Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

This is an author produced version of a paper published in
Preventive Veterinary Medicine.
This paper has been peer-reviewed but may not include the final publisher proof-corrections or pagination.

Citation for the published paper:
Lena-Mari Tamminen, Robert Söderlund, David A. Wilkinson, Maria Torsein, Erik Eriksson, Mikhail Churakov, Johan Dicksved, Linda J. Keeling, Ulf Emanuelson. (2019) Risk factors and dynamics of verotoxigenic Escherichia coli O157:H7 on cattle farms: An observational study combining information from questionnaires, spatial data and molecular analyses. *Preventive Veterinary Medicine*. Volume: 170, Number: 1 October.
https://doi.org/10.1016/j.prevetmed.2019.104726

Access to the published version may require journal subscription.
Published with permission from: Elsevier.

Standard set statement from the publisher:
© Elsevier, 2019 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

Epsilon Open Archive http://epsilon.slu.se

Risk factors and dynamics of verotoxigenic Escherichia coli O157:H7 on cattle farms: An observational study combining information from questionnaires, spatial data and molecular analyses

Lena-Mari Tamminen[a,*], Robert Söderlund[b], David A.Wilkinson[c,d,˙], Maria Torsein[e,ˀ], Erik Eriksson[b], Mikhail Churakov[f], Johan Dicksved[f], Linda J. Keeling[g], Ulf Emanuelson[a]

[a] Department of Clinical Sciences, Swedish University of Agricultural Sciences, Box 7054, SE-75007, Uppsala, Sweden
[b] National Veterinary Institute (SVA), SE-75189, Uppsala, Sweden
[c] Molecular Epidemiology and Public Health Laboratory (mEpilab), Infectious Disease Research Centre, Hopkirk Research Institute, Massey University, Private Bag 11-222, Palmerston North, New Zealand
[d] New Zealand Food Safety Science and Research Centre, Massey University, Palmerston North, New Zealand.
[e] Department of Animal Environment and Health, Swedish University of Agricultural Sciences, Box 234, SE-53223, Skara, Sweden
[f]Department of Animal Nutrition and Management, Swedish University of Agricultural Sciences, Box 7024, SE-75007, Uppsala, Sweden
[g] Department of Animal Environment and Health, Swedish University of Agricultural Sciences, Box 7068, SE-75007, Uppsala, Sweden

[*] Corresponding author: Lena-Mari Tamminen, Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden, SE-75007
Phone: +4618671428, E-mail: lena.mari.tamminen@slu.se

[˙] Present address: University of La Reunion, 97490, Reunion, France

[ˀ] Present address: Gökhem Bältaregården 2, 521 92 Falköping, Sweden

Abstract

The increasing number of human cases infected with a highly virulent type of verotoxigenic *Escherichia coli* (VTEC) O157:H7 in Sweden is the result of domestic transmission originating in regional clusters of infected cattle farms. To control the spread of the bacteria a comprehensive picture of infection dynamics, routes of transmission between farms and risk factors for persistence is urgently needed. The aim of the study was to investigate different aspects of the epidemiology of VTEC O157:H7 on the Swedish island of Öland by combining information from environmental sampling of VTEC O157:H7 from 80 farms with information from farmer questionnaires, spatial and molecular analyses. The farms were sampled in the spring and fall of 2014 and on four of them additional samples were collected during summer and winter. The results show a high prevalence of VTEC O157:H7 and a high proportion of strains belonging to the virulent clade 8. Farms that became infected between samplings were all located in an area with high cattle density. The most important risk factors identified are generally associated with biosecurity and indicate that visitors travelling between farms may be important for transmission. In addition, whole genome sequencing of a subset of isolates from the four farms where additional sampling was performed revealed ongoing local transmission that cannot be observed with a lower resolution typing method. Our observations also show that VTEC O157:H7 may persist in the farm environment for extended periods of time, suggesting that specific on-farm measures to reduce environmental prevalence and spread between groups of animals may be required in these cases.

Keywords: VTEC O157, EHEC, epidemiology, clade 8, transmission

Introduction

Verotoxin-producing *Escherichia coli* serotype O157:H7 (VTEC O157:H7) is a zoonotic pathogen causing public health concerns across the world (Majowicz et al., 2014). It belongs to the group enterohemoragic *E. coli* (EHEC) that, in addition to severe gastrointestinal disease, can cause serious complications such as hemolytic uremic syndrome (HUS) in children and the elderly (Karmali, 2004). In Sweden, these are often associated with a specific group of VTEC O157:H7 called clade 8, a strain known to cause more serious disease, with proportionally higher numbers of hospitalizations and cases of HUS (Manning et al., 2008). A recent international comparison suggested that all the included clade 8 isolates from Sweden were derived from a single introduction from North America around 1990 (Franz et al., 2018) and the important connection between infected cattle farms and human cases is well established (Eriksson et al., 2011; Söderlund et al., 2014). Fortunately, the overall prevalence of clade 8 in Sweden is relatively low but local clustering of VTEC O157:H7 and clade 8 can lead to high local prevalence within the cattle population (Widgren et al., 2015) and thus lead to an increased hazard for the surrounding human population. Historically, the presence of VTEC O157:H7 has been a problem in south western parts of Sweden, especially the county of Halland (Eriksson et al., 2005), but after 2011 this established pattern began to change. In 2013 the incidence of human cases in the eastern county of Kalmar, previously a low incidence area, had the highest incidence in the country (7.3 cases per 100 000 inhabitants) (Folkhälsomyndigheten, 2019). In addition, national surveillance identified clade 8 in slaughtered cattle from the island of Öland (part of Kalmar county) for the first time in 2014 (Unpublished data, National Veterinary Institute, Uppsala, Sweden).

It has been previously suggested that reducing on farm prevalence of VTEC O157:H7 is the most efficient way to control the disease in humans (Bell, 2002; LeJeune

and Wetzel, 2007). The importance of reducing transmission from cattle is also emphasised in the Swedish strategy for reducing human cases of VTEC O157:H7 which highlights the need for actions throughout the food chain, including control measures on infected farms (Socialstyrelsen, 2014). However, identification of infected animals and farms is difficult as cattle carry VTEC O157:H7 in their intestine without showing clinical symptoms (Chase-Topping et al., 2008). Also, farms have been shown to be transiently infected, often clearing the bacteria after 3-4 months (Widgren et al., 2015; Zhang et al., 2010), although there are examples where farms have been observed to be positive for longer periods of time (Fremaux et al., 2006; Herbert et al., 2014; Lahti et al., 2003; LeJeune et al., 2004; Tamminen et al., 2018). This variation in farm-persistence means that the need and usefulness of control measures differ between farms as some farms may not require any interventions to clear the bacteria. On farm persistence will also, in combination with transmission rate, influence local prevalence. Previous studies have indicated that long-distance transmission typically occurs through cattle trade but that ongoing local spread between farms is important for prevalence (Herbert et al., 2014; Widgren et al., 2018). A simulation study indicated that multiple transmission routes exist (Zhang et al., 2010). For example wildlife, like birds and flies (Ahmad et al., 2007; Cernicchiaro et al., 2012; Swirski et al., 2014; Synge et al., 2003), human activities, like purchase and movement of animals (Widgren et al., 2015) and taking animals to shows (Cernicchiaro et al., 2009), may all play a role. Still, the underlying drivers of local spread and persistence are poorly understood.

One of the reasons for this knowledge gap is that the microbiological analyses performed in the majority of earlier studies on VTEC O157:H7 have been limited to isolating the bacteria through culture-based methods, like direct plating on specific agars or immunomagnetic separation, followed by confirming the presence of virulence genes

via polymerase chain reaction (PCR). The availability of typing methods, like multi-locus variable number tandem repeat analysis (MLVA) and pulse field gel electrophoresis (PFGE), has led to new insights, showing that different types of VTEC O157:H7 may behave differently. Some are more or less likely to cause disease in humans, while certain variants are more likely to persist in the cattle population (Herbert et al., 2014; Söderlund et al., 2014). With whole genome sequencing (WGS) and the use of single nucleotide polymorphisms (SNPs) to characterise isolates, even more information is becoming available for the study of VTEC O157:H7 transmission and dynamics. Some recent studies have used these techniques to study host associations and international transmission events (Franz et al., 2018; Strachan et al., 2015).

The purpose of this observational study was to investigate the epidemiology of VTEC O157:H7 in cattle herds on the Swedish island of Öland. The main objective was to study prevalence on the island as well as the dynamics of clearance, persistence and new infection of VTEC O157:H7 between spring and fall 2014 in order to evaluate the need for, and appropriate structure of, control measures in the area. To provide further guidance on most efficient control measures, risk factors associated with the presence, infection and reinfection of VTEC O157:H7 were analysed and modern molecular techniques were used to explore persistence and local transmission of VTEC O157:H7.

