
Days in milk (SD)¶	8·04 (4·54)	6·33 (4·45)	9·25 (4·65)
Parity†			
1	12 (63%)	4 (21%)	3 (16%)
2	3 (37%)	4 (50%)	1 (13%)
≥3	9 (69%)	4 (31%)	0 (0%)
Disease†			
Absent	12 (57%)	8 (38%)	1 (5%)
Present	12 (63%)	4 (21%)	3 (16%)

(Continued)

Table 2 (*Continued*)

	Pattern, *n* = cow (40)			
	Never shed	Intermittent	Multi-day	
Treatment†				
None	13 (57%)	9 (39%)	1 (4%)	
Antibiotic	9 (82%)	2 (18%)	0 (0%)	
Other	2 (33%)	1 (17%)	3 (50%)	0·03

*Variables analysed with chi-square test.
†Variables analysed with Fisher's exact test.
‡Continuous variables assessed with Student's *t*-test.
§Continuous variables assessed with Wilcoxon rank test.
¶Continuous variables assessed with one-way ANOVA.

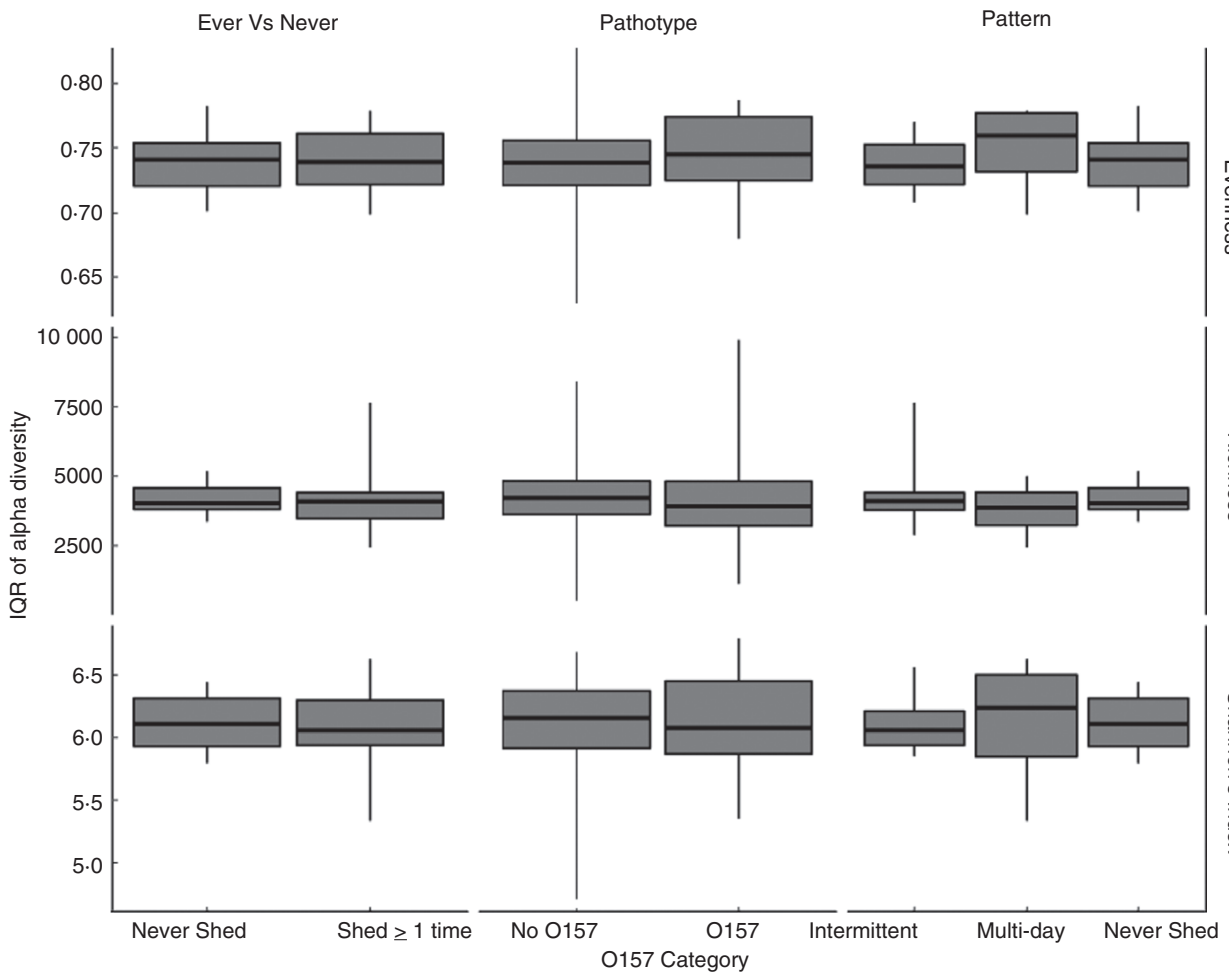

Figure 2 Interquartile range of normalized alpha-diversity measures (Shannon's, richness, evenness) defined by sample or cow O157 category. Boxes define values between the second and third quartiles, with the median value defined by a horizontal line. Whiskers define maximum and minimum diversity values respectively. For cow-level categories (ever vs never, pattern), each individual's diversity values were averaged.

