Epidemics 37 (2021) 100521

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Transmission dynamics of an antimicrobial resistant *Campylobacter jejuni* lineage in New Zealand's commercial poultry network

Sabrina S. Greening [a,*], Ji Zhang [b,c], Anne C. Midwinter [b], David A. Wilkinson [b,c], Ahmed Fayaz [b], Deborah A. Williamson [d], Marti J. Anderson [e], M. Carolyn Gates [a], Nigel P. French [b,c]

[a] *Epicentre, School of Veterinary Science, Massey University, Palmerston North, New Zealand*
[b] *mEpiLab, Infectious Disease Research Centre, School of Veterinary Science, Massey University, Palmerston North, New Zealand*
[c] *New Zealand Food Safety Science and Research Centre, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand*
[d] *Microbiological Diagnostic Unit and Public Health Laboratory, University of Melbourne, Parkville, Victoria, Australia*
[e] *New Zealand Institute for Advanced Study, Massey University, Auckland, New Zealand*

ARTICLE INFO

Keywords:
Campylobacter jejuni
Contact networks
Phylodynamics
PERMANOVA
Correlation matrices

ABSTRACT

Understanding the relative contribution of different between-farm transmission pathways is essential in guiding recommendations for mitigating disease spread. This study investigated the association between contact pathways linking poultry farms in New Zealand and the genetic relatedness of antimicrobial resistant *Campylobacter jejuni* Sequence Type 6964 (ST-6964), with the aim of identifying the most likely contact pathways that contributed to its rapid spread across the industry. Whole-genome sequencing was performed on 167*C. jejuni* ST-6964 isolates sampled from across 30 New Zealand commercial poultry enterprises. The genetic relatedness between isolates was determined using whole genome multilocus sequence typing (wgMLST). Permutational multivariate analysis of variance and distance-based linear models were used to explore the strength of the relationship between pairwise genetic associations among the *C. jejuni* isolates and each of several pairwise distance matrices, indicating either the geographical distance between farms or the network distance of transportation vehicles. Overall, a significant association was found between the pairwise genetic relatedness of the *C. jejuni* isolates and the parent company, the road distance and the network distance of transporting feed vehicles. This result suggests that the transportation of feed within the commercial poultry industry as well as other local contacts between flocks, such as the movements of personnel, may have played a significant role in the spread of *C. jejuni*. However, further information on the historical contact patterns between farms is needed to fully characterise the risk of these pathways and to understand how they could be targeted to reduce the spread of *C. jejuni*.

1. Introduction

Controlling the spread of any infectious disease requires decision-makers to have a good understanding of the mechanisms and pathways through which the infectious agent is spreading within the population. However, in practice, this can be difficult to achieve since many pathogens can utilise multiple transmission pathways to jump between hosts, with transmission modes often varying among pathogen strains and host populations (Antonovics et al., 2017). This creates a major challenge in many infectious disease outbreaks, as it becomes difficult to determine the relative contribution of each pathway to the spread of a disease, especially when contact tracing data is limited. This in turn makes it difficult for decision-makers to recommend targeted control strategies that are timelier and more cost-effective than broader approaches (Webster et al., 2017). A good example of this is *Campylobacter jejuni*; one of the leading causes of foodborne gastroenteritis worldwide (Kaakoush et al., 2015). This pathogen has complicated dynamics and multiple transmission pathways which makes it challenging to model and control (Koutsoumanis et al., 2016), despite the implementation of many targeted intervention strategies across different countries (Lin, 2009; Newell et al., 2011).

In New Zealand, source attribution models have identified a range of sources responsible for human campylobacteriosis cases with by far the largest proportion of human cases having been linked to the

* Corresponding author.
E-mail address: S.Greening@massey.ac.nz (S.S. Greening).

https://doi.org/10.1016/j.epidem.2021.100521
Received 13 August 2020; Received in revised form 5 August 2021; Accepted 7 November 2021
Available online 8 November 2021

consumption of poultry meat (Müllner et al., 2009). This resulted in the implementation of regulatory and voluntary control strategies along the poultry supply chain and, in the three years following their implementation from 2005 to 2008, there was a 50% reduction in human campylobacteriosis notifications (Sears et al., 2011; Nohra et al., 2020). However, despite continued control efforts, campylobacteriosis continues to be the most frequently reported foodborne illness in New Zealand, with rates as much as ten times greater than those reported in the United States and double those reported in other industrialised countries such as the United Kingdom (Olson et al., 2008). In addition to the high incidence of campylobacteriosis cases in New Zealand, routine sampling at a sentinel surveillance site in 2014 also detected a new strain of *C. jejuni*, identified as Sequence Type 6964 (ST-6964), that was resistant to both fluoroquinolones and tetracycline (French et al., 2019). This antimicrobial resistant strain subsequently became a dominant strain in both poultry and humans, affecting multiple poultry suppliers in the North Island (French et al., 2019). These findings not only represented a major epidemiological shift, with previous *Campylobacter* STs only found to be associated with individual poultry suppliers (Müllner et al., 2010) but also represented a significant change in the patterns of resistance (Williamson et al., 2015), similar to that seen in many other countries (Kaakoush et al., 2015). Further analysis using the molecular sequence data of 227*C. jejuni* ST-6964 isolates also revealed evolutionary changes in the accessory genome associated with a plasmid, phage insertions, and natural transformation (French et al. 2019). However, despite these insights into the population structure and evolutionary dynamics of *C. jejuni* ST-6964, the transmission dynamics of this strain remain unclear.

In similar studies that have used molecular sequence data to characterise the genetic diversity between sampled pathogen isolates, epidemiological linkages between closely related isolates have been used to help make inferences about unknown transmission dynamics (Gilbertson et al., 2018; Grenfell, 2004; Ypma et al., 2012). An increasing number of these studies integrating pathogen phylogenies in epidemiological investigations use host contact network data (Jombart et al., 2014; Leventhal et al., 2012; Stadler and Bonhoeffer, 2013). In infectious disease epidemiology, contact networks are used to capture all known interactions, or contacts, that are believed to contribute to disease spread within a population (Chaters et al., 2019; Sah et al., 2018; Silk et al., 2018). These data can be used in disease outbreaks alongside traditional contact tracing methods to help reconstruct transmission trees based on the assumption that the true transmission network will always be a subset of the larger contact network (Craft, 2015). However, these methods largely rely on the ability of researchers to correctly define all the contacts that are relevant for disease transmission, which in many cases is extremely challenging (Bansal et al., 2010; Eames et al., 2015). Therefore, many studies have instead relied on theoretical models based on simulated data, such as network-based disease simulation models that integrate pathogen sequence data into network analyses, with few real-world examples currently in the literature to validate novel methodologies (Enns and Brandeau, 2011; Eubank et al., 2004; Zhang et al., 2012).

In New Zealand, the recent emergence and rapid spread of *C. jejuni* ST-6964 presented a timely opportunity to collect pathogen whole-genome sequence data and farm-level contact network data and apply statistical approaches that explore the relative contribution of different transmission pathways to the spread of *C. jejuni* ST-6964. Given this, this study aimed to investigate the degree of congruence between multiple potential contact pathways and the genetic relatedness of these isolates in order to determine the most likely routes of transmission contributing towards the rapid spread of this strain.

2. Materials and methods

2.1. Sample collection, isolate culture, and whole-genome sequencing

As an extension of the poultry survey reported by Muellner et al. (2016), 922 swabs were taken from birds across 75 commercial poultry farms including farms from each of the four major poultry suppliers that account for over 90% of chicken meat production in New Zealand (Muellner et al., 2016). As samples were taken from poultry carcases post-slaughter at commercial poultry abattoirs it was advised by the Massey University Human Ethics Committee: Northern that animal ethical approval was not required. Each swab was taken from the pooled caecal contents of up to five chickens originated from the same farm, with swabs being collected from slaughter processing plants across New Zealand's North and South Island between May 2015 and July 2016. Swabs were delivered for bacterial culture to the Molecular Epidemiology and Public Health laboratory (^{m}EpiLab) at Massey University where cultures were grown in a microaerobic incubator (Don Whitley Scientific, Yorkshire, UK) at 42 °C on modified charcoal cefoperazone deoxycholate agar (mCCDA) (LabM, Lancashire, UK) containing ciprofloxacin (4 mg/litre) (Sigma, Missouri, USA) and tetracycline (16 mg/litre) (Sigma, Missouri, USA) for selective isolation of resistant *Campylobacter* colonies. Overall, 72.5% (668/922) of the swabs gave growth that resembled *Campylobacter* including swabs from across 54.7% (41/75) of the farms. From each positive plate, one to two single colonies were sub-cultured and genomic DNA was isolated on a JANUS automated workstation (PerkinElmer, Massachusetts, USA) using Chemagic magnetic bead technology (PerkinElmer, Massachusetts, USA), according to the manufacturer's instructions. *C. jejuni* ST-6964 isolates were confirmed by 7-gene multilocus sequence typing (Dingle et al., 2008) as described by Nohra et al. (2016).

For this study, a subset of 167 of the confirmed *C. jejuni* isolates were randomly selected for whole-genome sequencing (WGS). This subset included isolates from across 30 individual poultry farms belonging to three out of the four major poultry suppliers (hereafter anonymously referred to as "A", "B" and "C"); including 26 broiler flocks (*i.e.,* poultry growers) and four breeding flocks. For WGS, the DNA quality was assessed using Qubit™ dsDNA high sensitivity assay kits (Thermo Fisher Scientific Inc., Oregon, USA) before DNA libraries were prepared using a NexteraXT DNA preparation kit (Illumina, California, USA). 2 × 100 bp sequencing was performed on all 167 isolates using the NextSeq 500 platform (Illumina, California, USA), as previously described (Baines et al., 2016), with sequence data available from GenBank BioProject ID PRJNA520992 and PubMLST (https://pubmlst.org/campylobacter) no. 70207-12, 70229, 70230, 70232, 70233, 70252, 70253, and 78631-845.