Materials and methods

This study was part of a national surveillance effort financed by the Swedish Board of Agriculture. Environmental samples were collected from 80 cattle farms on Öland on two occasions, once in April and once in October 2014 (Figure 1). Öland is an island located on the east coast of Sweden and is 137 km long and up to 16 km wide. Sample size calculation using http://epitools.ausvet.com.au indicated that this could estimate true prevalence on Öland with 5 % precision and 90 % confidence level when the assumed

prevalence was 10 % (national average) (Humphry et al., 2004). Sampling was performed by the local livestock association who also recruited farmers across the island. The local livestock association staff phoned farmers before scheduled routine visits for e.g. dehorning or insemination and asked farmers to participate in the study. As motivation, farmers were offered a small financial compensation. The local livestock association continued recruiting until 80 farms across the island had been enrolled in the study. Two farmers declined to participate over the phone which means that 82 farmers in total were contacted.

On-farm sampling

Two environmental sampling techniques were used on all farms, as previously described by Widgren et al. (2013). Overshoe sampling (OS) was performed by fitting gauze soaked with phosphate buffered saline (PBS) over plastic overshoes and walking around in the pens. The gauze was rotated during sampling so the whole gauze was used and then each gauze was removed and the pair placed in a plastic bag. Collectors placed a new pair of plastic covers over their boots before each sampling to ensure no cross-contamination. While walking around the pen the person also collected a pooled fecal sample (PS) consisting of fresh faeces collected from 15-20 pick points on the floor or from the deep litter bedding. Approximately 1 cm^3 of feces was picked from each point and placed in a 100 ml plastic container. Samples were collected from two groups of animals; calves (from weaning up to six months of age) and young stock (approximately 6 -12 months of age). One PS and one OS was collected from each group, meaning that a total of two OS and two PS samples were collected per sampling occasion from each farm. Sampling was performed by personnel of the local livestock association. Samples were collected in the beginning of the working week (Monday-Wednesday) and shipped to the National

149 Veterinary Institute by standard post. Sample analysis started the day after sampling.

150 ***Analysis of VTEC O157:H7***

151 *Microbiological analysis*

152 For each sample (the pair of gauzes or 25 mg of feces), 225 ml of modified tryptic soy
153 broth (mTSB) (Oxoid) (supplemented with 20 mg/l of novobiocin) was added and mixed
154 with the sample in a stomacher. Samples were then pre-enriched at 41.5 °C ± 0.5 °C for
155 18–24 h. After pre-enrichment, immunomagnetic separation (IMS) was performed with
156 paramagnetic beads (Dynabeads anti-E. coli O157; Dynal) according to the
157 manufacturer´s instructions. IMS was performed either directly after 18–24 h of
158 incubation or after the pre-enriched broth had been stored in cold storage for 24–48 h at
159 4 °C. After IMS, the beads were spread out on sorbitol McConkey agar (Oxoid)
160 supplemented with 0.05 mg/l cefixime and 2.5 mg/l of potassium tellurite (CT-SMAC;
161 Dynal). After incubation at 37 °C for 18–24 h, the agar plates were screened for suspected
162 sorbitol negative colonies of E. coli O157. Up to 5 suspected colonies were picked for
163 agglutination with a latex kit (DR 622; Oxoid) and colonies which yielded a positive
164 agglutination were further tested biochemically using the API 20 E system (bioMérieux).
165 If positive for VTEC O157:H7, PCR according to Paton & Paton (1998) and Gannon and
166 others (1997) was performed to identify the presence of genes coding for verotoxin 1 and
167 2 (*vtx1* and *vtx2*) and intimin (*eaeA*). Belonging to clade 8 was determined by real-time
168 PCR as described by Söderlund et al. (2014).

169 *MLVA-typing*

170 Multi-locus variable number tandem repeat analysis typing (MLVA) analysis was
171 performed on all strains of VTEC O157:H7 as previously described (Söderlund et al.,

2014).

Whole genome sequencing

Four farms included in the study were part of a parallel research project, and from these additional samples were available. In addition to the spring and fall sampling previously described, these farms were visited three times during summer (in July, June and September) as well as once in December, as presented in Figure 2. Sampling of barn and pasture environments was performed around groups of calves and young stock by combining OS and PS as described above. Manure samples were collected from the manure pit. Samples were then enriched and treated as described above. Flies were caught in traps on pasture. At arrival to the National Veterinary institute they were placed in a stomacher bag and homogenized before enrichment as previously described. Whole genome sequencing was performed on 30 isolates of clade 8 recovered from these farms throughout the year (collection month presented in Figure 2). DNA was extracted using a DNeasy Blood & Tissue kit automated on a BioRobot system (Qiagen). Sequencing libraries were prepared using the Nextera XT kit and sequenced on an Illumina MiSeq system with 2 x 250 bp paired-end reads.

Processing, assembly and analysis of raw reads was performed using the Nullarbor pipeline in "accurate" mode (Seemann et al., 2017) using *E. coli* O157:H7 str. Sakai (NC_002695.2) as the reference. Recombination of the core genome was assessed in Gubbins (Croucher et al., 2015) and a phylogenetic tree based on core genome SNP-distance was generated in RAxML based on Maximum likelihood (model GTRGamma with 1000 bootstraps) (Stamatakis, 2014). The phylogenetic relationship was illustrated using Interactive Tree of Life (iTOL) software (Letunic and Bork, 2016)

Questionnaire

Information about the farms was collected through a questionnaire sent by post to farmers in October 2014 (around the time of the fall sampling), along with the documents necessary to receive compensation for participating in the study. Farmers that had not responded by the end of November 2014 were reminded by phone or email. The questionnaire (available in Swedish from the corresponding author) included questions about general herd characteristics, contacts with other farms, hygiene routines and specific events during the time between sample collections. The majority of the questions were closed but included room for additional comments. All questions about contacts included an additional row for stating which farms the contact was concerning. The questionnaire was developed in cooperation with a representative from Farm and Animal Health Services and reviewed by a veterinarian specialized in cattle medicine and herd health.

Data management

Data were entered in Microsoft Excel and exported to R Statistical software (R Core Team, 2018) where statistical analysis was performed. Coordinates representing the farm building of all cattle farms on the island were retrieved from the national registry for animal production sites at the Swedish Board of Agriculture through the national database "Geodata" (https://www.geodata.se). For each farm the number of neighbours was calculated in QGIS by summing the number of other cattle farms located within a 5 km radius. The radius was selected based on a previous Swedish study which found that infected farms within this distance increases risk of becoming infected (Widgren et al., 2015). Variables from the questionnaire were categorised as either "herd characteristics", i.e. variables that would stay the same over time, or "between sampling events", i.e.

219 specific events that had occurred between the spring and fall samplings (see
220 Supplementary material, Table 1). The 80 study farms were organized into four groups
221 based on their infection status: NN (negative at both samplings), NP (negative at the first
222 sampling and positive at the second), PN (positive at the first sampling and negative at
223 the second) and PP (positive at both samplings).

224 ***Statistical analysis***

225 To assess spatial clustering of positive herds for each of the two sampling occasions, we
226 used Cuzick-Edwards' kNN (k nearest neighbours) and Ripley's K function tests (Cuzick
227 and Edwards, 1990; Ripley, 1981). Both of these tests account for the underlying
228 population at risk and determine if the observed distribution of positive farms is
229 significantly different from a randomly simulated one. The two methods differ in the
230 choice of statistics that depend on: the number of neighbours and Euclidian distance,
231 respectively. The analysis was performed using "smacpod" R package (French,
232 2018) and random distributions of cases were simulated 1000 times.

233 The associations between general herd characteristic and presence of VTEC
234 O157:H7 at any sampling occasion were analysed using a generalized linear mixed model
235 fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature). Herd was included as
236 a random variable to account for the two sampling occasions and the model run with 25
237 iterations using the package lme4 (Bates et al., 2015). Multicollinearity among the
238 variables was checked using the variance inflation factor (VIF) in the car package (Fox
239 and Weisberg, 2011). A backwards model selection was performed using Akaike
240 Information Criterion (AIC). Non-significant ($p > 0.1$) variables were excluded one at a
241 time and the change in AIC evaluated. If AIC decreased the variable was left out. If AIC
242 remained the same the variable was kept. Confounding was controlled by reintroducing

each excluded variable to the final model and evaluating the change in AIC and in estimates of the other variables. The overall goodness of fit was assessed by Hosmer-Lemeshow test using the package "ResourceSelection" and splitting the data into 10 groups (Lele et al., 2019; Lemeshow and Hosmer, 1982). Area under the curve (AUC) was calculated using the package "pROC" (Turck et al., 2011). Residual errors of the model were analysed using Moran's I in the package "spdep" to assess spatial independence. In addition to analysing association between residual errors and 10 nearest neighbours, we also calculated bisquared weights based on Euclidean distance for the 10 nearest neighbours of each farm and analysed the association.