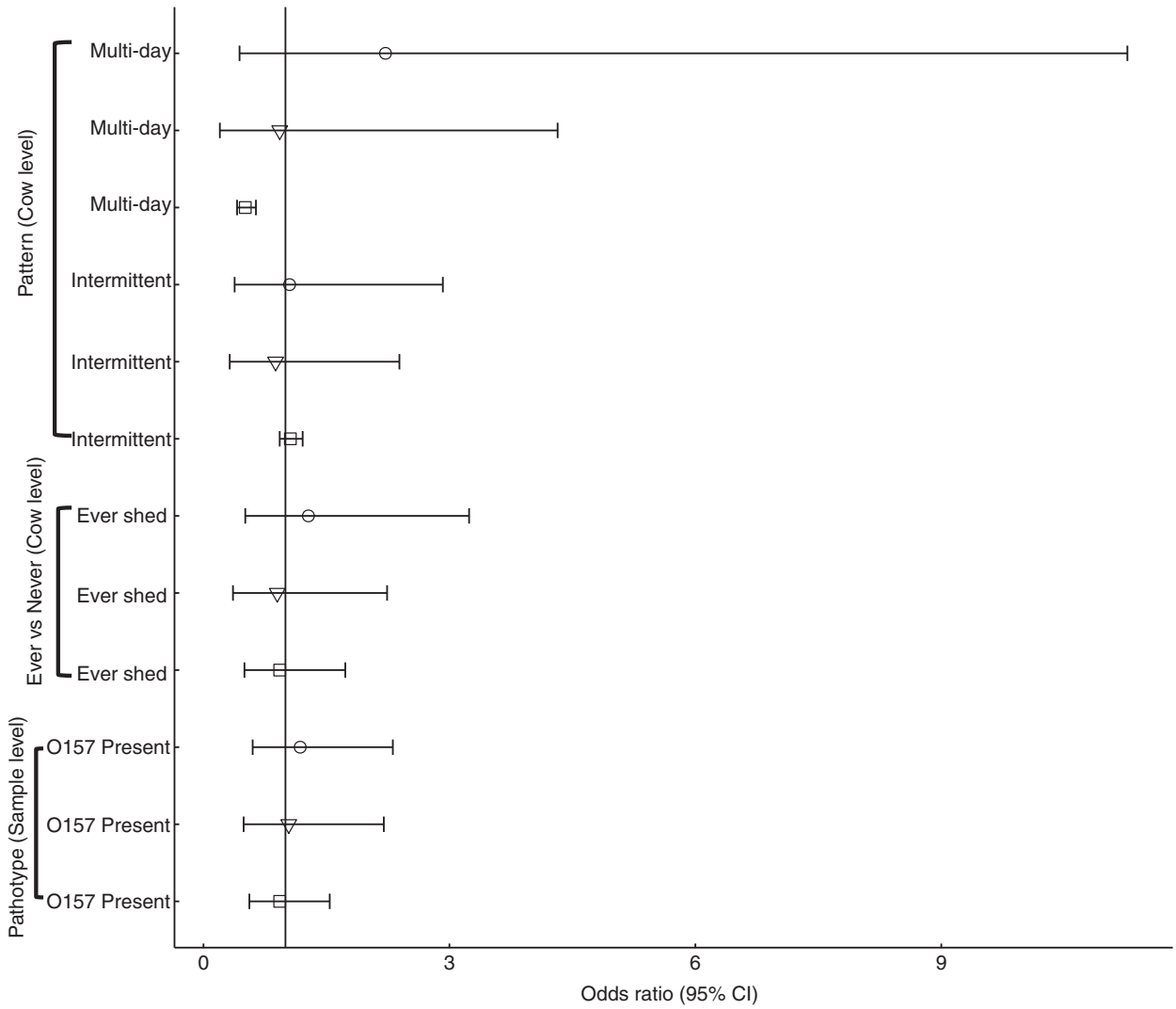

Figure 3 Associations between O157 metrics and evenness (○), Shannon's index (▽) and richness (□) measures using regression modelling. Odds ratios and confidence intervals were calculated from model coefficients using SE and interquartile range. For pattern and ever *vs* never, multinomial and logistic models were used to compare values to cows that never shed, with diversity measures averaged by cow. For pathotype, mixed logistic models with 'cow' as a random effect were used to compare O157-positive to -negative samples.

Differential abundance testing

To examine taxa that may be driving associations seen during diversity modelling, differential abundance testing was performed using sample-level O157 metrics. Both ZIG and negative binomial regression were used to evaluate differences in taxa defined at the species, genus and family levels. When aggregating taxa to these different taxonomic levels and filtering to those seen in 25% of samples, there were 195 unique species, 590 unique genus and 290 unique family-level taxa for analysis. Comparing samples with O157 to those without O157 and controlling for 'cow', one genus (*Moryella*; $P = 0{\cdot}04$) was seen to be higher in abundance and one species (*Bacillus coagulans*; $P = 0{\cdot}09$) was seen to be lower in abundance using ZIG models (Fig. 5a). Negative binomial models did not show similar conclusions; there were no family-, genus- or species-level taxa seen to have differential abundance between O157-positive and -negative samples (Fig. 5b).

We compared day prior *vs* shedding day ($n = 12$ pairs), as well as shedding day *vs* day after shedding ($n = 10$ pairs), by running Wilcoxon rank-sum tests on the differences in family, genus and species taxa abundance. After correcting for multiple testing, no specific taxa were observed to differ in days prior or after O157 shedding, when compared to the O157 shedding day.