2.2. Genome assembly and phylogenetic analyses

An outline of both the genome assembly process and the phylogenetic analyses performed are shown in Fig. 1, however, to summarise, first the Illumina raw reads of the selected 167*C. jejuni* isolates underwent a quality control check using the QCtool pipeline developed by Mauro Truglio (unpublished; https://github.com/mtruglio/QCtool) to ensure potential contaminants such as PhiX control reads and adapter sequences were removed. The quality control report was then checked manually to exclude further poor-quality sequencing results which might have been deleterious in the subsequent analysis. Genome *de novo* assembly was completed using the software tool Spades (*v*3.10.1) (Bankevich et al., 2012) with the assembly run in "careful mode" to ensure mismatches were corrected. Assemblies were then annotated with Prokka (*v*1.12) (Seemann, 2014), and *C. jejuni* isolates were confirmed *in silico* by comparing the full length of the 16S ribosomal ribonucleic acid (rRNA) gene to the nucleotide collection (nr/nt) database hosted in GenBank using the online nucleotide BLAST (BLASTN) tool (Camacho et al., 2009).

Preliminary genotyping was then performed with a multilocus

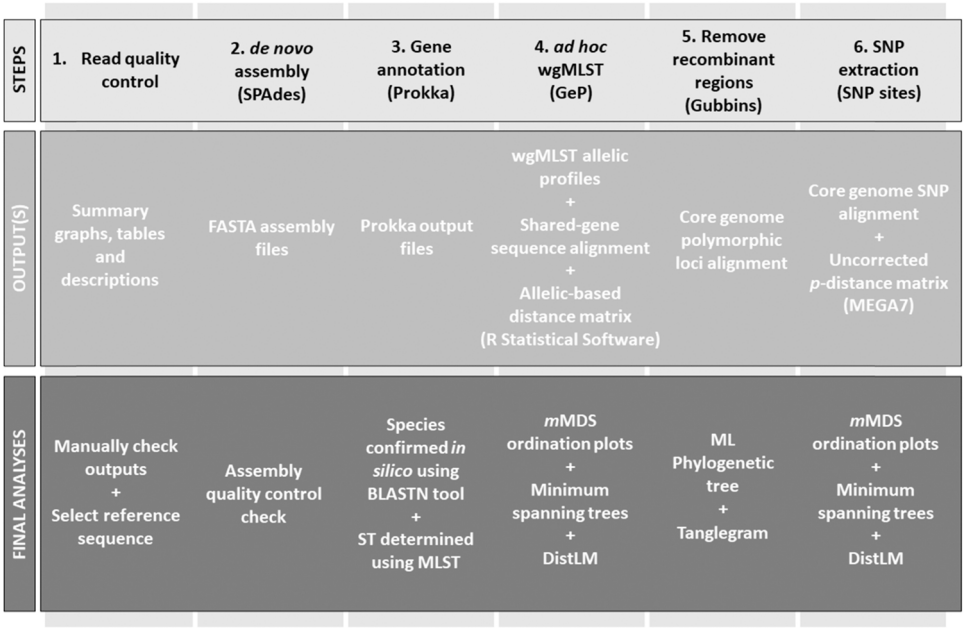

Fig. 1. Outline of the phylogenetic analysis showing the outputs of each step (1−6) including the genetic distance matrices used in the final distance-based redundancy models.

sequence type (MLST) analysis using a script developed by Torsten Seemann (unpublished). The relationship between the isolates was then determined by an *ad hoc* whole-genome MLST (wgMLST) analysis using Genome Profiler (GeP, *v*2.2) (Zhang et al., 2015) to convert assembly sequence data into wgMLST allelic profiles. For this analysis, a draft genome sequence of ST-6964 was selected as a reference sequence (PubMLST no. 78757) from the list of assemblies based on the quality control checks. Next, hypothetical recombinant regions were removed from the shared-gene sequence alignment concatenation produced by GeP (*v*2.2) (Zhang et al., 2015) using Gubbins (*v*2.2.0) (Croucher et al., 2015) to mitigate the effects of horizontal sequence transfer mechanisms on phylogenetic reconstructions (Boto, 2010). After the removal of recombinant regions, the alignment of polymorphic loci was used to construct a maximum-likelihood phylogenetic tree rooted by outgroup using the R package *phangorn* (Schliep, 2011) and allowing the genetic relatedness of isolates to be visualised.

2.3. Genetic distance matrices

To represent the relationship between each pair of *C. jejuni* ST-6964 isolates, two different genetic distance matrices were generated with one matrix representing the uncorrected *p*-distance between each isolate and the other matrix representing an allelic-based distance measure. The uncorrected *p*-distance was calculated by finding the proportion of nucleotide sites that differed between each pair of isolates without making any corrections for multiple substitutions at the same site or differences in the evolutionary rates among sites. This calculation was performed using MEGA7 (*v*7.0.26) (Kumar et al., 2016) and was based on a core-genome SNP alignment that was produced by extracting the single nucleotide polymorphic (SNP) loci from the filtered alignment using SNP sites (Page et al., 2016) (Fig. 1). The allelic-based distance was calculated from the wgMLST allelic profiles that had been produced by GeP (*v*2.2) (Zhang et al., 2015). The allelic profiles were used to compute a distance matrix based on the Manhattan distance measure (Craw, 2011) using R Statistical software (Fig. 1). Results obtained using these two matrices were highly consistent, so we report here only the results obtained using the allelic-based distance matrix with results from the *p*-distance matrix provided in the Supplementary Materials.

2.4. Model matrices

An additional five model matrices were constructed with the aim of relating these, both individually and collectively, to the genetic distances described above. Model matrices included: (i) the geographical Euclidean distance (based on latitude and longitude) between each pair of farms from which isolates were collected, (ii) the road distance between farms, and the network distance between farms based on the contact networks constructed from the on- and off-farm movements of (iii) feed, (iv) live birds and hatching eggs, and (v) waste and litter; with the methods for defining these described below.

The geographical distances between farms were calculated by obtaining a postal address for each farm from a database of commercial poultry producers registered with either the Poultry Industry Association of New Zealand (PIANZ) or the Egg Producers Federation of New Zealand (EPF). The addresses provided in the database were checked using Google Maps (2017) to make sure they specified a poultry production site (indicated by the presence of poultry sheds) and not the producers' residential address. Co-ordinates were collected and the two pairwise distance matrices (Euclidean distances and road distances) were calculated using the R packages *geosphere* (Hijmans, 2019) and *gmapsdistance* (Melo et al., 2018), respectively. The matrices were then expanded to express the geographical distance between each pair of isolates with a value of zero in the matrix indicating that the isolates were sampled from the same farm. The matrices were used to create hierarchical dendrograms representing the geographical distance between sequenced isolates using the R package *ape* (Paradis and Schliep, 2019) which were used, in turn, to construct tanglegrams to assist in visualising relationships between matrices (see the section "Relating genetic distances to model matrices").

The network distances between farms were calculated by constructing several contact networks from the reported on- and off-farm movements for each of the following: (i) feed, (ii) live birds and hatching eggs, or (iii) waste and litter. In each network, nodes represented the farms from which isolates were sampled, with an undirected edge linking nodes (*i.e.,* farms) utilising the same transport company (Fig. 2). The reported movements were obtained from the results of an industry survey administered to all active poultry producers in New Zealand registered with either PIANZ or EPF as of June 2016. The survey

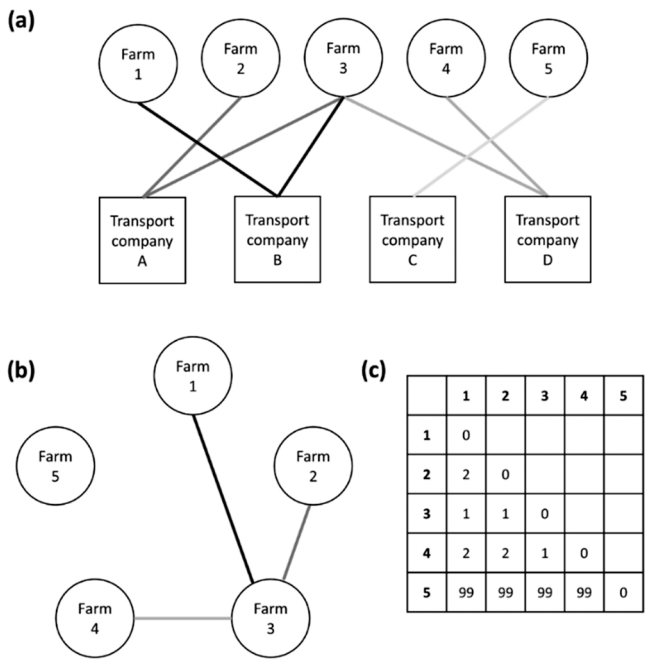

Fig. 2. Schematic diagram showing the construction of (a) bimodal and (b) unimodal network graphs. In network (a) there are two node types representing either farms (circles) or transporting companies (squares) moving on- and off-farm. In network (b) the farms have been directly linked if they have a transporting company in common. The shortest path (SP) between any two farms is the least number of steps needed to get to one farm from another in the unimodal network graph. For example, in network (b) Farm 4 can be reached from Farm 2 *via* Farm 3 with a SP equal to 2 as shown in (c) which is the shortest path matrix when considering all the transporting companies with 1 indicating there is a direct link between two farms in (b) and 99 indicating no link.

was based on a previous questionnaire conducted in 2006 (Lockhart et al., 2010) and modified in collaboration with PIANZ, EPF and the Ministry for Primary Industries (MPI) with the aim of collecting information on the farm demographics, contact patterns, and biosecurity practices of New Zealand commercial poultry operations. The study was judged to be low risk through peer evaluation and consequently was not formally reviewed by any of the University's Human Ethics Committees. Full details of the survey design and implementation have been described elsewhere (Greening et al., 2020) and a copy of the complete survey questionnaire is provided in the Supplementary Materials.