All variables, including "between sampling events", were used to study risk factors for persistence of VTEC O157:H7 on a farm by comparing farms that cleared themselves of the bacteria between the spring and fall sampling with farms that remained positive in fall. Similarly farms negative in spring and which remained negative were compared to farms where the bacteria was introduced over summer to study risk factors for introduction. Farms that were positive in spring and farms where infection was introduced over summer were relatively few and due to the small sample size, analysis was limited to Wilcoxon rank test, using the "coin" package (Hothorn et al., 2006), for the quantitative variables and Fisher's Exact test for the qualitative variables using the package "hypergea" (Boenn, 2018). In addition a comparison of "between sampling events" of farms positive in fall and farms negative in fall was performed as described above.

A matrix of genetic distance, as extracted from the Maximum Likelihood phylogenetic tree generated in RAxML, between the 30 whole genome sequenced isolates from the four farms was created. From this pairwise distances between all isolates was extracted rendering 465 observations. The association between genetic distance between

each pair of isolates and geographical distance between the collection points (i.e. distance between farms or 0 km for isolates from the same farm) was analysed using a linear model in the R base package. Difference in days between collection was calculated between all isolates and included as a fixed effect in the model to account for strain development over time. In addition to geographical distance a model comparing driving distance (retrieved through Google maps) and genetic distance was also fitted. Normality of residuals and signs of heteroscedasticity were graphically assessed through diagnostic plots.

Results

Presence of VTEC O157:H7 on the 80 farms

Results of spring and fall samplings including results from MLVA typing are presented in Figure 1. In spring, VTEC O157:H7 was found on 21 farms; all isolates except two belonged to clade 8. In fall, the number of positive farms was again 21 and all isolates belonged to clade 8. Thus, no seasonal difference in prevalence between spring and fall was observed. Three of the 80 farms declined to participate in the follow-up sampling and of these 1 had been positive for clade 8 in spring. Of the farms negative in spring 44 were negative at both samplings (NN) and 13 became positive during summer (NP). Of the farms positive in spring, eight were positive also on the second sampling (PP) and 12 became negative (PN). As seen in Figure 1, there was strong similarity in MLVA profiles between farms indicating a recent introduction and rapid spread of the bacteria between farms. In total five clusters of clade 8 were found, although the differences between them were very small. The dominating cluster (150-A1) was found all over the island. Multiple MLVA types were identified on four farms on the same sampling occasion (2 farms with 2 MLVA types and 2 farms with 3 MLVA-types). The two isolates that did not belong to clade 8 were found in the south and north of the island. We detected strong spatial

clustering of positive farms in the fall, while in the spring their distribution was random (Figure 3).

Risk factors for presence of VTEC O157:H7

Completed questionnaires were received from 55 of the 80 farms. Thirteen farms were positive for the bacteria in spring and 14 farms were positive in the fall sampling. Out of these 14 farms, 6 had been positive in the spring sampling. All responses to the questionnaire can be found in the supplementary material (Table S1). Between-farm contacts stated by farmers are presented in Figure 4.

Presence of VTEC O157:H7 at any sampling occasion

After model selection, the final model of herd characteristics associated with the presence of VTEC O157:H7 contained 4 variables presented in Table 1. We tested for spatial autocorrelation of the residuals using Moran's I, which was not significant, indicating that the assumption of independence was fulfilled. Being a large farm with many animals, having several neighbours and using reproductive services (meaning that the farmer continuously used artificial insemination services provided by the local livestock association) was significantly associated with the presence of VTEC O157:H7 on farms. In addition, having a cat on the farm was retained in the model as removing it increased AIC.

Table 1. Results from the final logistic regression model[a] for risk factors associated with presence of VTEC O157:H7 on a farm at any sampling occasion with farm ID included as a random effect[b].

	OR	Estimate	(SE)	p-value
Cat (yes)	3.0	1.092	(0.650)	$p < 0.10$
Use reproductive services (yes)	4.4	1.487	(0.735)	$p < 0.05$
Number of cattle	2	0.700	(0.264)	$p < 0.01$
Neighbours within 5 km	1.15	0.148	(0.066)	$p < 0.05$

[a]Hosmer-Lemeshow goodness of fit test was 8.79 with 8 d.f and $p = 0.36$, AUC was 87%.
[b]Variance explained by farm was 0.7 (with standard deviation 0.83)

Clearance, introduction and persistence of VTEC O157:H7

A selection of variables (with p-values < 0.15) from the comparison of farms that were negative for VTEC O157:H7 on both sampling occasions (NN) and farms where infection was introduced during summer (NP) and the comparison between farms positive on both occasions (PP) to farms that cleared infection over summer (NP) are presented in Table 2. Similarly a selection of variables (with p-values < 0.15) related to "in between sampling events" and comparison of farm status in the fall sampling are presented in Table 3.

Table 2. Risk factors associated with new infection or clearance of VTEC O157:H7. For quantitative variables arithmetic mean and quartiles (25^{th} : 75^{th}) are presented.
(*NN= negative on both sampling occasions, NP=negative in spring, positive fall, PN=positive spring, negative fall, PP = positive on both occasions, [a] indicates Wilcoxon-Mann-Whitney test, [b] indicates Fisher Exact test)*

Risk factor		**NN** *n=33*	**NP** *n=9*	**OR (95 % CI)**	**p-value**
Number of cattle	(Quant)	245 (150:301)	296 (247:350)		0.14[a]
Number of neighbours (within 5 km)	(Quant)	18 (15:21)	23 (24:25)		<0.05 [a]
Horse	Yes	3	3	5	0.10 [b]
	No	30	6	(0.8-31.3)	
Purchased animals	Yes	3	4	7.4	<0.05 [b]
	No	30	5	(1.0-68.0)	
Any known contact with positive farm	Yes	7	5	4.4	0.09 [b]
	No	26	4	(0.7-29.3)	
		PN *n=7*	**PP** *n=6*		
Number of cattle	(Quant)	194 (138:230)	435 (313:553)		<0.01 [a]
Number of neighbours (within 5 km)	(Quant)	25 (23:26)	17 (16:18)		<0.05 [a]
Type of farm	Milk	5	1		<0.05 [b]
	Combination	2	5		
Any known contact with positive farm	Yes	0	4	27	<0.05 [b]
	No	7	2	(1.0-698.8)	
Visits to other farms the passed 5 months	Yes	0	3	15	0.07 [b]
	No	7	3	(0.6-376.7)	

Table 3. Comparison between farms positive for VTEC O157:H7 in the fall sampling and farms negative in the fall sampling using Fisher Exact test).
(*FN=negative in fall sampling, FP=positive in fall sampling,*

Risk factor	**FN** *n=40*	**FP** *n=15*	**OR (95 % CI)**	**p-value**

Purchased animals	Yes	3	5	4.4	0.10
	No	36	10	(0.8-26.5)	
Any known contact	Yes	7	9	6.8	<0.01
with positive farm	No	33	6	(1.6-32.3)	
Share Agricultural	Yes	20	12	3.3	0.07
Machines	No	20	3	(0.9-24.8)	

Whole genome sequencing

Average SNP distance between isolates from the 4 farms are presented in Figure 2 and show that isolates from farm 1, 2 and 4 generally had shorter SNP distance between each other compared to isolates from Farm 3 that were more distant. Distance between isolates within the same farm varied and was smallest on Farm 4 where average SNP distance was 64. On Farm 1 and 2 it was 109 and 108 respectively whereas the isolates from Farm 3 had an average distance of 216. This pattern is also seen in the phylogenetic tree of the core genome (Fig. 5). On Farm 3, highly similar isolates of VTEC O157:H7 were found in the September, October and December sampling. These were isolated from different sources, including environmental sampling of pasture, from flies on the pasture as well as the barn and the manure pit. On this farm there was also another group of similar isolates collected from the barn and on pasture in the May, October and December samplings that were more closely related to the isolates from the other farms. Isolates from Farm 1, 2 and 4 showed high genetic relatedness but generally clustered within farm and sampling date. Isolates did not have a clear environmental niche and closely related isolates were retrieved from samples collected from different sources.

The association between genetic distance and distance between farms (Fig. 6) was highly significant ($p < 0.001$), but as R^2 was only 0.33 it is clear that the distance only explains part of the variation observed. The model improved slightly when using driving distance instead of geographical distance (adjusted R^2 increased from 0.33 to 0.37). However, the association was attributable to the dissimilarity of the isolates from Farm 3

which was located furthest away from the other farms and the significant association disappeared when isolates from this farm were removed from the analysis.