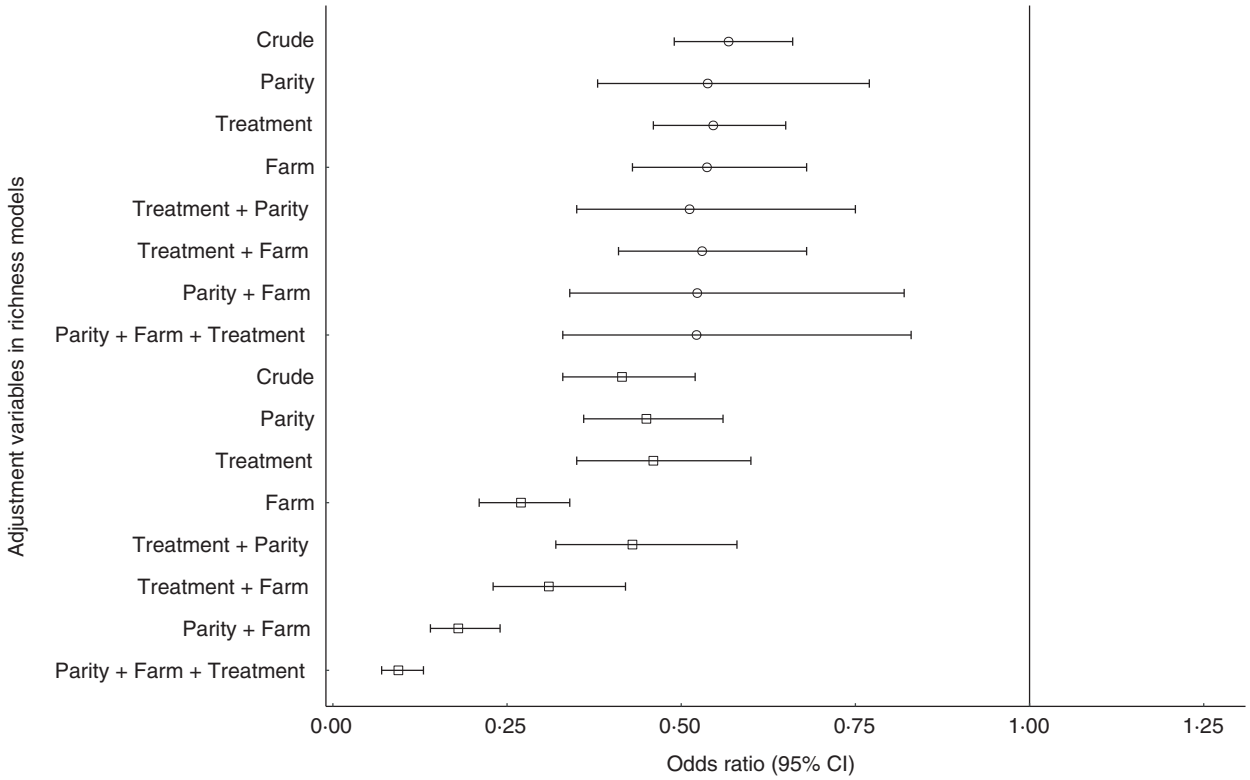

Figure 4 Crude and confounder adjusted estimates of association between O157 intermittent (○) and multi-day (□) shedding pattern and richness, omitting a single high-richness cow. Odds ratios and confidence intervals were calculated from model coefficients using SE and interquartile range. Multinomial models were used to compare values of cows that were intermittent or multi-day shedders to cows that never shed O157.

Discussion

Faecal bacterial communities influence the health and disease status of animals, and have been studied specifically in cows in terms of animal growth and development. Although the roles these communities play in pathogen presence has been investigated in mouse and other comparative models, less is known about their influence on the risk of O157 shedding in dairy cows. The current pilot study aimed to elucidate bacterial community differences associated with O157 presence and with the shedding pattern cows display by tracking communities in a cohort of shedding and nonshedding early lactation cows over the course of 5 days.

Reported shedding prevalence of O157 in dairy cows has been highly variable (1–72·7%) (Menrath *et al.* 2010; Cernicchiaro *et al.* 2012; Ferreira *et al.* 2014). We estimated a prevalence of 10% EHEC-positive samples and 2% aEPEC-positive samples during our 5-day study period. Previous work on the farms under study showed annual early lactation prevalence of 3·7 and 3% for EHEC and aEPEC respectively (Stenkamp-Strahm *et al.* 2017a). It is difficult to compare shedding rates between the current study and previous ones, as current samples were only obtained during the month of July and repeatedly from the same individuals. However, 35% of the early lactation cows shed EHEC at some point during a 5-day period. A study by Williams *et al.* (2015) sampled nulliparous dairy heifers at pasture for two periods over 8 days, and each individual was seen to shed O157 at least once during this time frame. The current project sampled primiparous and multiparous cows, and demonstrated that high rates of O157 shedding may take place during the early stages of lactation.

In the current work, sorbitol and nonsorbitol fermenting strains were initially isolated, as O157 with both characteristics can result in bovine-associated human outbreaks (King *et al.* 2014). After PCR confirmation, non-*eaeA*+ strains were not considered further, as previous identification of similar *rfb*+ strains from these herds showed them to be H12 or H45 (Stenkamp-Strahm *et al.* 2017a). Although virulence of these strains is possible, they were not the focus of the study. The authors chose to look at both aEPEC and EHEC O157 strains together during downstream analyses. This was due in part to the low number of samples with each specific pathotype, which impeded individual (bacteria level) analyses. Furthermore, there is evidence for an evolutionary