As not all farms from which isolates were collected completed the survey (*i.e.,* there were non-responders), two networks were constructed for each type of on- and off-farm movement: one including only the farms that responded (hereafter referred to as the "empirical network"), and one including all the farms with sampled isolates (hereafter referred to as the "imputed network"). For the latter, missing links caused by non-responders were inferred based on information supplied by an industry representative who was able to provide the names of the transporting companies that each non-responder was most likely utilising, based on their expert opinion. Results obtained using empirical *versus* imputed networks were highly congruent, so we report here only the results obtained using the imputed networks, with results from the empirical networks provided in the Supplementary Materials.

Network graphs were constructed using the R package *igraph* (Csardi and Nepusz, 2006) based on a force-based algorithm proposed by Fruchterman and Reingold (1991). We report here the network degree centrality and betweenness centrality for each network but note that other network statistics for these data have been described elsewhere (Greening et al., 2020). Network graphs were used to produce pairwise distance matrices with values representing the shortest path (*i.e.,* the minimum number of links) between each pair of farms in each network (Fig. 2). If a pair of farms were completely unconnected in the network, a numeric value of 99 was recorded in the matrix to represent a very large distance between pairs of isolates that were unconnected. In addition to network matrices, a community analysis was performed in which all the on- and off-farm movements were combined to produce a single network. Communities could then be identified, with farms in the same community having more internal links between them than external links to other communities within the network. This analysis was completed using a link community detection algorithm (Ahn et al., 2010) in the R package *linkcomm* (Kalinka and Tomancak, 2011).

2.5. Relating genetic distances and model matrices

The relationship between the genetic distance matrix and each of the individual model distance matrices was examined using a non-parametric Mantel test (Mantel, 1967). A robust version of the Mantel test was implemented in the RELATE routine in PRIMER (*v*7.0) (Clarke and Gorley, 2015), using Spearman's rank correlation (*rho*) as a measure of matrix correlation. Model matrices that had a statistically significant relationship with the genetic distance matrix ($p < 0.05$, 9999 permutations) were explored further in the formal linear models described in the next section "Explaining variation in genetic distances". To further examine if an association remained after removing any potential effects of individual farms and their parent company on the genetic distances, a residual genetic distance matrix (removing the effects of parent company and farms nested within parent company) was obtained using the method described in Anderson (2017). Mantel tests were then repeated (using the RELATE routine) to examine the correlation between this residual genetic distance matrix and each of the model matrices. To visualise the relative strengths of matrix associations, second-stage metric multidimensional scaling (*m*MDS) ordination plots (Somerfield and Clarke, 1995) were generated in PRIMER *v*7.0 (Clarke and Gorley, 2015) after calculating the matrix correlations between all pairs of distance matrices (*i.e.,* the allelic-based distance matrix, the geographic distance matrices and the network distance matrices). This allowed us to see not only the proximity of each model matrix to the genetic distance matrix but also showed visually the similarities among the various different model matrices. To help visualise these relationships, a minimum spanning tree was added as an overlay to the *m*MDS ordination. To examine if closely related isolates shared other common factors, additional *m*MDS ordination plots (Kruskal and Wish, 1978) were generated with plots mapping isolates in a two-dimensional Euclidean space in a manner that preserved the rank order of dissimilarities present in the underlying genetic distance matrices. Different colours were used to easily identify isolates sharing a common factor such as the parent company (*i.e.,* poultry supplier) and network community.

In addition to the Mantel tests and *m*MDS ordination plots, the relationship between pathogen phylogeny and the geographical proximity between all the farms (*i.e.,* Euclidean distance and road distance) was visualised using tanglegrams. Tanglegrams were constructed using the R package *dendextend* (Galili, 2015) such that isolates on the maximum-likelihood phylogenetic tree were connected *via* auxiliary lines to the corresponding isolate on the hierarchical dendrograms described above representing either the Euclidean distance or road distance between the isolates. To optimise each tanglegram, a two-tree crossing minimisation technique was used to minimise the number of crossings between the auxiliary lines. Tanglegrams were annotated to indicate the parent company of the farm from which the isolates were sampled, and a Spearman's rank correlation coefficient was calculated between the trees' cophenetic distances matrices (Sokal and Rohlf, 1962).

2.6. Explaining variation in genetic distances

To find parsimonious models to explain variation in genetic distances

using the information provided in the geographical and/or network distance matrices, we first generated Euclidean coordinates that would capture the information contained in each model distance matrix. This was done by calculating a number (m) of coordinate axes using mMDS to represent the information in each distance matrix. We found that, in every case, $m = 2$ or 3 was sufficient to capture the salient information contained in each of the model distance matrices (stress < 0.01). We also created regression coordinates *i.e.,* in the form of analysis of variance (ANOVA) contrasts, that coded for two additional factors; parent company and farms (nested in parent company) in order to fit the matrix models and ANOVA factors in a regression setting. A distance-based redundancy analysis (Legendre and Anderson, 1999; McArdle and Anderson, 2001) of genetic distances *versus* the network models and ANOVA factors was run using the DISTLM routine in the PERMANOVA+ add-on package (Anderson et al. 2008) for PRIMER (*v*7.0) (Clarke and Gorley, 2015). Note that the set of coordinates corresponding to specific network distance matrices (or ANOVA factors) were kept together as a group for model selection. Marginal tests were then performed on each set individually (with p-values obtained using 9999 unconstrained permutations) with forward selection based on R^2 used to uncover potential redundancies and regions of overlap in the explanatory power of regression variable sets (with p-values for sequential conditional tests at each step obtained using permutation of residuals under a reduced model, Freedman and Lane, 1983). Next, we performed model selection and searched for the best overall model (that is, combinations of sets of regression coordinates that best explain variation in the genetic distance among isolates), using the multivariate analogue to Akaike's "An Information Criterion" (AIC) (Anderson et al., 2008). During model selection, we forced the inclusion of two important known factors; parent company and farm (as ANOVA terms in the model).

3. Results

3.1. Sample collection

The geographical distribution amongst the selected subset of farms was limited to three regions in the North Island of New Zealand; with four farms from both suppliers A and B located in the Auckland region, two farms from supplier B and eleven farms from supplier C located in the Waikato region, and nine farms from supplier A located in the Taranaki region (Fig. 3). The number of selected samples from each supplier was unevenly distributed, with 57.5% (96/167) of the isolates sampled from supplier A, 35.9% (60/167) of isolates sampled from supplier B and 6.6% (11/167) sampled from supplier C.

3.2. Phylogenetic analyses

The rooted maximum-likelihood phylogenetic tree reconstructed from the core polymorphic loci alignment produced by Gubbins (*i.e.,* based on the core-genome MLST) is shown in Fig. 3. In this analysis, 230 polymorphic loci (including indels) were identified among the 167 isolates, including 204 SNP loci. The average SNP difference was 8.9 (minimum: 0, maximum 47) and the average allele difference was 12.5 (minimum: 0, maximum: 63). Within the phylogeny, there were three distinct genetic clusters, with the majority of isolates sampled from farms belonging to either the same poultry supplier or network community grouping within the same genetic cluster (Fig. 4), with the exception of a few single isolates that cluster with isolates sampled from farms belonging to neither the same supplier nor the same community. However, the tanglegram generated from the rooted maximum-likelihood phylogenetic tree and hierarchical dendrogram based on Euclidean distance shows that these single isolates are located within the same geographical region as other isolates in the cluster (Fig. 5). Despite this, there was only a weak correlation between the phylogenetic