Discussion

Presence of VTEC O157:H7 and clade 8 on Öland

In this study, VTEC O157:H7 was detected on 26 % and 27 % of the sampled farms during the spring and fall, respectively. This is higher than in previous national studies that reported 8.9 % (Eriksson et al., 2005) and 6.1 – 13.6 % (Widgren et al., 2015). This study used the same sampling scheme as the study by Widgren et al. (2015), a method that has been shown to reliably identify herds with animals shedding VTEC O157:H7 (Widgren et al., 2013). This could indicate that this region differs from other regions included in the earlier studies. However, previous studies have shown that prevalence varies between years and the results may represent an unusual year and not regional differences (Widgren et al., 2015). It should also be noted that the farms included in this study were not selected at random, which might have led to selection bias due to convenience sampling. Thus, the prevalence observed should be interpreted with caution.

In addition to the high prevalence, the proportion of positive farms where clade 8 was present was also very high (95 %) compared to previous national studies in other regions where the observed proportion has varied between 0 – 55 % (Söderlund et al., 2014; Widgren et al., 2015). Due to the association between human cases of VTEC O157:H7 and cattle density (Frank et al., 2008; Kistemann et al., 2004) the high presence of this virulent strain should be considered an important threat to public health. This is particularly important on Öland as it is a major food-producing region as well as a popular area for recreational activities in the summer.

373 Farmers on Öland differ from the majority of Swedish farmers as they often own
374 multiple small areas of land spread across the island instead of one large area centred
375 around a barn. This means that animals are frequently transported around the island to
376 different pastures during summer and pastures of different farms are often located close
377 to each other with only simple fences separating the animals. While this would lead us to
378 expect bidirectional contact between farms, a large number of one-way contacts are
379 present in the network based on answers from the questionnaire (Fig 4). It is likely that
380 multiple contacts occur between farms, especially between neighbouring pastures, and
381 this information is perhaps not best-captured by our questionnaire as it requires farmers
382 to provide excessively thorough catalogues of land-ownership adjacencies.

383 ***Risk factors and transmission of VTEC O157:H7***

384 Previous international studies have shown higher levels of VTEC O157:H7 in summer
385 and fall (Barkocy-Gallagher et al., 2003; Schouten et al., 2005 and in Sweden a study
386 found that the probability of detecting VTEC O157:H7 on dairy farms increases in the
387 third and fourth quarter of the year (Widgren et al., 2015). In this study no clear
388 differences in proportion of positive farms were observed between the spring and fall
389 periods. However, analysis of the spatial clustering of positive herds (Fig 3) revealed a
390 strong clustering in the fall but not in the spring sampling. This might suggest that local
391 transmission is more intensive during summer months compared to winter, when animals
392 are generally kept inside. For example cattle could be encountering new strains on pasture
393 and bringing them home, as observed through the whole genome sequencing of isolates
394 from Farm 3.

395 The analyses of the responses from the questionnaires support previous findings
396 that larger farm size and the purchase of animals increase the risk of having VTEC

O157:H7 on a farm (Herbert et al., 2014; Widgren et al., 2015). The increased risk associated with the use of reproductive services may be linked to receiving visitors that travel frequently between farms in the area. Implementation of biosecurity measures for these local movements may be an important target for controlling VTEC O157:H7. However, considering other routes, like birds, flies and purchase of animals (Ahmad et al., 2007; Cernicchiaro et al., 2009; Schouten et al., 2004; Wilson et al., 1993) may also be necessary. In addition, it cannot be excluded that the association reflects an unmeasured effect related to difference between farmers that choose to use reproductive services and those that carry out the task themselves, as many farmers in Sweden choose to do.

The association between genetic distance and geographic distance observed between the sequenced isolates also indicate that local transmission through movement of humans and vehicles may be of potential importance. However, as this analysis included a limited selection of isolates from a small number of farms, and that the geographically distant Farm 3 heavily influenced the association, results from this analysis should be interpreted carefully. Still, it is interesting that the model improved slightly when road distance was used compared to geographical distance between the farms and this should be further explored in future studies with genetic distances available for correlation with a larger number of pairwise geographical distances. It is also obvious from the presented data, that even best-resolution typing techniques (WGS) have limitations in a region with highly related genotypes. In these settings genetic diversity resulting from separate sources of infection can be indistinguishable from diversity that has emerged within an individual farm. When it comes to tracing the source of isolates back to farms, e.g. from a human case, in an outbreak situation, it is also clear that an

isolate cannot be reliably attributed to a single farm simply based on sequencing results. Therefore, source attribution will have to rely on epidemiological evidence.

It is also interesting that the presence of a cat on the farm was weakly associated with presence of VTEC O157:H7. It has previously been shown that cats and cattle from the same farm can carry comparable types of VTEC (Joris et al., 2013). Hence, cats might serve as a disease vector as they move around freely within, and potentially between nearby farms. The free movement of such animals between farms makes evaluation of the associated risk indirect as cats could be a hazard to both farms that report keeping cats and farms that report not keeping cats, potentially leading to underestimation of the risk observed in this study. Thus, studies directly looking at VTEC carriage in cats would be required to elucidate any role they play in the dissemination of these agents.

Persistence or reinfection?

Previous studies using MLVA and PFGE have identified that related strains may persist on farms and hypothesised that the farm was the reservoir of the pathogen in these cases (Joris et al., 2013; Lahti et al., 2003; LeJeune et al., 2004; Sanderson et al., 2006). In this study, MLVA also indicated persistence between sampling occasions but when looking in more detail using whole genome sequencing there are examples of several strains that appear to jump between three of the farms, indicating ongoing transmission or the continuous presence of multiple strains on the same farms. This insight into the transmission dynamics of VTEC O157:H7 would not have been possible using other typing techniques. The close genetic relationship observed between the isolates in this study thus highlights the need for maximum-resolution typing strategies to differentiate between closely related strains of VTEC O157:H7 (and other organisms). This is particularly true in relatively closed systems such as the studied farms where the majority

445 of relevant circulating strains are homogenous and likely to have derived from a recent
446 common ancestor.

447 The only farm where isolates were consistently related over time was Farm 3.
448 Although the limited number of farms with available sequences does not allow firm
449 conclusions to be drawn about persistence and re-infection risks, the two observed
450 patterns generate new hypotheses when considering the risk factors identified from our
451 analysis of farming practices. We may be identifying a mix of risk factors associated with
452 new infection as well as persistence. For example, the underlying reason behind the risk
453 associated with increasing number of animals may be related to having enough animals
454 on farm to get a circulation of the bacteria. This may explain the highly significant
455 difference in size between the farms that cleared infection during summer and those that
456 remained positive. However, large farms in this particular area of Sweden may also have
457 their animals spread out on pastures on several parts of the island and thereby have a
458 larger contact network. It has also been shown that larger farms have increased number
459 of professional visits compared to smaller farms (Nöremark et al., 2013). Thus, the risk
460 for introduction of new strains is likely higher on larger farms.

461 Untangling these relationships will require additional studies including WGS
462 techniques in the future. In addition to understanding to which extent persistence occurs,
463 the potential role of persistently infected farms in sustaining bacterial circulation in an
464 area may be important to consider. Identifying and understanding the drivers behind
465 persistence on farms may also be of particular importance because of the association
466 between persisting strains and clinical disease in humans (Herbert et al., 2014). It is also
467 important to recognize that both patterns exist when considering control of the pathogen,
468 as farms with persisting isolates likely require other control measures than farms where
469 new strains are frequently introduced.

Implementation of interventions and on-farm measures on Öland

As a response to the wide spread of clade 8 on the island authorities (including national agencies as well as the local municipality) jointly generated information campaigns. One was targeted to the public including information about hand hygiene when in contact with cattle. In addition, information notices were put up on entrances to cattle pastures around the island. Farmers were informed about the bacteria and how to prevent transmission to humans visiting their farms. In addition farms where the bacteria had been identified were offered repeated sampling during 2015 and, if they remained positive, advice on how to reduce the infectious pressure on their farms were provided. These recommendations were mainly targeted on minimizing contact between animal groups and other measures previously described in Tamminen et al. (2018), but additional advice based on the results from this study is now being developed. For example control of flies is now being included. The frequent transmission between farms has also shifted the national public health strategy from focusing on individual farms to considering high risk areas and highlighted the importance of biosecurity measures within these areas.