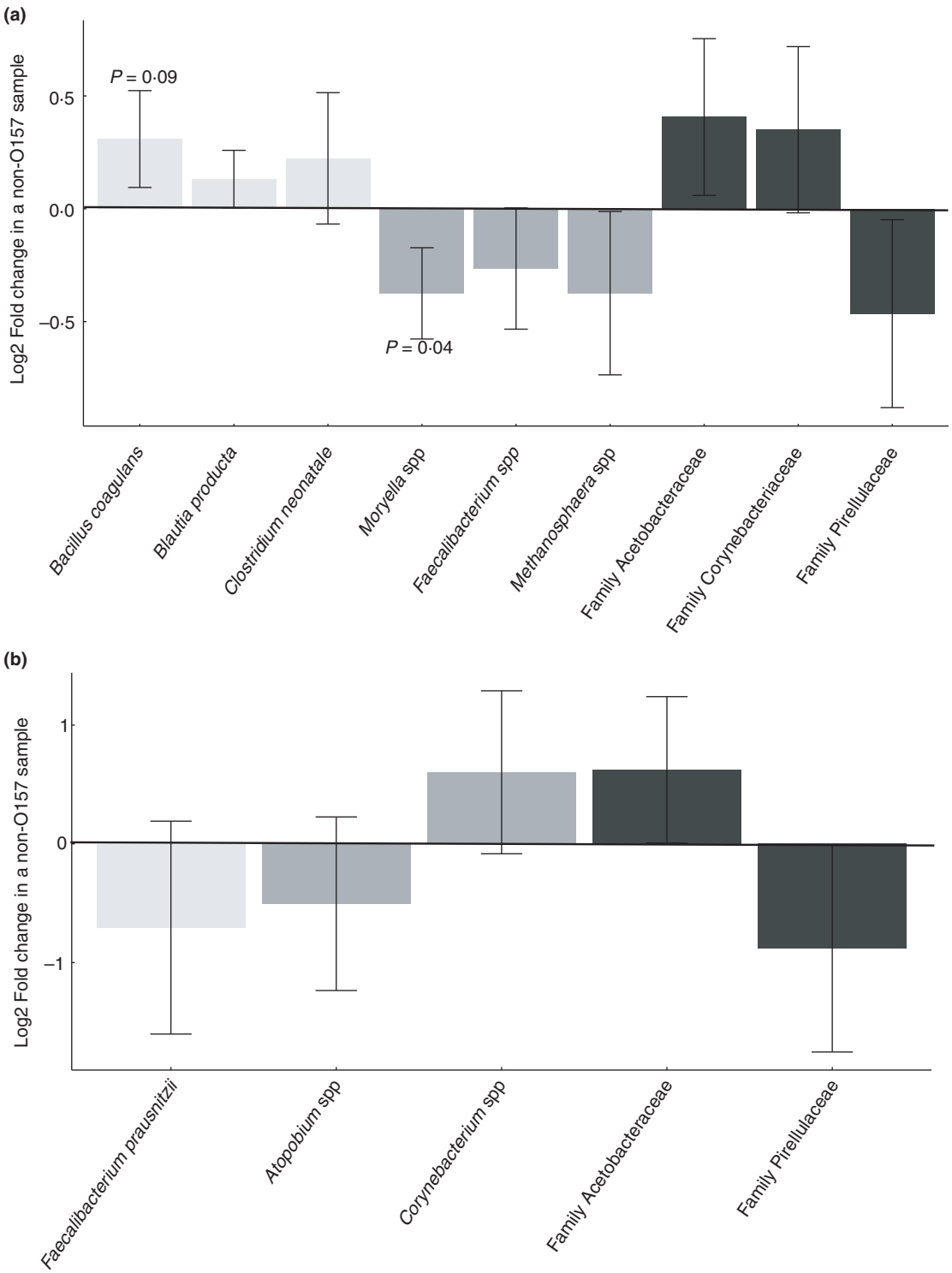

Figure 5 Log2 fold changes and 95% confidence intervals of selected species (light grey), genus (medium grey) and family (black)-level bacterial taxa in faecal samples that do not harbour pathogenic O157 compared to faecal samples that do. Using a zero-inflated Gaussian model (a), two taxa were seen to be significantly different from O157 samples after FDR correction at $P < 0{\cdot}1$; *Bacillus coagulans* ($P = 0{\cdot}04$) and *Moryella* ($P = 0{\cdot}09$). When using a negative binomial model (b), no taxa were seen to be significantly different from O157-negative samples after FDR correction.

relationship between aEPEC and EHEC O157 and high sequence homology (Cookson *et al.* 2010). Given this information, and the fact that *stx* genes reside on mobile genetic elements and harbour an ability to be 'shared' among genetically similar bacteria, the authors hypothesized that the strains likely share a niche in the bovine GI microbiota.

A few phylum-level community members slightly less abundant in O157-positive compared to -negative samples (Actinobacteria, Bacteroidetes, Spirochaetes) were likely offset by a higher proportion of Firmicutes that defined these pathogen-harbouring communities. Regardless, the early lactation cows of this study showed similarities in beta diversity when ordinating OTU counts, and in the proportions of different taxa present. This was regardless of life-history features and farm sampled. Contrary to this finding, a study by Chopyk *et al.* (2016) showed that EHEC positive cattle hide communities clustered separately from those that did not contain EHEC. That study subtracted the OTUs found in faeces by those measured on hides, and is not directly comparable to results of this analysis. Faecal microbial communities in cows have been seen previously to become more similar as animals age, with calf communities being highly variable compared to those of adults or calf dams (Zhao *et al.* 2013; Dill-McFarland *et al.* 2017). Other cow studies have shown that the feeding operation and type of ration may be more important than the history of the cow itself for causing rumen and faecal microbiota changes (Shanks *et al.* 2011). The farms in the current study fed similar rations to their early lactation cattle. We propose that the lack of separation based on beta diversity and only slight taxa proportion differences seen by O157 status of these faecal samples is likely attributable to herd nutritional and managerial uniformity. These outcomes presented an ideal platform for measuring O157-related changes in abundance of specific taxa within the samples, and modelling diversity measures with O157 shedding metrics.

When modelling bacterial diversity measures, cows that were classified as intermittent or multi-day O157 shedders were seen to have lower average richness than cows that never shed during the study period. These findings remained when controlling for life-history features of cows shown to influence shedding and faecal-level alpha diversity. Research has shown that the presence of different commensal microbes can reduce the proliferation of invading pathogens through competitive exclusion for space, nutrient depletion and creation of antimicrobial compounds. Results from studies looking at cattle environments suggest that O157 tends to thrive where there is less microbial diversity. For instance, O157 was seen to survive better in manure-amended soil that was autoclaved *vs* manure-amended soil that was not autoclaved (Jiang *et al.* 2002). A study that used denaturing gradient gel electrophoresis to classify beef microbial communities containing STEC showed that Shannon's diversity was lower in STEC-positive compared to -negative samples (Zhao *et al.* 2013). Likewise, bacterial diversity was lower in cattle hide samples that contained STEC compared to those that did not (Chopyk *et al.* 2016). Contrary to these results, a study by Xu *et al.* (2014) revealed that 'super-shedding' ($\geq 10^4$ CFU per gram of faeces) feedlot steers had significantly higher richness compared to their O157-negative pen-mates. Changes in outcomes between that and the current study likely have to do with no cattle in the current study 'super-shedding', sequencing depth for analysis, differences in study animals and sequencing approaches.