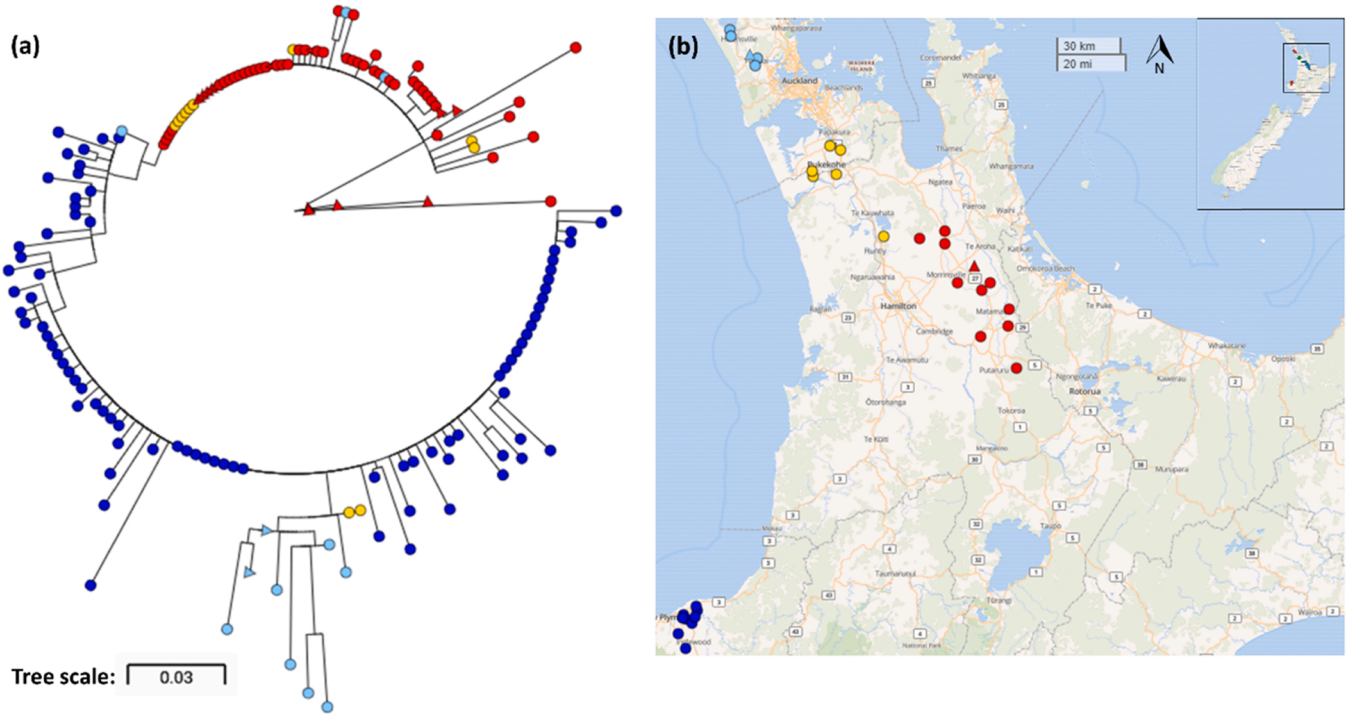

Fig. 3. (a) A maximum likelihood phylogenetic tree rooted by outgroup showing the genetic relatedness between 167 *C. jejuni* ST-6964 isolates based on the core polymorphic loci alignment following core-genome MLST. Production type is indicated by the shape of the terminal node (triangles: poultry breeder and circles: poultry grower) with the three colours indicating the poultry supplier (blue: supplier A, red: supplier B, and yellow: supplier C) to which the farm belongs. Nodes on the phylogeny also correspond to the points on (b) a map showing the geographical distribution of the 30 New Zealand commercial poultry enterprises from which sampled isolates were found positive for *C. jejuni* ST-6964. The figure has been created using the online tool Microreact (Argimón et al., 2016).

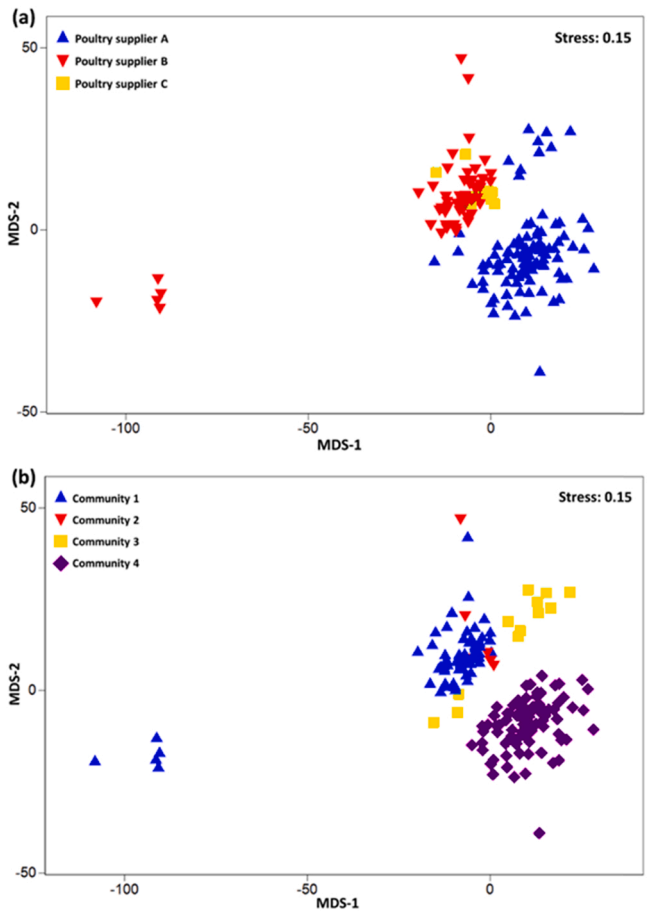

Fig. 4. Two-dimensional metric multidimensional scaling (*m*MDS) ordination plots based on the allelic dissimilarity matrix between 167*C. jejuni* ST-6964 isolates following wgMLST. Isolates are coloured according to (a) parent company and (b) network community.

distances and geographical distances with a Spearman *rho* value of 0.387 for Euclidean distance, and 0.385 for road distance. The tanglegram generated from the maximum-likelihood phylogenetic tree and the hierarchical dendrogram based on road distance is shown in the Supplementary Materials (Supplementary Fig. 1).

3.3. Characteristics of model matrices

The p-distance between isolates ranged from 0 to 0.16 (mean = 0.04) and the allelic-based Manhattan distance between isolates ranged from 0 to 129.78 (mean = 26.55). In the geographical distance matrices, the Euclidean distance between the farms ranged from 1.31 km to 282.08 km (mean = 131.73 km) whilst the road distance ranged from 1.32 km to 418.79 km (mean = 180.98 km). In the network distance matrices based on the imputed networks, no shortest path between any two farms was greater than 3; however, many farms remained unconnected. For example, in the feed network, 59.1% of the potential pathways between farms did not exist, while 39.8% of the pathways were a direct link between two farms. Both the live bird and waste networks had 43.4% of the pathways that were non-existent, while 33.8% and 36.3% were direct links between two farms, respectively. This suggests that in these networks, a greater number of farms were connected than not connected, although there were not so many direct links. The network graphs, shown in Supplementary Fig. 2, also indicated that live bird and waste networks were more cohesive, having larger betweenness centrality scores (Table 1). The high proportion of farms directly linked within the feed network reflects the small number of transporting companies operating within the network. The network graphs showed a clustering of farms around each other of the three companies operating within the feed network with no links between the clusters (Supplementary Fig. 2). In contrast, the live bird and waste networks had a greater number of links between clusters (*i.e.*, individual farms are using more than one transporting company). The community analysis performed on the imputed network combining all the on- and off-farm movements identified four communities (Supplementary Fig. 3) with 56.7% (17/30) of the farms belonging to the largest community. The *m*MDS plot, based on the allelic-based distance matrix, suggested that pathogens in the same network community have similar genetic structures (Fig. 4b).

3.4. Relating genetic distances and model matrices

All model matrices showed a significant relationship ($p < 0.001$) with the unconstrained allelic dissimilarity matrix. The Spearman rank matrix correlation coefficients ranged from 0.438 to 0.632 (Table 2). In addition, parent company was significantly associated with the genetic structure of isolates, and there was also significant variation due to individual farms (Supplementary Materials, Table 1). After removing the effects of farm and parent company, only the feed network had a significant matrix correlation with the residual allelic dissimilarity matrix ($p < 0.05$). The second-stage *m*MDS plot showed that Euclidean distance and road distance matrices were very highly correlated, as were the live bird and waste networks, whilst minimum spanning trees indicated that the feed network had the closest relationship with the allelic dissimilarities among pathogen isolates (Fig. 6).

3.5. Explaining variation in genetic distances

When considered individually (marginal tests), road distance accounted for the largest proportion of the variation in the allelic dissimilarity matrix, followed by the Euclidean distance and the feed network (Table 3). Sequential tests showed that, after the addition of road distance, farm, parent company, and feed network, the remaining networks added little to explain genetic variation (Table 4; $p > 0.50$). Indeed, the model with the lowest AIC value (out of the class of models that included perforce the parent company and the farm) included only feed network and road distance (Table 5). A similar result was found for the empirical networks (Supplementary Materials, Table 4–8). However, there was a slight discrepancy in the models built using the alternative genetic distance measure, the *p*-distance, which did not include the feed network in the model having the lowest AIC value (Supplementary Materials, Tables 9–18).

4. Discussion

This study combined network data with molecular data in PERMANOVA and DistLM analyses to identify transmission pathways important to the spread of *C. jejuni* in the New Zealand poultry industry. Overall, results showed that the genetic distance between isolates from different poultry suppliers was greater than the distance between isolates from the same supplier. This finding supports the association between poultry supplier and pathogen phylogeny that has been reported in previous studies (McTavish et al., 2008; Müllner et al., 2010).

After removing the effects of farm and supplier, this study also found an association between the network distance of transporting feed vehicles and the genetic relatedness of the 167*C. jejuni* ST-6964 isolates sampled across 30 farms. Although the feed network accounted for a smaller contribution to the variation in genetic distances in the distance-based redundancy analysis, compared to farm, supplier and road distances, these results suggest that the transportation of feed within the commercial poultry industry may play a role in the spread of *C. jejuni* between flocks. Such an association could be attributed to transporting vehicles and/or the feed itself acting as a vector. However, it is also