Conclusion

This study reports an unusually high prevalence of VTEC O157:H7 and high proportion of clade 8 on the studied farms on Öland island, which is a significant public health concern. Presence of VTEC O157:H7 was positively associated with the previously known risk factors: size of farm and number of close neighbours. In addition, risk factors related to biosecurity, such as using reproductive services and having a cat on the farm, were also identified as important. All the collected isolates were genetically similar, reinforcing the need for using whole genome sequencing techniques to study local transmission dynamics of VTEC O157:H7.

Acknowledgements

The authors would like to thank the Swedish board of Agriculture for providing funding of study and the farmers that agreed to participate. We would also like to thank the Farm and Animal Health (Gård och Djurhälsan) for organising sampling and inviting the authors to take part in the study. The authors would also like to thank the EHEC and molecular diagnostics laboratories at the Swedish veterinary institute (SVA) for excellent technical assistance.

Funding: This work was supported by the Swedish board of Agriculture, through Farm and Animal Health Services. David A. Wilkinson was funded by the New Zealand Food Safety Science and Research Centre.

506 **References**

507 Ahmad, A., Nagaraja, T.G., Zurek, L., 2007. Transmission of Escherichia coli O157:H7
508 to cattle by house flies. Prev. Vet. Med. 80, 74–81.
509 https://doi.org/10.1016/j.prevetmed.2007.01.006

510 Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects
511 Models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01

512 Bell, C., 2002. Approach to the control of entero-haemorrhagic Escherichia coli
513 (EHEC). Int. J. Food Microbiol. 78, 197–216.
514 https://doi.org/https://doi.org/10.1016/S0168-1605(02)00188-5

515 Boenn, M., 2018. hypergea: Hypergeometric Tests. R package version 1.3.6,
516 https://CRAN.R-project.org/package=hypergea.

517 Cernicchiaro, N., Pearl, D.L., Ghimire, S., Gyles, C.L., Johnson, R.P., LeJeune, J.T.,
518 Ziebell, K., McEwen, S. a, 2009. Risk factors associated with Escherichia coli
519 O157:H7 in Ontario beef cow-calf operations. Prev. Vet. Med. 92, 106–15.
520 https://doi.org/10.1016/j.prevetmed.2009.07.004

521 Cernicchiaro, N., Pearl, D.L., McEwen, S. a, Harpster, L., Homan, H.J., Linz, G.M.,
522 Lejeune, J.T., 2012. Association of wild bird density and farm management factors
523 with the prevalence of E. coli O157 in dairy herds in Ohio (2007-2009). Zoonoses
524 Public Health 59, 320–9. https://doi.org/10.1111/j.1863-2378.2012.01457.x

525 Chase-Topping, M.E., Gally, D., Low, C., Matthews, L., Woolhouse, M., 2008. Super-
526 shedding and the link between human infection and livestock carriage of
527 Escherichia coli O157. Nat. Rev. Microbiol. 6, 904–12.
528 https://doi.org/10.1038/nrmicro2029

529 Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D.,
530 Parkhill, J., Harris, S.R., 2015. Rapid phylogenetic analysis of large samples of
531 recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res.
532 43, e15. https://doi.org/10.1093/nar/gku1196

533 Cuzick, J., Edwards, R., 1990. Spatial Clustering for Inhomogeneous Populations. J. R.
534 Stat. Soc. Ser. B 52, 73–96. https://doi.org/10.1111/j.2517-6161.1990.tb01773.x

535 Eriksson, E., Aspan, A., Gunnarsson, A., Vågsholm, I., 2005. Prevalence of verotoxin-
536 producing Escherichia coli (VTEC) O157 in Swedish dairy herds. Epidemiol.
537 Infect. 133, 349–358. https://doi.org/10.1017/S0950268804003371

538 Eriksson, E., Söderlund, R., Boqvist, S., Aspan, a, 2011. Genotypic characterization to
539 identify markers associated with putative hypervirulence in Swedish Escherichia
540 coli O157:H7 cattle strains. J. Appl. Microbiol. 110, 323–32.
541 https://doi.org/10.1111/j.1365-2672.2010.04887.x

542 Folkhälsomyndigheten, 2019. Enterohemorragisk E. coli infektion (EHEC) [WWW
543 Document]. URL http://www.folkhalsomyndigheten.se/amnesomraden/statistik-
544 och-undersokningar/sjukdomsstatistik/enterohemorragisk-e-coli-infektion-ehec/
545 (accessed 5.21.19).

546 Fox, J., Weisberg, S., 2011. An R Companion to Applied Regression, Second. ed. Sage,
547 Thousand Oaks CA.

548 Frank, C., Kapfhammer, S., Werber, D., Stark, K., Held, L., 2008. Cattle Density and
549 Shiga Toxin-Producing *Escherichia coli* Infection in Germany: Increased Risk for
550 Most but Not All Serogroups. Vector-Borne Zoonotic Dis. 8, 635–644.
551 https://doi.org/10.1089/vbz.2007.0237

552 Franz, E., Rotariu, O., Lopes, B.S., Macrae, M., Bono, J.L., Laing, C., Gannon, V.,
553 Söderlund, R., Hoek, A.H.A.M. Van, Friesema, I., French, N.P., George, T., Biggs,
554 P.J., Jaros, P., Rivas, M., Chinen, I., Campos, J., Jernberg, C., Gobius, K., Mellor,
555 G.E., Chandry, P.S., Perez-reche, F., 2018. Phylogeographic analysis reveals
556 multiple international transmission events have driven the global emergence of
557 Escherichia coli O157:H7. Clin. Infect. Dis. ciy919, 1–24.
558 https://doi.org/10.1093/cid/ciy919/5146342

559 French, J., 2018. smacpod: Statistical Methods for the Analysis of Case-Control Point
560 Data. R package version 2.0.4. https://cran.r-project.org/package=smacpod

561 Fremaux, B., Raynaud, S., Beutin, L., Rozand, C.V., 2006. Dissemination and
562 persistence of Shiga toxin-producing Escherichia coli (STEC) strains on French
563 dairy farms. Vet. Microbiol. 117, 180–191.
564 https://doi.org/10.1016/j.vetmic.2006.04.030

565 Gannon, V., D'souza, S., Graham, T., King, R.K., Rahn, K., Read, S., 1997. Use of the
566 flagellar H7 gene as a target in multiplex PCR assays and improved specificity in
567 identification of enterohemorrhagic Escherichia coli strains. J. Clin. Microbiol. 35,
568 656–662.

569 Herbert, L.J., Vali, L., Hoyle, D. V, Innocent, G., McKendrick, I.J., Pearce, M.C.,
570 Mellor, D., Porphyre, T., Locking, M., Allison, L., Hanson, M., Matthews, L.,
571 Gunn, G.J., Woolhouse, M.E.J., Chase-Topping, M.E., 2014. E. coli O157 on
572 Scottish cattle farms: Evidence of local spread and persistence using repeat cross-
573 sectional data. BMC Vet. Res. 10, 95. https://doi.org/10.1186/1746-6148-10-95

574 Hothorn, T., Hornik, K., Van De Wiel, M.A., Zeileis, A., 2006. A lego system for

conditional inference. Am. Stat. 60, 257–263. https://doi.org/10.1198/000313006X118430

Humphry, R.W., Cameron, A., Gunn, G.J., 2004. A practical approach to calculate sample size for herd prevalence surveys. Prev. Vet. Med. 65, 173–188. https://doi.org/10.1016/j.prevetmed.2004.07.003

Joris, M.-A., Verstraete, K., De Reu, K., De Zutter, L., 2013. Longitudinal Follow-Up of the Persistence and Dissemination of EHEC on Cattle Farms in Belgium. Foodborne Pathog. Dis. 10, 295–301. https://doi.org/10.1089/fpd.2012.1277

Karmali, M.A., 2004. Infection by Shiga toxin-producing Escherichia coli: an overview. Mol. Biotechnol. 26, 117–122. https://doi.org/10.1385/MB:26:2:117

Kistemann, T., Zimmer, S., Vågsholm, I., Andersson, Y., 2004. GIS-supported investigation of human EHEC and cattle VTEC O157 infections in Sweden: Geographical distribution, spatial variation and possible risk factors. Epidemiol. Infect. 132, 495–505. https://doi.org/10.1017/S0950268803001729

Lahti, E., Ruoho, O., Rantala, L., Hänninen, M., Honkanen-buzalski, T., 2003. Longitudinal Study of Escherichia coli O157 in a Cattle Finishing Unit. Appl. Environ. Microbiol. 69, 554–561. https://doi.org/10.1128/AEM.69.1.554

LeJeune, J., Besser, T., Rice, D., Berg, J., Stilborn, R.P., Hancock, D., 2004. Longitudinal study of fecal shedding of Escherichia coli O157: H7 in feedlot cattle: predominance and persistence of specific clonal types despite massive cattle population turnover. Appl. Environ. Microbiol. 70, 377–384. https://doi.org/10.1128/AEM.70.1.377