Shannon's diversity index takes into account both the richness of samples and the relative abundance of taxa. We speculate that Shannon's index was not associated with O157 categories, while richness was, because the proportions of taxa within samples were very similar. Furthermore, we were unable to establish temporality between O157 presence and richness due to the nature of microbial community data collection, which took place only during a 5-day window. This is perhaps a reason why analysis at the sample level (pathotype) did not reveal associations between O157 presence and richness, while those at the cow level (pattern of shedding) did demonstrate these associations. Perhaps it is not the richness of the community during the shedding event, but the community immediately preceding, or preceding for a span of days, that influences eventual O157 colonization. Based on our conclusions, we can hypothesize that O157 is either able to gradually occupy an empty niche in less diverse bacterial communities, or acts to outcompete other bacteria and reduce or change the community diversity by its presence alone, prior to colonization. It remains possible that diversity outcomes are dependent on faecal quantities or exposure dose of the O157 pathogen, especially given the single outlier cow that shed while having a GI community with very high richness. More research is required to know if this is the case.

There is an interest in identifying specific faecal community members that are associated with the presence or absence of O157. Although previous microbiome studies have identified OTUs in communities that harbour STEC (Xu *et al.* 2014; Chopyk *et al.* 2016), this study is the first of its kind to measure changes in community members aggregated to specific taxa levels (family, genus, species) via robust modelling of normalized sequence counts. Perhaps the most important change in this approach is that no sequencing data have been omitted, as is commonly done in studies that rarefy counts to the lowest sample.

Using ZIG modelling of counts normalized with CSS, we identified one genus, *Moryella*, more abundant in O157-positive samples. Currently, the genus *Moryella* contains only one defined species: *Moryella indoligenes* (Carlier *et al.* 2007). This is an indole-producing, strictly anaerobic bacterium isolated from clinical abscesses in humans, presumably originating from within the human GI tract (Carlier *et al.* 2007). Other information on this species and genus remain scant, although the bacteria seem to be closely related to species of the genus *Clostridium*, and are weakly saccharolytic. The production of indole by *Moryella* sp. may be a reason why it exists in concert with O157. Indole is produced by Gram-positive and -negative bacteria, including *E. coli* and many pathogens, and may act as an interspecies signalling molecule (Lee and Lee 2010). Other diverse activities of the molecule include mediating virulence genes of EPEC, modifying EHEC genetic activation to enhance attaching and effacing lesion formation and stabilizing *E. coli* plasmids to maintain high copy numbers (Chant and Summers 2007; Hirakawa *et al.* 2009). *In vitro* the molecule can also decrease surface colonization and motility, and down-regulate genes that cause EHEC biofilm formation (Bansal *et al.* 2007). The exact mechanism by which *Moryella* sp. associated indole may influence O157 presence in our study requires further research.

The ZIG model also showed that *B. coagulans* was more abundant in non-O157 containing communities. *Bacillus coagulans* is a Gram-positive mobile facultative anaerobe that produces lactic acid and the bacteriocin-like inhibitory substance coagulin (Hyronimus *et al.* 1998). This species is currently marketed as a beneficial probiotic for humans, dogs, pigs, horses and cows. The bacterium is used as an add-in to supplements that contain electrolytes or vitamins and minerals, and is advertised to support calf health and growth, promote optimal digestion and ease the transition to lactation in dairy cows. The herds of this study were not supplementing their stock with any product that may have contained *B. coagulans*, or other probiotics. Our differential abundance result is supported by a recent study that showed *B. coagulans* supplementation reducing the total faecal coliform count in rats (Haldar and Gandhi 2016). It is possible that the lactic acid producing properties of this species form the basis for its negative correlation with O157. Lactic acid, produced by strains like *B. coagulans* and also *Lactobacillus* species, has been shown to permeabilize the Gram-negative membrane of pathogenic bacteria like *Salmonella*, *E. coli* and *Pseudomonas aeruginosa* (Alakomi *et al.* 2000). Strains in the genus *Bifidobacterium* and *Lactobacillus* were seen to be antagonistic against pyelonephritic *E. coli in vitro* (Hutt *et al.* 2006). In another study, the background flora in ground beef products, which contained a large percentage of lactic acid-producing bacteria, inhibited both the aerobic and anaerobic growth of O157 (Vold *et al.* 2000). *Lactobacillus* sp. supplementation has also been studied specifically in light of O157 shedding in cows. Strains in this genus were seen to reduce, but not eliminate, faecal shedding of O157 by beef feedlot steers in two separate studies (Brashears *et al.* 2003; Peterson *et al.* 2007). The exact mechanism by which bacteria like *Lactobacillus* sp. and *B. coagulans* are antagonistic towards O157 in the GI communities of cattle remains to be determined, but merits further investigation.