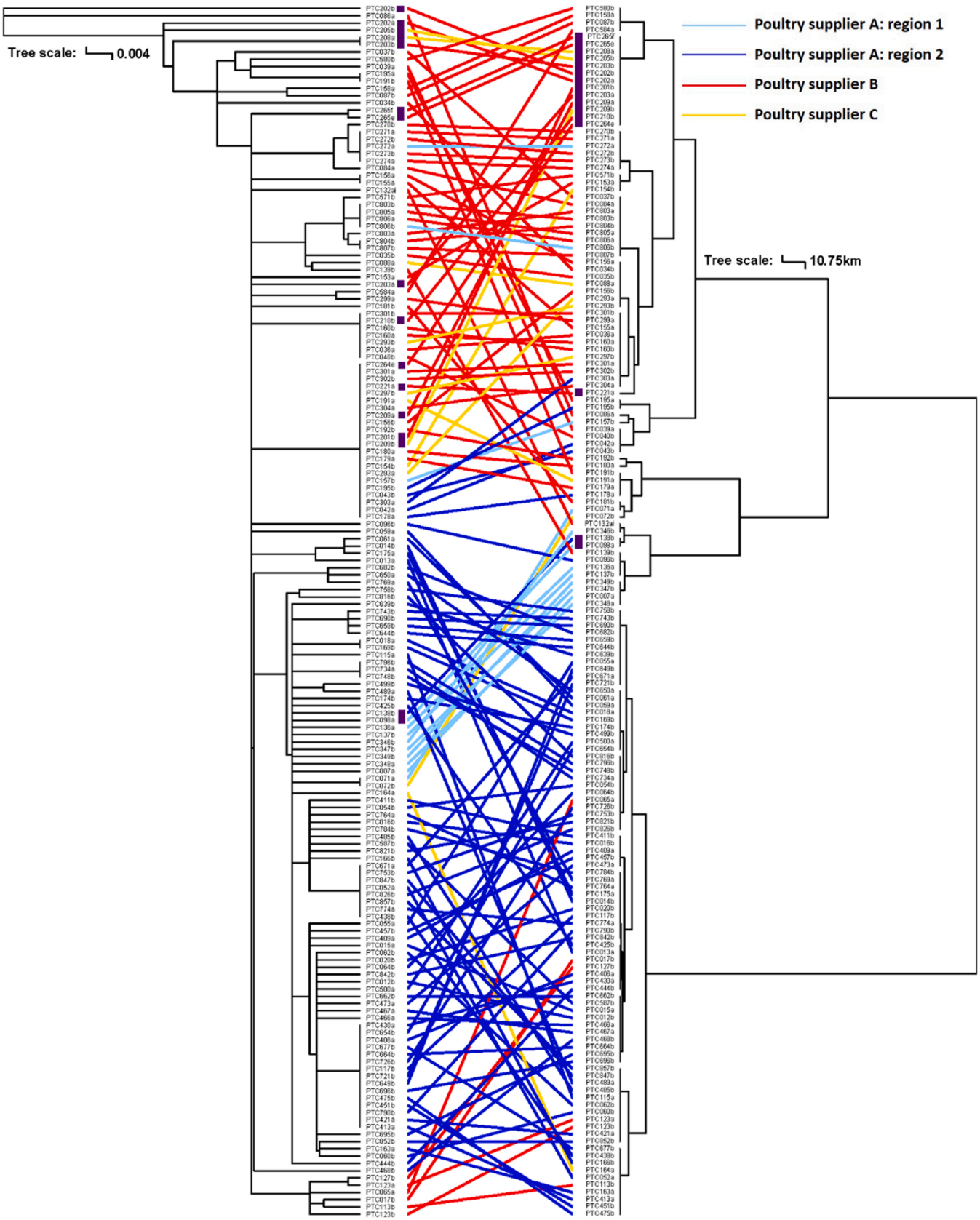

Fig. 5. A tanglegram containing a maximum-likelihood phylogenetic tree rooted by outgroup showing the population structure of 167 *C. jejuni* ST-6964 isolates based on the core polymorphic loci alignment following core-genome MLST (left) compared to a dendrogram representing the Euclidean distances between the farms from which isolates were sampled (right). The colour of the connecting line indicates the poultry supplier of each farm (A, B, or C) with farms belonging to supplier A located in two geographical regions (regions 1 and 2) in comparison to poultry suppliers B and C whose farms are geographically clustered in one region. A purple square following the isolate ID indicates the farm is a poultry breeder whilst no square indicates it is a poultry grower.

Table 1
Network statistics describing both the empirical network (*i.e.,* the network constructed from 16 survey responses reporting all on- and off-farm movements over a one-year period) and the imputed network (*i.e.,* the network constructed using all 30 farms from which isolates were sampled, with missing edges imputed based on expert opinion) constructed from all on- and off-farm movements relating to (i) feed, (ii) live birds and hatching eggs or (iii) waste and litter within the New Zealand commercial poultry industry. Network metrics include the "Degree", indicating the total number of on- and off-farm movements on a single farm in the network and, "Betweenness" indicating the frequency a farm is on the shortest path between any two other farms in the network.

Network metric	Network	Empirical network	Imputed network
Number of nodes	All	16	30
Number of Edges	Feed	208	455
	Live birds & hatching eggs	457	1121
	Waste & litter	367	776
	Combined	1032	2352
Mean degree (min-max)	Feed	6.25 (0–8)	11.53 (3–16)
	Live birds & hatching eggs	57.12 (4–94)	74.73 (18–109)
	Waste & litter	5.63 (1–8)	10.53 (4–17)
	Combined	6.75 (3–8)	13.73 (8–20)
Mean betweenness (min-max)	Feed	0.06 (0–0.25)	0.17 (0–0.45)
	Live birds & hatching eggs	0.56 (0–4)	3.57 (0–90)
	Waste & litter	1.13 (0–9)	3.13 (0–28)
	Combined	0.19 (0–1)	1.33 (0–6)

Table 2
Spearman's rank matrix correlation (*rho*) between each model matrix and (i) the allelic dissimilarity matrix between 167 *C. jejuni* ST-6964 isolates and (ii) the residual allelic dissimilarity matrix after fitting the ANOVA factors of parent company (n = 3) and farm (n = 30) nested within parent company, with *p*-values obtained using 9999 permutations.

	(i) Unconstrained matrix		(ii) Residual matrix	
	rho	p-value	rho	p-value
Feed	0.632	0.0001	0.040	0.0370
Live birds	0.623	0.0001	0.001	0.4154
Waste	0.614	0.0001	0.005	0.3336
Road distance	0.584	0.0001	-0.005	0.5491
Euclidean distance	0.580	0.0001	-0.024	0.8023
Parent company	0.438	0.0001	–	–

important to consider that the transportation of feed may be a proxy for other contact networks that have not been captured in this analysis. For instance, those farms that share the same feed companies, may also share the same catching companies used in the thinning or depopulation of flocks, which may result in similarities between the networks (*i.e.,* the movement of transporting feed vehicles and the movement of catching companies).

A correlation between the different networks can be seen in the *m*MDS plots, for example between the live bird and waste networks. This is not surprising given the vertical integration of most broiler operations in New Zealand with individual producers relying on suppliers to control all stages from primary production and processing to distribution. However, the network of transporting feed vehicles remains the most disparate in comparison to all the other transporting networks in this study, while showing the closest relationship with the genetic distance matrices of the *C. jejuni* isolates. This does not account for networks that are absent from this study, highlighting the importance for future studies to collect more information on farm practices particularly those that are known to play a role in the spread of *Campylobacter* such as thinning (Ellis-Iversen et al., 2012; van Wagenberg et al., 2016).

Study results also showed an association between the road and Euclidean distance between farms and the genetic relatedness of the 167 *C. jejuni* ST-6964 isolates. This association suggests that local

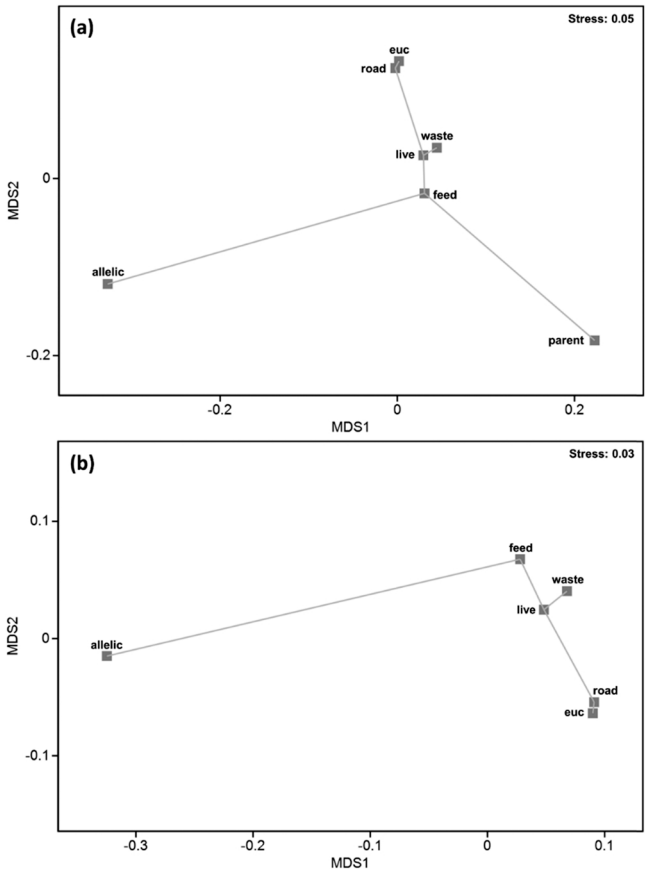

Fig. 6. Second-stage metric multidimensional scaling (*m*MDS) ordination plots showing proximities (rank matrix correlations) between the allelic dissimilarity matrix ("allelic") and each of several model distance matrices: "euc" = Euclidean distance; "road" = road distance; "feed" = feed network distance; "live" = live birds network distance; "waste" = waste and litter network distance; "parent" = parent company distance. Superimposed on each plot is the minimum spanning tree (light grey lines). The *m*MDS plot shown in (a) includes parent company whereas (b) does not include parent company.

Table 3
Individual distance-based redundancy analysis models to explain variation in allelic dissimilarities among 167 *C. jejuni* ST-6964 isolates in response to each of two ANOVA factors (parent company or farm nested in parent company) or sets of regression coordinates corresponding to the geographic position (Euclidean distance or road distance) or the network model matrices of interest (feed, live birds or waste), with *p*-values for each of these separate marginal tests obtained using 9999 unrestricted permutations. "Prop" gives the proportion of the total variation explained whilst "*df*" gives the numerator (regression) and denominator (residual) degrees of freedom for the test. Models have been presented in order of decreasing R^2 values.