LeJeune, J.T., Wetzel, A.N., 2007. Preharvest control of Escherichia coli O157 in cattle. J. Anim. Sci. 85, E73-80. https://doi.org/10.2527/jas.2006-612

Lele, S.R., Keim, J.L., Solymos, P., 2019. ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data. R package version 0.3-4. https://cran.r-project.org/package=ResourceSelection

Lemeshow, S., Hosmer, D.W., 1982. A review of goodness of fit statistics for use in the development of logistic regression models. Am. J. Epidemiol. 115, 92–106. https://doi.org/10.1093/oxfordjournals.aje.a113284

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245. https://doi.org/10.1093/nar/gkw290

Majowicz, S.E., Scallan, E., Jones-Bitton, A., Sargeant, J.M., Stapleton, J., Angulo, F.J., Yeung, D.H., 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, 447–55. https://doi.org/10.1089/fpd.2013.1704

Manning, S.D., Motiwala, A.S., Springman, A.C., Qi, W., Lacher, D.W., Ouellette, L.M., Mladonicky, J.M., Somsel, P., Rudrik, J.T., Dietrich, S.E., Zhang, W., Swaminathan, B., Alland, D., Whittam, T.S., 2008. Variation in virulence among clades of Escherichia coli O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. U. S. A. 105, 4868–73. https://doi.org/10.1073/pnas.0710834105

Nöremark, M., Frössling, J., Lewerin, S.S., 2013. A survey of visitors on swedish livestock farms with reference to the spread of animal diseases. BMC Vet. Res. 9.

https://doi.org/10.1186/1746-6148-9-184

Paton, J., Paton, A., 1998. Pathogenesis and diagnosis of Shiga toxin-producing Escherichia coli infections. Clin. Microbiol. Rev. 11, 450–479.

R Core Team, 2018. R: A Language and Environment for Statistical Computing.

Ripley, B.D., 1981. Spatial statistics. John Wiley and Sons, New York.

Sanderson, M.W., Sargeant, J.M., Shi, X., Nagaraja, T.G., Zurek, L., Alam, M.J., 2006. Longitudinal emergence and distribution of Eschenchia coli O157 genotypes in a beef feedlot. Appl. Environ. Microbiol. 72, 7614–7619. https://doi.org/10.1128/AEM.01412-06

Schouten, J.M., Bouwknegt, M., Van De Giessen, a. W., Frankena, K., De Jong, M.C.M., Graat, E. a M., 2004. Prevalence estimation and risk factors for Escherichia coli O157 on Dutch dairy farms. Prev. Vet. Med. 64, 49–61. https://doi.org/10.1016/j.prevetmed.2004.03.004

Seemann, T., Goncalves da Silva, A., Bulach, D., Schultz, M., Kwong, J., Howden, B., 2017. Nullarbor v 1.2.8. https://github.com/tseemann/nullarbor.

Socialstyrelsen, 2014. Infektion med EHEC/VTEC: Ett nationell strategidokument. [WWW Document]. URL http://www.socialstyrelsen.se/nyheter/2014december/strategiskaminskariskenforehec-infektionhosmanniskor (Accessed 2019-01-03).

Söderlund, R., Jernberg, C., Ivarsson, S., Hedenström, I., Eriksson, E., Bongcam-Rudloff, E., Aspán, A., 2014. Molecular typing of Escherichia coli O157:H7 isolates from Swedish cattle and human cases: population dynamics and virulence. J. Clin. Microbiol. 52, 3906–12. https://doi.org/10.1128/JCM.01877-14

642 Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-
643 analysis of large phylogenies. Bioinformatics 30, 1312–1313.
644 https://doi.org/10.1093/bioinformatics/btu033

645 Strachan, N.J.C., Rotariu, O., Lopes, B., Macrae, M., Fairley, S., Laing, C., Gannon, V.,
646 Allison, L.J., Hanson, M.F., Dallman, T., Ashton, P., Franz, E., Van Hoek,
647 A.H.A.M., French, N.P., George, T., Biggs, P.J., Forbes, K.J., 2015. Whole
648 Genome Sequencing demonstrates that Geographic Variation of Escherichia coli
649 O157 Genotypes Dominates Host Association. Sci. Rep. 5, 1–10.
650 https://doi.org/10.1038/srep14145

651 Swirski, A.L., Pearl, D.L., Williams, M.L., Homan, H.J., Linz, G.M., Cernicchiaro, N.,
652 Lejeune, J.T., 2014. Spatial Epidemiology of Escherichia coli O157: H7 in Dairy
653 Cattle in Relation to Night Roosts Of Sturnus vulgaris (European Starling) in Ohio,
654 USA (2007-2009). Zoonoses Public Health 61, 427–435.
655 https://doi.org/10.1111/zph.12092

656 Synge, B.A., Chase-Topping, M.E., Hopkins, G.F., McKendrick, I.J., Thomson-Carter,
657 F., Gray, D., Rusbridge, S.M., Munro, F.I., Foster, G., Gunn, G.J., 2003. Factors
658 influencing the shedding of verocytotoxin-producing Escherichia coli O157 by
659 beef suckler cows. Epidemiol. Infect. 130, 301–312.
660 https://doi.org/10.1017/S0950268802008208

661 Tamminen, L.-M., Fransson, H., Tråvén, M., Aspán, A., Alenius, S., Emanuelson, U.,
662 Dreimanis, I., Törnquist, M., Eriksson, E., 2018. Effect of on-farm interventions in
663 the aftermath of an outbreak of hypervirulent verocytotoxin-producing Escherichia
664 coli O157:H7 in Sweden. Vet. Rec. 182, 516. https://doi.org/10.1136/vr.104223

665 Turck, N., Vutskits, L., Sanchez-Pena, P., Robin, X., Hainard, A., Gex-Fabry, M.,
666 Fouda, C., Bassem, H., Mueller, M., Lisacek, F., Puybasset, L., Sanchez, J.-C.,
667 2011. pROC: an open-source package for R and S+ to analyze and compare ROC
668 curves. BMC Bioinformatics 8, 12–77. https://doi.org/10.1007/s00134-009-1641-y

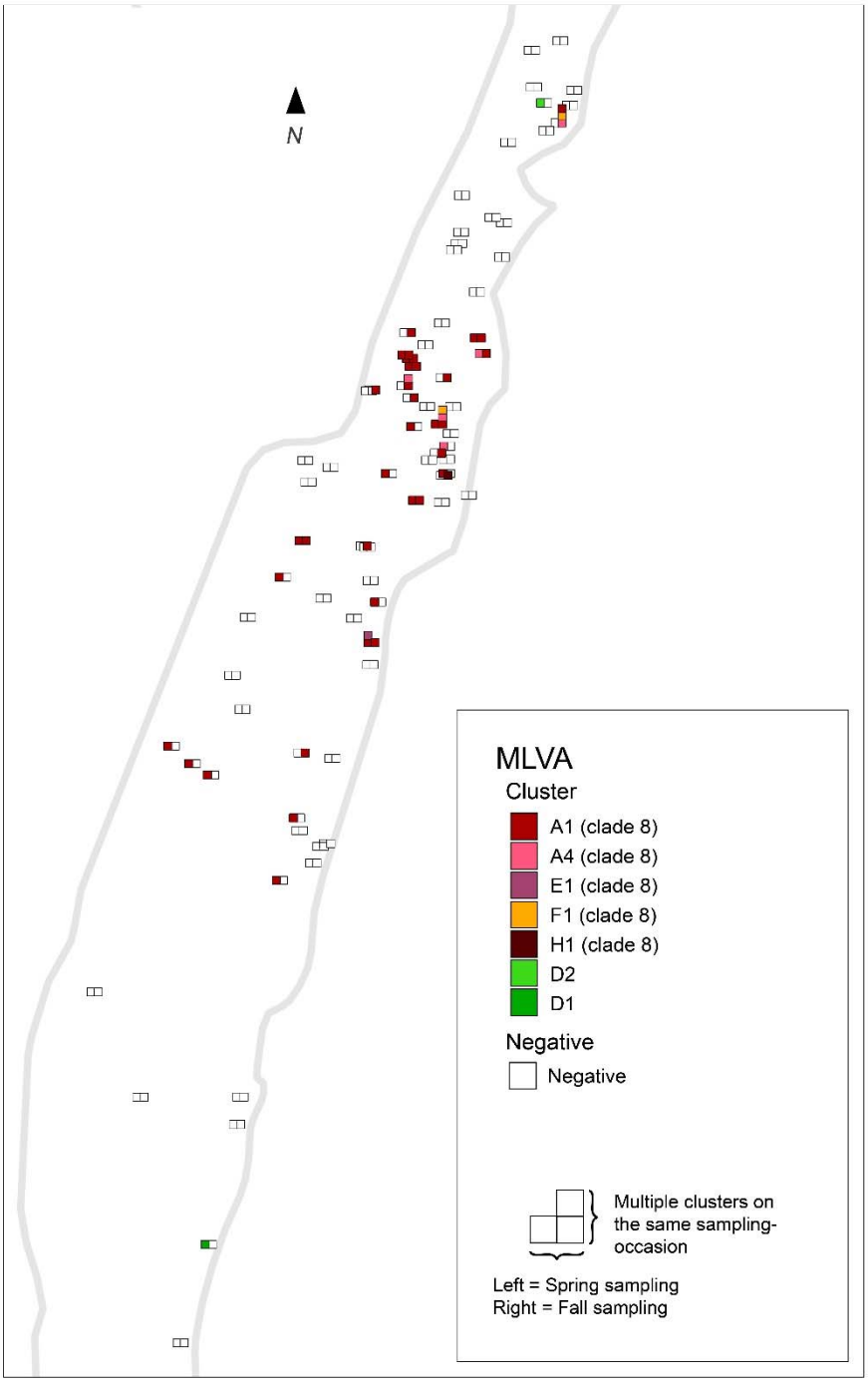
669 Widgren, S., Engblom, S., Emanuelson, U., Lindberg, A., 2018. Spatio-temporal
670 modelling of verotoxigenic Escherichia coli O157 in cattle in Sweden: exploring
671 options for control. Vet. Res. 49, 78. https://doi.org/10.1186/s13567-018-0574-2