Because these changes in taxa were seen with the shedding event itself, we could not determine if they were prior to or a result of O157 presence. We sought to assess the differences in microbial community taxa of shedding animals on days preceding and following their shedding event, in an effort to establish temporality. Due to the pilot nature of the project and small number of matched sample pairs, the statistical rank test did not reveal significant changes in any taxa. Day-to-day changes in GI communities of animals with unchanged diet or environment are likely minimal. True assessment of differences in O157 between days would require a high-powered match test, so these results cannot conclusively determine if taxa differences exist before and after shedding events. When using a differential abundance modelling strategy with a negative binomial distribution, the taxa changes seen via ZIG modelling were not upheld. Although many of the same taxa were close to significance in both models, we cannot overlook the implication of this outcome. There is a lack of consensus in the research community regarding the best approaches to measure differential microbial abundance using 16S rRNA data (McMurdie and Holmes 2014; Thorsen *et al.* 2016; Weiss *et al.* 2017). Both modelling approaches utilized in the current study presented different strengths. ZIG models have been shown to have higher false-positive rates, but concurrently higher detection sensitivity, compared to others. Given this pilot study comparison between relatively few ($n = 40$) individual cows, and similarities in overall community due to diet and environment, the authors chose to use the high-sensitivity ZIG model in concert with a more conservative negative binomial. Changes in outcome between models may also be due to differences in the normalization strategies; ZIG in metagenomeSeq utilizes a cumulative sum-scaling approach, while DeSeq2 uses an approach common to RNASeq analyses involving internal normalization with geometric means. To validate the O157-associated changes of taxa seen in this study future work should employ *in vitro* experiments, as well as additional MCS measure of shedding cows (beef and dairy) that are managed in different ways.

Mitigating shedding in cattle is one approach to reducing outbreaks of human O157 disease. Historically, understanding the dynamics of cow O157 colonization has proven difficult, and blocked the progression towards developing cow management or treatment strategies that reduce shedding. Understanding the cow microbiota and changes surrounding shedding events can theoretically lead to a better understanding of shedding dynamics, while also revealing potential treatment strategies. Results of the current work showed that regardless of life-history features known to influence shedding, lower average bacterial richness was associated with an intermittent or multi-day shedding pattern in dairy cows. The species *B. coagulans* was negatively associated with the presence of O157 in faecal samples, while the genus *Moryella* was positively associated with O157. These species may be used as targets for therapeutic agents (*Moryella*), or as therapeutic agents (*B. coagulans*). *In vitro* work is required to confirm and validate these taxa findings first, however, and investigate the mechanisms by which these bacteria influence O157. Taken together, results of this study suggest that the cow microbiome plays a role in O157 pathogen presence. Modulating microbial communities may be a useful approach to reduce O157 shedding in cows.

Acknowledgements

The authors thank the participating dairies for their involvement in this study and Dr. Joshua Schaeffer and Amanda VanDyke for their contribution and help during study design and farm sampling. The authors also acknowledge Dr. Lyndsey Linke and Roberta Magnuson for their help with the laboratory characterization of faecal samples. This work was funded by USDA NIFA Health and Disease Formula Funds (CRCFY2015; 1004896), MoBio and the Earth Microbiome Consortium's 2nd Place Award for 'Innovative Microbiome Projects', a cooperative agreement between Colorado State University and USDA:APHIS:VS: CEAH (15-9200-0443CA) and the High Plains Intermountain Center for Agricultural Health and Safety by a grant from CDC NIOSH (5U54OH008085).

Conflict of Interest

No conflict of interest declared.

References

Alakomi, H.L., Skytta, E., Saarela, M., Mattila-Sandholm, T., Latva-Kala, K. and Helander, I.M. (2000) Lactic acid permeabilizes gram-negative bacteria by disrupting the outer membrane. *Appl Environ Microbiol* **66**, 2001–2005.

Ayaz, N.D., Gencay, Y.E. and Erol, I. (2014) Prevalence and molecular characterization of sorbitol fermenting and non-fermenting *Escherichia coli* O157:H7(+)/H7(-) isolated from cattle at slaughterhouse and slaughterhouse wastewater. *Int J Food Microbiol* **174**, 31–38.

Bansal, T., Englert, D., Lee, J., Hegde, M., Wood, T.K. and Jayaraman, A. (2007) Differential effects of epinephrine, norepinephrine, and indole on *Escherichia coli* O157:H7 chemotaxis, colonization, and gene expression. *Infect Immun* **75**, 4597–4607.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodol)* **57**, 289–300.

Brashears, M.M., Galyean, M.L., Loneragan, G.H., Mann, J.E. and Killinger-Mann, K. (2003) Prevalence of *Escherichia coli* O157:H7 and performance by beef feedlot cattle given Lactobacillus direct-fed microbials. *J Food Prot* **66**, 748–754.

Buffie, C.G. and Pamer, E.G. (2013) Microbiota-mediated colonization resistance against intestinal pathogens. *Nat Rev Immunol* **13**, 790–801.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. and Knight, R. (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G. *et al.* (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, **7**, 335–336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J. *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621–1624.

Carlier, J.P., K'Ouas, G. and Han, X.Y. (2007) *Moryella indoligenes* gen. nov., sp. nov., an anaerobic bacterium isolated from clinical specimens. *Int J Syst Evol Microbiol* **57**, 725–729.

Carrothers, J.M., York, M.A., Brooker, S.L., Lackey, K.A., Williams, J.E., Shafii, B., Price, W.J., Settles, M.L. *et al.* (2015) Fecal microbial community structure is stable over time and related to variation in macronutrient and micronutrient intakes in lactating women. *J Nutr* **145**, 2379–2388.

CEAH, USDA. (2003) *Escherichia coli* O157 on US Dairy Operations. Fort Collins: CEAH.

Cernicchiaro, N., Pearl, D.L., McEwen, S.A., Harpster, L., Homan, H.J., Linz, G.M. and Lejeune, J.T. (2012) Association of wild bird density and farm management factors with the prevalence of *E. coli* O157 in dairy herds in Ohio (2007-2009). *Zoonoses Public Health* **59**, 320–329.

Chant, E.L. and Summers, D.K. (2007) Indole signalling contributes to the stable maintenance of *Escherichia coli* multicopy plasmids. *Mol Microbiol* **63**, 35–43.

Chevalier, C., Stojanovic, O., Colin, D.J., Suarez-Zamorano, N., Tarallo, V., Veyrat-Durebex, C., Rigo, D., Fabbiano, S. *et al.* (2015) Gut microbiota orchestrates energy homeostasis during cold. *Cell* **163**, 1360–1374.