Model matrix	Prop.	Pseudo-F	*df*	*p*-value
Parent Company	0.2649	29.55	3, 164	0.0001
Waste	0.2854	32.76	3, 164	0.0001
Farm	0.2894	2.10	28, 139	0.0117
Live birds	0.2912	33.69	3, 164	0.0001
Feed	0.3184	25.39	4, 163	0.0001
Euclidean distance	0.3458	43.33	3, 164	0.0001
Road distance	0.4112	37.95	4, 163	0.0001

contacts may also play a role in pathogen spread, although the exact mechanism cannot be deduced without further information since many local pathways such as the movement of wildlife (Craven et al., 2000; Nichols, 2005) or personnel (Meerburg and Kijlstra, 2007; Slader et al., 2002) are known to contribute to the spread of *C. jejuni*. Additional

Table 4
Distance-based redundancy analysis to explain variation in allelic dissimilarities among 167 *C. jejuni* ST-6964 isolates in response to the factors and sets of regressors listed in Table 3, but here conditional tests were done in a sequential stepwise fashion under forward selection based on R^2. Each test used 9999 permutations of residuals under a reduced model. "Prop" gives the proportion of additional variation explained by adding that set of variables to the model, "Cumul" tracks the cumulative explained variation with each added step, and "*df*" provides the regression and residual degrees of freedom. Note: from step 5 onwards, the additional explained variation is < 1%.

Step		Prop.	Cumul.	*df*	Pseudo-F	*p*-value
1	+Road	0.4112	0.4112	4, 163	37.95	0.0001
2	+Farm	0.1475	0.5588	31, 136	1.684	0.0438
3	+Parent company	0.0156	0.5743	33, 134	2.453	0.0497
4	+Feed	0.0192	0.5935	36, 131	2.063	0.0839
5	+Live birds	0.0052	0.5987	38, 129	0.8342	0.5017
6	+Euclidean	0.0030	0.6018	40, 127	0.4805	0.6538
7	+Waste	0.0025	0.6042	42, 125	0.3891	0.8696

Table 5
Top four distance-based redundancy models obtained on the basis of the multivariate analogue to the Akaike Information Criterion (AIC) (see text) to explain variation in allelic dissimilarities among 167 *C. jejuni* ST-6964 isolates. Two ANOVA factors (parent company and farms nested in parent company) were included in all potential models *a priori*.

Model selections	R^2	No. Sets	AIC
Feed network, road distance, parent company and farm	0.5935	4	975.23
Road distance, parent company and farm	0.5743	3	976.94
Feed network, live bird network, road distance, parent company and farm	0.5987	5	977.09
Euclidean distance, feed network, road distance, parent company and farm	0.5964	5	978.06

sampling may help to discern between the different local mechanisms, for instance, a study by Ridley et al. (2011) was able to match *Campylobacter* sequences sampled from farms with those isolated from the lunch bags of employees; emphasizing the movement of personnel as a major risk factor.

In addition to adding different sampling sites, increasing the number of samples per farm as well as the number of farms using longitudinal sampling, will increase the sensitivity of detection, aid phylogenetic reconstruction (Nabhan and Sarkar, 2012), and provide a more complete representation of the distribution, evolution, and transmission of the pathogen of interest. The farms sampled in this study are likely to represent a relatively small sample of all the infected farms, sampled over a relatively short time period, with limited information on likely periods of exposure to transmission from other farms or sources. For this reason, no attempts were made to infer transmission between farms using, for example, an ancestral state reconstruction approach such as a structured coalescent model (*e.g.,* De Maio et al., 2016). However, future studies may want to consider using a phylodynamic framework as new methods are developed that incorporate contact network data and other epidemiological variables, to evaluate the role of direct and indirect transmission pathways (Firestone et al., 2020; Featherstone et al., 2021).

Further considerations could also include the comparison of accessory genes and antimicrobial resistance (AMR) data ranging from a simple comparison between the presence of AMR genes to using AMR data in phylodynamic models (Ingle et al., 2021). For this study, AMR data was not explored further as all isolates were previously shown to be genotypically and phenotypically resistant to tetracycline and fluoroquinolones (French et al., 2019). Furthermore, out of eight mobile genetic elements that have previously been examined, only three (the pTet-like plasmid and the *C. jejuni* integrated elements CJIE1v and CJIE4) were differentially distributed among the isolates in this study (French et al., 2019). This helps to confirm the low genetic diversity between the isolates, particularly those isolated from the same poultry supplier.

Another limitation of our study was the low response rate to the industry survey, creating the potential for both non-response bias and recall bias. Overall, only 53.3% (16/30) of the farms in the study had movement data that could be used to construct the contact networks, and although movement data for non-responders was collected based on expert opinion, it is hard to quantify the impact of this non-response bias. However, the similarity in results obtained using the imputed and the empirical networks provided some reassurance regarding the network imputation methods used. By using a survey questionnaire, the study results are also likely to suffer from limitations inherent to most questionnaires, such as recall bias, with many previous studies showing a poor correlation between survey responses and on-farm practices (Bewsell, 2010; Racicot et al., 2012; Sax et al., 2003). In this study, producers were asked to recall all on- and off-farm movements over a one-year period. This was thought to be appropriate due to the steady nature of the poultry industry and the relationship between parent company and on-farm contractors, making it easier for producers to name all the transporting companies that deliver goods or services such as feed. However, over a one-year period, it may be hard for producers to recall more informal movements, such as the on- and off-farm movement of farm visitors, or irregular movements such as the redistribution of left-over feed at the end of a production cycle. By only constructing networks using the small proportion of broiler farms that were sampled here, artefacts may also arise in community analysis. For instance, farms identified within a community may only belong to that community when considering a particular subset of network nodes and edges (Shizuka and Farine, 2016).

In addition to the effects of network imputation, this study also looked at the effect of using two different genetic distance measures on the relationship between the model matrices and the genetic relatedness of the *C. jejuni* isolates. The discrepancies between the final models fitted to different genetic distance dissimilarity matrices create some doubt as to the exact contribution of transporting feed vehicles in the spread *C. jejuni* ST-6964. These discrepancies may be explained, however, by the resolution captured within each of the matrices. Overall, there should be greater resolution within the allelic dissimilarity matrix produced using wgMLST compared to the *p*-distance matrix as the latter was produced using a SNP-based approach, which considers only the shared genome with recombinant regions removed (Allard et al., 2016; Schürch et al., 2018; Zhang et al., 2015). The relatively low resolution of the SNP based approach is also compounded by the small timeframe over which isolates were collected. Similar discrepancies depending on the genetic distance used have also been described in previous studies. For example, Shirk et al. (2017) used population genetic simulations to assess the accuracy of 16 individual-based genetic distance metrics. Results from this study found that most metrics performed well when sample size and genetic structure was high but less so when sample size and genetic structure was low. A similar comparison between multiple genetic distance metrics, including more than just the SNP-based and allelic distance matrix presented in this study, to assess their relative performance and maximize model selection accuracy may aid future studies.

Given these limitations, it is clear that further work is needed to understand why and how these contact networks may be important for the spread of *C. jejuni.* In particular, being able to identify the attributes within the feed network and the pathways captured by both road distance and Euclidean distance contributing to transmission would help inform targeted control activities within the commercial poultry industry. Future research will rely on both the participation of producers and innovative technology to capture and track all on- and off-farm movements that may be contributing to the spread of disease. For example, Global Positioning Systems (GPS) may be used to track the spatial pattern of transporting vehicles within livestock contact

networks, or new recording systems may be developed to aid the producer to track all vehicles entering their farm, such as the OnSide® smartphone-based app (http://www.onside.co.nz/). Further research into on-farm biosecurity is also important to see if producers are identifying these pathways as major risks for transmission and if they are implementing any risk management strategies to target these pathways such as the disinfection of vehicles. Without this additional information, it will be difficult to recommend suitable control strategies that may target pathways contributing to the spread of *C. jejuni* within the commercial poultry industry.

5. Conclusion

This study highlights the importance of both local transmission and contact networks in the spread of *C. jejuni* within the commercial poultry industry. This includes not only transmission between farms sharing the same parent company but also between farms in different supply chains. Although this spread is likely to be facilitated by the small number of companies servicing the industry, particularly those few delivering and transporting feed, additional research is required to fully characterise these risk pathways and to develop appropriate control strategies that would reduce the spread of *C. jejuni* within the commercial poultry industry.

Funding

This work was supported by the Poultry Industry Association of New Zealand (PIANZ), the Egg Producers' Federation of New Zealand (EPF), the Ministry for Primary Industries (MPI), Massey University School of Veterinary Science, and the Biological Heritage National Science Challenge. In addition, JZ, DAW, and NPF were funded by the New Zealand Food Safety Science & Research Centre (NZFSSRC).

CRediT authorship contribution statement

ACM, DAW (Massey University), and DAW (University of Melbourne) were responsible for sample collection, processing, and whole-genome sequencing, while AF was responsible for the database design and entry of poultry survey data. SG conducted the analyses with guidance from CG, JZ, NPF, and MJA. SG also drafted the manuscript which all authors provided feedback on and approved for submission.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

The authors are grateful to all the poultry producers who took the time to complete the survey questionnaire and the industry representative who provided estimations of transport company utilisation.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.epidem.2021.100521.

References

Ahn, Y.-Y., Bagrow, J.P., Lehmann, S., 2010. Link communities reveal multiscale complexity in networks. Nature 466, 761–764. https://doi.org/10.1038/nature09182.

Allard, M.W., Strain, E., Melka, D., Bunning, K., Musser, S.M., Brown, E.W., Timme, R., 2016. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. J. Clin. Microbiol. 54, 1975–1983. https://doi.org/10.1128/JCM.00081-16.