672 Widgren, S., Eriksson, E., Aspan, A., Emanuelson, U., Alenius, S., Lindberg, A., 2013.
673 Environmental sampling for evaluating verotoxigenic Escherichia coli O157: H7
674 status in dairy cattle herds. J. Vet. Diagn. Invest. 25, 189–98.
675 https://doi.org/10.1177/1040638712474814

676 Widgren, S., Söderlund, R., Eriksson, E., Fasth, C., Aspan, A., Emanuelson, U.,
677 Alenius, S., Lindberg, A., 2015. Longitudinal observational study over 38 months
678 of verotoxigenic Escherichia coli O157:H7 status in 126 cattle herds. Prev. Vet.
679 Med. 121, 343–352. https://doi.org/10.1016/j.prevetmed.2015.08.010

680 Wilson, J.B., McEwen, S. a., Clarke, R.C., Leslie, K.E., Waltner-Toews, D., Gyles,
681 C.L., 1993. Risk factors for bovine infection with verocytotoxigenic Escherichia
682 coli in Ontario, Canada. Prev. Vet. Med. 16, 159–170.
683 https://doi.org/10.1016/0167-5877(93)90063-Y

684 Zhang, X.S., Chase-Topping, M.E., McKendrick, I.J., Savill, N.J., Woolhouse, M.E.J.,
685 2010. Spread of E. coli O157 infection among Scottish cattle farms: Stochastic
686 models and model selection. Epidemics 2, 11–20.
687 https://doi.org/10.1016/j.epidem.2010.02.001



688

689 **Figure 1.** Presence of VTEC O157:H7 established by environmental sampling of 80
690 Swedish farms in spring and fall 2014. Colour represents MLVA-type of isolates.
691 Location of farms have been nudged and presentation of the coastline indistinct to avoid
692 identification of individual farms. The island is 137 km long and 16 km wide (at the
693 widest point).
694

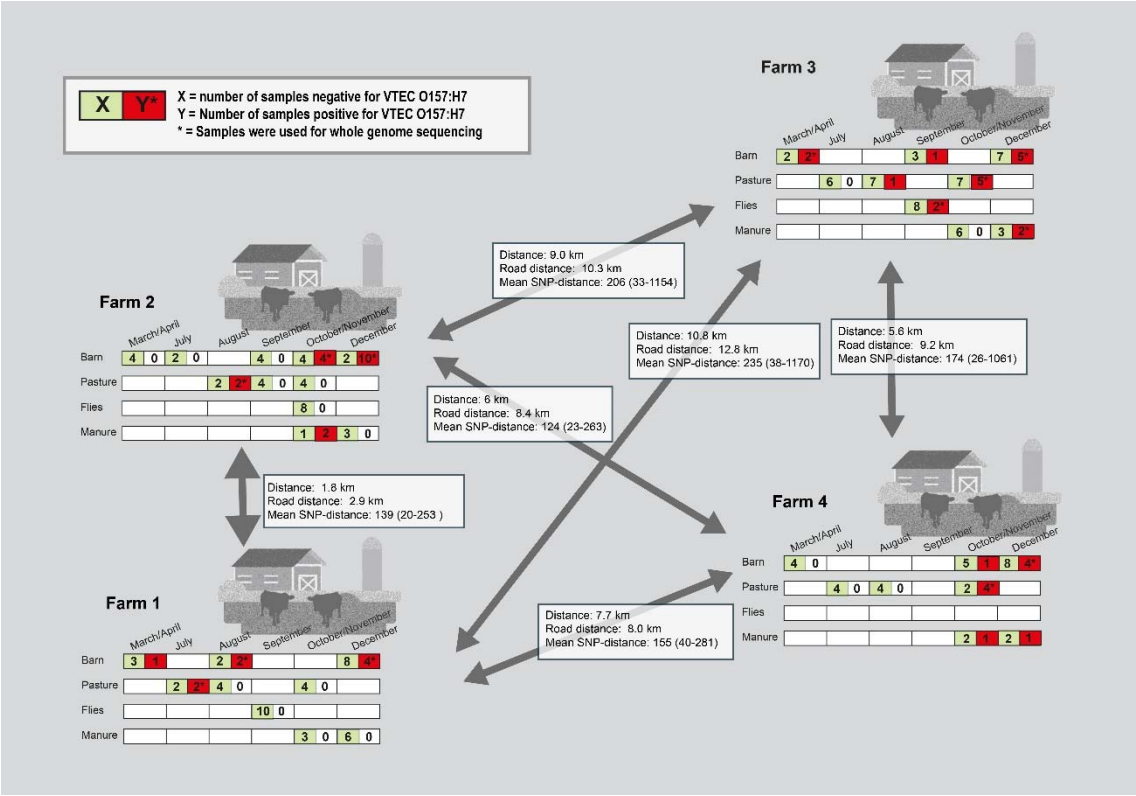
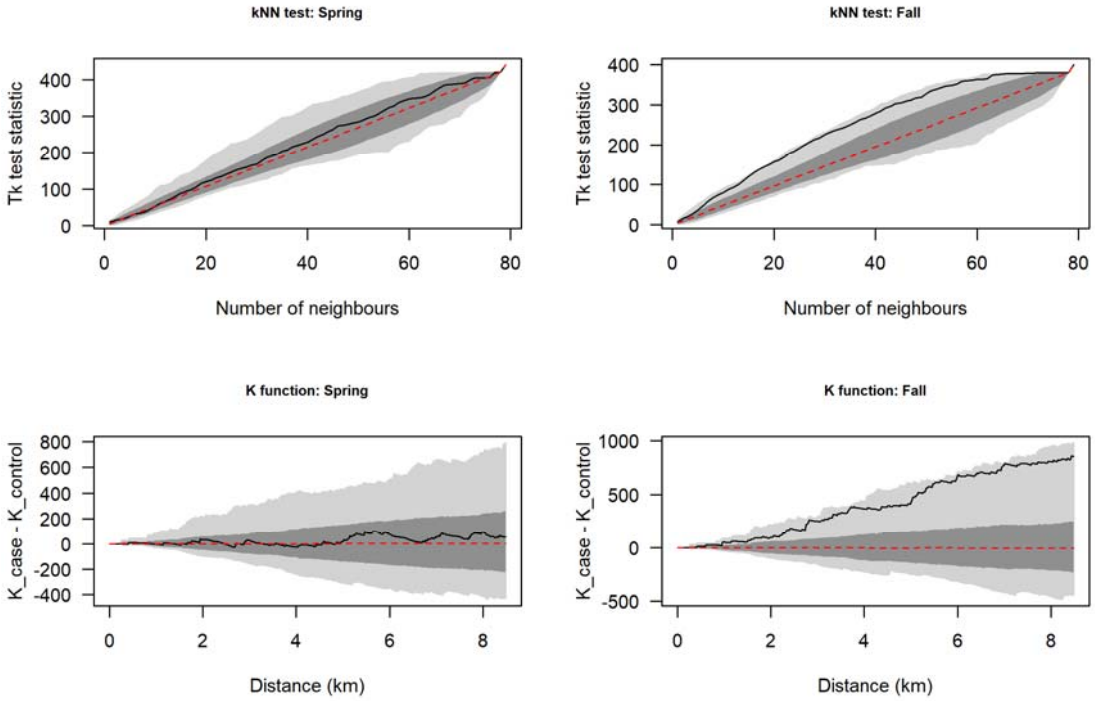