Cho, I. and Blaser, M.J. (2012) The human microbiome: at the interface of health and disease. *Nat Rev Genet* **13**, 260–270.

Chopyk, J., Moore, R.M., DiSpirito, Z., Stromberg, Z.R., Lewis, G.L., Renter, D.G., Cernicchiaro, N., Moxley, R.A. *et al.* (2016) Presence of pathogenic *Escherichia coli* is correlated with bacterial community diversity and composition on pre-harvest cattle hides. *Microbiome* **4**, 9.

Cobbaut, K., Berkvens, D., Houf, K., De Deken, R. and De Zutter, L. (2009) *Escherichia coli* O157 prevalence in different cattle farm types and identification of potential risk factors. *J Food Prot* **72**, 1848–1853.

Cobbold, R.N., Hancock, D.D., Rice, D.H., Berg, J., Stilborn, R., Hovde, C.J. and Besser, T.E. (2007) Rectoanal junction colonization of feedlot cattle by *Escherichia coli* O157:H7 and its association with supershedders and excretion dynamics. *Appl Environ Microbiol* **73**, 1563–1568.

Cookson, A.L., Cao, M., Bennett, J., Nicol, C., Thomson-Carter, F. and Attwood, G.T. (2010) Relationship between virulence gene profiles of atypical enteropathogenic *Escherichia coli* and Shiga toxin-producing *E. coli* isolates from cattle and sheep in New Zealand. *Appl Environ Microbiol* **76**, 3744–3747.

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M. and Finlay, B.B. (2013) Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin Microbiol Rev* **26**, 822–880.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D. *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069–5072.

Dill-McFarland, K.A., Breaker, J.D. and Suen, G. (2017) Microbial succession in the gastrointestinal tract of dairy cows from 2 weeks to first lactation. *Sci Rep* **7**, 40864.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461.

Ferreira, M.R., Freitas Filho, E.G., Pinto, J.F., Dias, M. and Moreira, C.N. (2014) Isolation, prevalence, and risk factors for infection by shiga toxin-producing *Escherichia coli* (STEC) in dairy cattle. *Trop Anim Health Prod* **46**, 635–639.

Gilbert, J.A., Jansson, J.K. and Knight, R. (2014) The Earth Microbiome project: successes and aspirations. *BMC Biol* **12**, 69.

Haldar, L. and Gandhi, D.N. (2016) Effect of oral administration of *Bacillus coagulans* B37 and *Bacillus pumilus* B9 strains on fecal coliforms, *Lactobacillus* and *Bacillus* spp. in rat animal model. *Vet World* **9**, 766–772.

Hirakawa, H., Kodama, T., Takumi-Kobayashi, A., Honda, T. and Yamaguchi, A. (2009) 'Secreted indole serves as a signal for expression of type III secretion system translocators in enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* **155**, 541–550.

Hutt, P., Shchepetova, J., Loivukene, K., Kullisaar, T. and Mikelsaar, M. (2006) Antagonistic activity of probiotic lactobacilli and bifidobacteria against entero- and uropathogens. *J Appl Microbiol* **100**, 1324–1332.

Hyronimus, B., Le Marrec, C. and Urdaci, M.C. (1998) Coagulin, a bacteriocin-like inhibitory substance produced by *Bacillus coagulans* I4. *J Appl Microbiol* **85**, 42–50.

Jami, E. and Mizrahi, I. (2012) Similarity of the ruminal bacteria across individual lactating cows. *Anaerobe* **18**, 338–343.

Jewell, K.A., McCormick, C.A., Odt, C.L., Weimer, P.J. and Suen, G. (2015) Ruminal bacterial community composition in dairy cows is dynamic over the course of two lactations and correlates with feed efficiency. *Appl Environ Microbiol* **81**, 4697–4710.

Jiang, X., Morgan, J. and Doyle, M.P. (2002) Fate of *Escherichia coli* O157:H7 in manure-amended soil. *Appl Environ Microbiol* **68**, 2605–2609.

King, L.A., Loukiadis, E., Mariani-Kurkdjian, P., Haeghebaert, S., Weill, F.X., Baliere, C., Ganet, S., Gouali, M. *et al.* (2014) Foodborne transmission of sorbitol-fermenting *Escherichia coli* O157:[H7] via ground beef: an outbreak in northern France, 2011. *Clin Microbiol Infect* **20**, O1136–O1144.

Lee, J.H. and Lee, J. (2010) Indole as an intercellular signal in microbial communities. *FEMS Microbiol Rev* **34**, 426–444.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550.

Majowicz, S.E., Scallan, E., Jones-Bitton, A., Sargeant, J.M., Stapleton, J., Angulo, F.J., Yeung, D.H. and Kirk, M.D. (2014) Global incidence of human shiga toxin–producing *Escherichia coli* infections and deaths: a systematic review and knowledge synthesis. *Foodborne Pathog Dis* **11**, 447–455.

Mao, S., Zhang, M., Liu, J. and Zhu, W. (2015) Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. *Sci Rep* **5**, 16116.

March, S.B. and Ratnam, S. (1986) Sorbitol-MacConkey medium for detection of *Escherichia coli* O157:H7 associated with hemorrhagic colitis. *J Clin Microbiol* **23**, 869–872.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R. *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**, 610–618.

McMurdie, P.J. and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217.

McMurdie, P.J. and Holmes, S. (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* **10**, e1003531.

Mechie, S.C., Chapman, P.A. and Siddons, C.A. (1997) A fifteen month study of *Escherichia coli* O157:H7 in a dairy herd. *Epidemiol Infect* **118**, 17–25.

Menrath, A., Wieler, L.H., Heidemanns, K., Semmler, T., Fruth, A. and Kemper, N. (2010) Shiga toxin producing *Escherichia coli*: identification of non-O157:H7-Super-Shedding cows and related risk factors. *Gut Pathog* **2**, 7.