Anderson, M.J., Gorley, R.N., Clarke, K.R., 2008. PERMANOVA+ for PRIMER: A Guide to Software and Statistical Methods. Primer-E Ltd,, Plymouth, UK.

Anderson, M.J., 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). John Wiley & Sons, Ltd., Chichester, UK, pp. 1–15 (Wiley StatsRef: Statistics Reference Online).

Antonovics, J., Wilson, A.J., Forbes, M.R., Hauffe, H.C., Kallio, E.R., Leggett, H.C., Longdon, B., Okamura, B., Sait, S.M., Webster, J.P., 2017. The evolution of transmission mode. Philos. Trans. R. Soc. B Biol. Sci. 372, e20160083 https://doi.org/10.1098/rstb.2016.0083.

Argimón, S., Abudahab, K., Goater, R.J.E., Fedosejev, A., Bhai, J., Glasner, C., Feil, E.J., Holden, M.T.G., Yeats, C.A., Grundmann, H., Spratt, B.G., Aanensen, D.M., 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb. Genom. 2, e000093 https://doi.org/10.1099/mgen.0.000093.

Baines, S.L., Howden, B.P., Heffernan, H., Stinear, T.P., Carter, G.P., Seemann, T., Kwong, J.C., Ritchie, S.R., Williamson, D.A., 2016. Rapid emergence and evolution of *Staphylococcus aureus* clones harbouring *fusC*-containing staphylococcal cassette chromosome elements. Antimicrob. Agents Chemother. 60, 2359–2365. https://doi.org/10.1128/AAC.03020-15.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Bansal, S., Read, J., Pourbohloul, B., Meyers, L.A., 2010. The dynamic nature of contact networks in infectious disease epidemiology. J. Biol. Dyn. 4, 478–489. https://doi.org/10.1080/17513758.2010.503376.

Bewsell, D., 2010. Determining Individuals' Response to New Zealand Biosecurity. Lincoln University,, Lincoln, New Zealand.

Boto, L., 2010. Horizontal gene transfer in evolution: facts and challenges. Proc. R. Soc. B Biol. Sci. 277, 819–827. https://doi.org/10.1098/rspb.2009.1679.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinform. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Chaters, G.L., Johnson, P.C.D., Cleaveland, S., Crispell, J., de Glanville, W.A., Doherty, T., Matthews, L., Mohr, S., Nyasebwa, O.M., Rossi, G., Salvador, L.C.M., Swai, E., Kao, R.R., 2019. Analysing livestock network data for infectious disease control: an argument for routine data collection in emerging economies. Philos. Trans. R. Soc. B Biol. Sci. 374, e20180264 https://doi.org/10.1098/rstb.2018.0264.

Clarke, K., Gorley, R., 2015. PRIMER v7: User Manual/Tutorial. PRIMER-E, Plymouth, UK.

Craft, M.E., 2015. Infectious disease transmission and contact networks in wildlife and livestock. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 370, e20140107 https://doi.org/10.1098/rstb.2014.0107.

Craven, S.E., Stern, N.J., Line, E., Bailey, J.S., Cox, N.A., Fedorka-Cray, P., 2000. Determination of the incidence of *Salmonella* spp., *Campylobacter jejuni*, and *Clostridium perfringens* in wild birds near broiler chicken houses by sampling intestinal droppings. Avian Dis 44, 715–720. https://doi.org/10.2307/ad.1593118.

Craw, S., 2011. Manhattan Distance. In: Sammut, C., Webb, G.I. (Eds.), Encyclopaedia of Machine Learning. Springer, Boston, Massachusetts, USA. https://doi.org/10.1007/978-0-387-30164-8_506.

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., Harris, S.R., 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res 43, e15. https://doi.org/10.1093/nar/gku1196.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal. Complex Systems 1695 (http://igraph.org).

De Maio, N., Wu, C.-H., Wilson, D.J., 2016. SCOTTI: Efficient reconstruction of transmission within outbreaks with the structured coalescent. PLoS Comput. Biol. 12, e1005130 https://doi.org/10.1371/journal.pcbi.1005130.

Dingle, K.E., McCarthy, N.D., Cody, A.J., Peto, T.E.A., Maiden, M.C.J., 2008. Extended sequence typing of *Campylobacter spp.*, United Kingdom. Emerg. Infect. Dis. 14, 1620–1622. https://doi.org/10.3201/eid1410.071109.

Eames, K., Bansal, S., Frost, S., Riley, S., 2015. Six challenges in measuring contact networks for use in modelling. Epidemics 10, 72–77. https://doi.org/10.1016/j.epidem.2014.08.006.

Ellis-Iversen, J., Ridley, A., Morris, V., Sowa, A., Harris, J., Atterbury, R., Sparks, N., Allen, V., 2012. Persistent environmental reservoirs on farms as risk factors for *Campylobacter* in commercial poultry. Epidemiol. Infect. 140, 916–924. https://doi.org/10.1017/S095026881100118X.

Enns, E.A., Brandeau, M.L., 2011. Inferring model parameters in network-based disease simulation. Health Care Manag. Sci. 14, 174–188. https://doi.org/10.1007/s10729-011-9150-2.

Eubank, S., Guclu, H., Anil Kumar, V.S., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban social networks. Nature 429, 180–184. https://doi.org/10.1038/nature02541.

Featherstone, L.A., Di Giallonardo, F., Holmes, E.C., Vaughan, T.G., Duchêne, S., 2021. Infectious disease phylodynamics with occurrence data. Methods Ecol Evol. 1–10. https://doi.org/10.1111/2041-210X.13620.

Firestone, S.M., Hayama, Y., Lau, M.S.Y., Yamamoto, T., Nishi, T., Bradhurst, R.A., Demirhan, H., Stevenson, M.A., Tsutsuiet, T., 2020. Transmission network reconstruction for foot-and-mouth disease outbreaks incorporating farm-level covariates. PLoS ONE 15, e0235660. https://doi.org/10.1371/journal.pone.0235660.

Freedman, D., Lane, D., 1983. A nonstochastic interpretation of reported significance levels. J. Bus. Econ. Stat 1, 292–298. https://doi.org/10.1080/07350015.1983.10509354.

French, N.P., Zhang, J., Carter, G.P., Midwinter, A.C., Biggs, P.J., Dyet, K., Gilpin, B.J., Ingle, D.J., Mulqueen, K., Rogers, L.E., Wilkinson, D.A., Greening, S.S., Muellner, P., Fayaz, A., Williamson, D.A., 2019. Genomic analysis of fluoroquinolone- and tetracycline-resistant *Campylobacter jejuni* sequence type 6964 in humans and poultry, New Zealand, 2014–2016. Emerg. Infect. Dis. 25, 2226–2234. https://doi.org/10.3201/eid2512.190267.

Fruchterman, T.M.J., Reingold, E.M., 1991. Graph drawing by force-directed placement. Softw. Pract. Exp. 21, 1129–1164. https://doi.org/10.1002/spe.4380211102.

Galili, T., 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31, 3718–3720. https://doi.org/10.1093/bioinformatics/btv428.

Gilbertson, M.L.J., Fountain-Jones, N.M., Craft, M.E., 2018. Incorporating genomic methods into contact networks to reveal new insights into animal behaviour and infectious disease dynamics. Behaviour 155, 759–791. https://doi.org/10.1163/1568539X-00003471.

Greening, S.S., Mulqueen, K., Rawdon, T.G., French, N.P., Gates, M.C., 2020. Estimating the level of disease risk and biosecurity on commercial poultry farms in New Zealand. N. Z. Vet. J. 18, 1–11. https://doi.org/10.1080/00480169.2020.1746208.

Grenfell, B.T., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303, 327–332. https://doi.org/10.1126/science.1090727.

Hijmans, R.J., 2019. geosphere: spherical trigonometry. R package version 1.5–10. ⟨https://CRAN.R-project.org/package=geosphere⟩.

Ingle, D.J., Howden, B.P., Duchene, S., 2021. Development of phylodynamic methods for bacterial pathogens. Trends Microbiol 29, 788–797. https://doi.org/10.1016/j.tim.2021.02.008.

Jombart, T., Cori, A., Didelot, X., Cauchemez, S., Fraser, C., Ferguson, N., 2014. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. PLoS Comput. Biol. 10, e1003457 https://doi.org/10.1371/journal.pcbi.1003457.

Kaakoush, N.O., Castaño-Rodríguez, N., Mitchell, H.M., Man, S.M., 2015. Global epidemiology of *Campylobacter* infection. Clin. Microbiol. Rev. 28, 687–720. https://doi.org/10.1128/CMR.00006-15.

Kalinka, A.T., Tomancak, P., 2011. linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. Bioinforma. 27, 2011–2012. https://doi.org/10.1093/bioinformatics/btr311.

Koutsoumanis, K.P., Lianou, A., Gougouli, M., 2016. Latest developments in foodborne pathogens modelling. Curr. Opin. Food Sci. 8, 89–98. https://doi.org/10.1016/j.cofs.2016.04.006.

Kruskal, J.B., Wish, M., 1978. Multidimensional Scaling. Sage University Paper Series on Quantitative Applications in the Social Sciences. Sage Publications,, Newbury Park, California.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol 33, 1870–1874. https://doi.org/10.1093/molbev/msw054.

Legendre, P., Anderson, M.J., 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol. Monogr. 69, 1–24. https://doi.org/10.1890/0012-9615.

Leventhal, G.E., Kouyos, R., Stadler, T., Wyl, V., von, Yerly, S., Böni, J., Cellerai, C., Klimkait, T., Günthard, H.F., Bonhoeffer, S., 2012. Inferring epidemic contact structure from phylogenetic trees. PLoS Comput. Biol. 8, e1002413 https://doi.org/10.1371/journal.pcbi.1002413.