Nguyen, Y. and Sperandio, V. (2012) Enterohemorrhagic *E. coli* (EHEC) pathogenesis. *Front Cell Infect Microbiol* **2**, 90.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minhcin, P.R., O'Hara, R.B. *et al.* (2017) vegan: Community Ecology Package. R package version 2.4-3. https://CRAN.R-project.org/package=vegan

Paton, A.W. and Paton, J.C. (1998) Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for stx1, stx2, eaeA, enterohemorrhagic *E. coli* hlyA, rfbO111, and rfbO157. *J Clin Microbiol* **36**, 598–602.

Paulson, J.N., Stine, O.C., Bravo, H.C. and Pop, M. (2013) Differential abundance analysis for microbial marker-gene surveys. *Nat Methods* **10**, 1200–1202.

Peterson, R.E., Klopfenstein, T.J., Erickson, G.E., Folmer, J., Hinkley, S., Moxley, R.A. and Smith, D.R. (2007) Effect of *Lactobacillus acidophilus* strain NP51 on *Escherichia coli* O157:H7 fecal shedding and finishing performance in beef feedlot cattle. *J Food Prot* **70**, 287–291.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Rideout, J.R., He, Y., Navas-Molina, J.A., Walters, W.A., Ursell, L.K., Gibbons, S.M., Chase, J., McDonald, D. *et al.* (2014) Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* **2**, e545.

Robinson, S.E., Wright, E.J., Hart, C.A., Bennett, M. and French, N.P. (2004) Intermittent and persistent shedding of *Escherichia coli* O157 in cohorts of naturally infected calves. *J Appl Microbiol* **97**, 1045–1053.

Schmidt, H., Scheef, J., Huppertz, H.I., Frosch, M. and Karch, H. (1999) *Escherichia coli* O157:H7 and O157:H(-) strains that do not produce Shiga toxin: phenotypic and genetic characterization of isolates associated with diarrhea and hemolytic-uremic syndrome. *J Clin Microbiol* **37**, 3491–3496.

Sekirov, I., Tam, N.M., Jogova, M., Robertson, M.L., Li, Y., Lupp, C. and Finlay, B.B. (2008) Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infect Immun* **76**, 4726–4736.

Shanks, O.C., Kelty, C.A., Archibeque, S., Jenkins, M., Newton, R.J., McLellan, S.L., Huse, S.M. and Sogin, M.L. (2011) Community structures of fecal bacteria in cattle from different animal feeding operations. *Appl Environ Microbiol* **77**, 2992–3001.

Shere, J.A., Bartlett, K.J. and Kaspar, C.W. (1998) Longitudinal study of *Escherichia coli* O157:H7 dissemination on four dairy farms in Wisconsin. *Appl Environ Microbiol* **64**, 1390–1399.

Stenkamp-Strahm, C., Mc, C.C., Rao, S., Magnuson, R., Hyatt, D.R. and Linke, L. (2017a) Climate, lactation, and treatment factors influence faecal shedding of *Escherichia coli* O157 pathotypes in dairy cows. *Epidemiol Infect* **145**, 115–125.

Stenkamp-Strahm, C., McConnel, C., Hyatt, D.R., Magnuson, R., Tenneson, P. and Linke, L. (2017b) Prevalence of *Escherichia coli* O157 shedding in preweaned calves on Colorado dairies. *J Food Prot* **80**, 990–993.

Thorsen, J., Brejnrod, A., Mortensen, M., Rasmussen, M.A., Stokholm, J., Al-Soud, W.A., Sorensen, S., Bisgaard, H. *et al.* (2016) Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. *Microbiome* **4**, 62.

Venegas-Vargas, C., Henderson, S., Khare, A., Mosci, R.E., Lehnert, J.D., Singh, P., Ouellette, L.M., Norby, B. *et al.* (2016) Factors associated with shiga toxin-producing *Escherichia coli* shedding by dairy and beef cattle. *Appl Environ Microbiol* **82**, 5049–5056.

Vold, L., Holck, A., Wasteson, Y. and Nissen, H. (2000) High levels of background flora inhibits growth of *Escherichia coli* O157:H7 in ground beef. *Int J Food Microbiol* **56**, 219–225.

Wang, G., Clark, C.G. and Rodgers, F.G. (2002) Detection in *Escherichia coli* of the genes encoding the major virulence factors, the genes defining the O157:H7 serotype, and components of the type 2 Shiga toxin family by multiplex PCR. *J Clin Microbiol* **40**, 3613–3619.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**, 5261–5267.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R. *et al.* (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27.

Wells, S., Fedorka-Cray, P., Besser, T., McDonough, P. and Smith, B. (1998) *E. coli* O157 and Salmonella- Status on US Dairy Operations. Fort Collins, CO: CEAH.

Williams, K.J., Ward, M.P., Dhungyel, O.P. and Hall, E.J. (2015) Risk factors for *Escherichia coli* O157 shedding and super-shedding by dairy heifers at pasture. *Epidemiol Infect* **143**, 1004–1015.

Xu, Y., Dugat-Bony, E., Zaheer, R., Selinger, L., Barbieri, R., Munns, K., McAllister, T.A. and Selinger, L.B. (2014) *Escherichia coli* O157:H7 super-shedder and non-shedder feedlot steers harbour distinct fecal bacterial communities. *PLoS ONE* **9**, e98115.

Zhao, L., Tyler, P.J., Starnes, J., Bratcher, C.L., Rankins, D., McCaskey, T.A. and Wang, L. (2013) Correlation analysis

of Shiga toxin-producing *Escherichia coli* shedding and faecal bacterial composition in beef cattle. *J Appl Microbiol* **115**, 591–603.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Directed acyclic graph of factors hypothesized to confound or mediate the association between alpha diversity and O157 outcome.

Table S1. Associations between cow covariates and alpha-diversity measures using regression modelling.

Table S2. Crude and confounder adjusted estimates of associations between O157 category and alpha diversity.