Lin, J., 2009. novel approaches for *Campylobacter* control in poultry. Foodborne Pathog. Dis. 6, 755–765. https://doi.org/10.1089/fpd.2008.0247.

Lockhart, C.Y., Stevenson, M.A., Rawdon, T.G., Gerber, N., French, N.P., 2010. Patterns of contact within the New Zealand poultry industry. Prev. Vet. Med. 95, 258–266. https://doi.org/10.1016/j.prevetmed.2010.04.009.

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

McArdle, B.H., Anderson, M.J., 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 82, 290–297. https://doi.org/10.1890/0012-9658.

McTavish, S.M., Pope, C.E., Nicol, C., Sexton, K., French, N., Carter, P.E., 2008. Wide geographical distribution of internationally rare *Campylobacter* clones within New Zealand. Epidemiol. Infect. 136, 1244–1252. https://doi.org/10.1017/S0950268807009892.

Meerburg, B.G., Kijlstra, A., 2007. Role of rodents in the transmission of *Salmonella* and *Campylobacter*. J. Sci. Food Agric. 87, 2774–2781. https://doi.org/10.1002/jsfa.3004.

Melo, R., Rodriguez, D., Zarruk, D., 2018. gmapsdistance: distance and travel time between two points from Google maps. ⟨https://CRAN.R-project.org/package=gmapsdistance⟩.

Muellner, P., Kells, N., Campbell, D., 2016. Risk profile: the emergence of *Campylobacter jejuni* ST-6964 in poultry in New Zealand and its associated antimicrobial resistance. MPI Technical Paper No: 2016/16. Wellington, New Zealand.

Müllner, P., Collins-Emerson, J.M., Midwinter, A.C., Carter, P., Spencer, S.E., van der Logt, P., Hathaway, S., French, N.P., 2010. Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. Appl Env. Microbiol 76, 2145–2154. https://doi.org/10.1128/AEM.00862-09.

Müllner, P., Spencer, S.E.F., Wilson, D.J., Jones, G., Noble, A.D., Midwinter, A.C., Collins-Emerson, J.M., Carter, P., Hathaway, S., French, N.P., 2009. Assigning the source of human campylobacteriosis in New Zealand: A comparative genetic and epidemiological approach. Infect. Genet. Evol. 9, 1311–1319. https://doi.org/10.1016/j.meegid.2009.09.003.

Nabhan, A.R., Sarkar, I.N., 2012. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. Brief. Bioinform. 13, 122–134. https://doi.org/10.1093/bib/bbr014.

Newell, D.G., Elvers, K.T., Dopfer, D., Hansson, I., Jones, P., James, S., Gittins, J., Stern, N.J., Davies, R., Connerton, I., Pearson, D., Salvat, G., Allen, V.M., 2011. Biosecurity-based interventions and strategies to reduce *Campylobacter spp*. on poultry farms. Appl. Environ. Microbiol. 77, 8605–8614. https://doi.org/10.1128/AEM.01090-10.

Nichols, G.L., 2005. Fly transmission of *Campylobacter*. Emerg. Infect. Dis. 11, 361–364. https://doi.org/10.3201/eid1103.040460.

Nohra, A., Grinberg, A., Midwinter, A.C., Marshall, J.C., Collins-Emerson, J.M., French, N.P., 2016. Molecular epidemiology of *Campylobacter coli* strains isolated from different sources in New Zealand between 2005 and 2014. Appl. Environ. Microbiol. 82, 4363–4370. https://doi.org/10.1128/AEM.00934-16.

Nohra, A., Grinberg, A., Marshall, J.C., Midwinter, A.C., Collins-Emerson, J.M., French, N.P., 2020. Shifts in the molecular epidemiology of *Campylobacter jejuni* infections in a sentinel region of New Zealand following implementation of food safety interventions by the poultry industry. e01753-19 Appl. Environ. Microbiol. 86. https://doi.org/10.1128/AEM.01753-19.

Olson, C.K., Ethelberg, S., van Pelt, W., Tauxe, R.V., 2008. Epidemiology of *Campylobacter jejuni* infections in industrialized nations (Campylobacter Pg.). ASM Press,, Washington, DC, pp. 163–189 (Campylobacter Pg.).

Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. Genom 2, e000056. https://doi.org/10.1099/mgen.0.000056.

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Racicot, M., Venne, D., Durivage, A., Vaillancourt, J.-P., 2012. Evaluation of the relationship between personality traits, experience, education and biosecurity compliance on poultry farms in Québec, Canada. Prev. Vet. Med. 103, 201–207. https://doi.org/10.1016/j.prevetmed.2011.08.011.

Ridley, A., Morris, V., Gittins, J., Cawthraw, S., Harris, J., Edge, S., Allen, V., 2011. Potential sources of *Campylobacter* infection on chicken farms: contamination and control of broiler-harvesting equipment, vehicles and personnel. J. Appl. Microbiol 111, 233–244. https://doi.org/10.1111/j.1365-2672.2011.05038.x.

Sah, P., Otterstatter, M., Leu, S.T., Leviyang, S., Bansal, S., 2018. Revealing mechanisms of infectious disease transmission through empirical contact networks. bioRxiv, e169573. https://doi.org/10.1101/169573.

Sax, L.J., Gilmartin, S.K., Bryant, A.N., 2003. Assessing response rates and nonresponse bias in web and paper surveys. Res. High. Educ 44, 409–432. https://doi.org/10.1023/A:1024232915870.

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinforma. 27, 592–593. https://doi.org/10.1093/bioinformatics/btq706.

Schürch, A.C., Arredondo-Alonso, S., Willems, R.J.L., Goering, R.V., 2018. Whole-genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. Clin. Microbiol. Infect. 24, 350–354. https://doi.org/10.1016/j.cmi.2017.12.016.

Sears, A., Baker, M.G., Wilson, N., Marshall, J., Muellner, P., Campbell, D.M., Lake, R.J., French, N.P., 2011. Marked campylobacteriosis decline after interventions aimed at poultry, New Zealand. Emerg. Infect. Dis. 17, 1007–1015. https://doi.org/10.3201/eid1706.101272.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma. 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Shirk, A.J., Landguth, E.L., Cushman, S.A., 2017. A comparison of individual-based genetic distance metrics for landscape genetics. Mol. Ecol. 17, 1308–1317. https://doi.org/10.1111/1755-0998.12684.

Shizuka, D., Farine, D.R., 2016. Measuring the robustness of network community structure using assortativity. Anim. Behav. 112, 237–246. https://doi.org/10.1016/j.anbehav.2015.12.007.

Silk, M.J., Drewe, J.A., Delahay, R.J., Weber, N., Steward, L.C., Wilson-Aggarwal, J., Boots, M., Hodgson, D.J., Croft, D.P., McDonald, R.A., 2018. Quantifying direct and indirect contacts for the potential transmission of infection between species using a multilayer contact network. Behaviour 155, 731–757. https://doi.org/10.1163/1568539X-00003493.

Slader, J., Domingue, G., Jørgensen, F., McAlpine, K., Owen, R.J., Bolton, F.J., Humphrey, T.J., 2002. Impact of transport crate reuse and of catching and processing on *Campylobacter* and *Salmonella* contamination of broiler chickens. Appl. Environ. Microbiol. 68, 713–719. https://doi.org/10.1128/AEM.68.2.713-719.2002.

Sokal, R.R., Rohlf, F.J., 1962. The comparison of dendrograms by objective methods. Taxon 11, 33–40. https://doi.org/10.2307/1217208.

Somerfield, P., Clarke, K., 1995. Taxonomic levels, in marine community studies, revisited. Mar. Ecol. Prog. Ser. 127, 113–119. https://doi.org/10.3354/meps127113.

Stadler, T., Bonhoeffer, S., 2013. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. Philos. Trans. R. Soc. B Biol. Sci. 368, e20120198 https://doi.org/10.1098/rstb.2012.0198.

van Wagenberg, C.P.A., van Horne, P.L.M., Sommer, H.M., Nauta, M.J., 2016. Cost-effectiveness of *Campylobacter* interventions on broiler farms in six European countries. Microb. Risk Analy 2–3, 53–62. https://doi.org/10.1016/j.mran.2016.05.003.

Webster, J., Borlase, A., Rudge, J., 2017. Who acquires infection from whom and how? Disentangling multi-host and multi-mode transmission dynamics in the 'elimination' era. Philos. Trans. R. Soc. B Biol. Sci 372, e20160091. https://doi.org/10.1098/rstb.2016.0091.

Williamson, D.A., Dyet, K., Heffernan, H., 2015. Antimicrobial Resistance in Human Isolates of *Campylobacter jejuni*. The Institute of Environmental Science and Research Limited, Porirua, New Zealand.

Ypma, R.J., Bataille, A.M., Stegeman, A., Koch, G., Wallinga, J., van Ballegooijen, W.M., 2012. Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data. Proc Biol Sci 279, 444–450. https://doi.org/10.1098/rspb.2011.0913.

Zhang, J., Halkilahti, J., Hänninen, M.-L., Rossi, M., 2015. Refinement of whole-genome multilocus sequence typing analysis by addressing gene paralogy. J. Clin. Microbiol. 53, 1765–1767. https://doi.org/10.1128/JCM.00051-15.

Zhang, T., Lees, M., Kwoh, C.K., Fu, X., Lee, G.K.K., Goh, R.S.M., 2012. A contact-network-based simulation model for evaluating interventions under "what-if' scenarios in epidemics. Proceedings of the 2012 Winter Simulation Conference (WSC) Pg. 1–12. Berlin, Germany. https://doi.org/10.1109/WSC.2012.6465056.