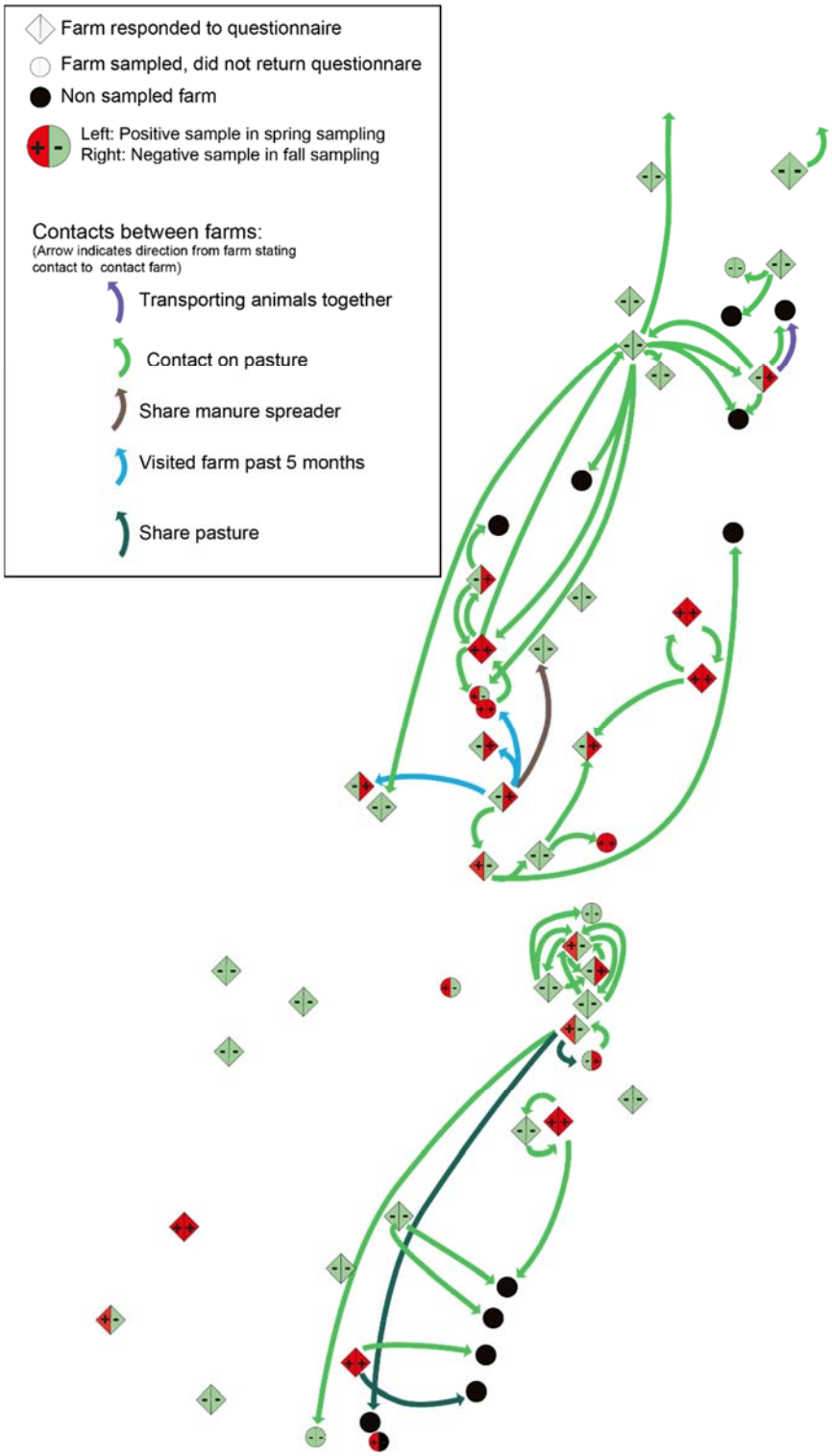
Figure 2. Additional sampling occasions and types of samples collected from the four farms that were part of a parallel research project during 2014. From sampling occasions with positive samples marked with * isolates were used for whole genome sequencing.



701

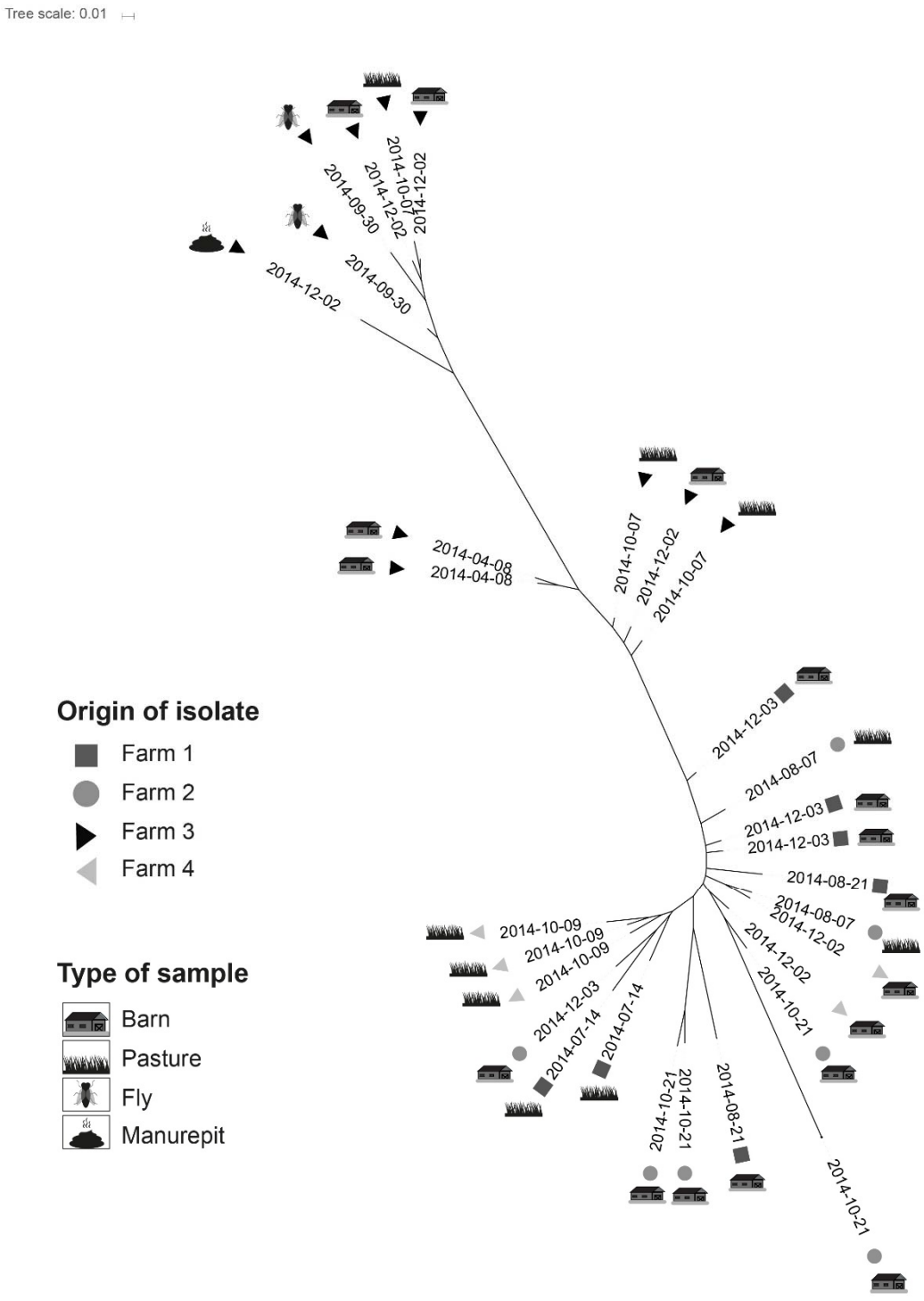
702 **Figure 3.** Spatial clustering of positive herds in the spring (left) and the fall (right). Top
703 row shows results of the Cuzick-Edwards' kNN test: clustering is observed when the
704 test statistic for the observed distribution (solid black line) exceeds upperbound of the
705 95% envelope of test statistics for simulated distributions (darkgrey area). Bottom row
706 shows results of K function test: here, the difference between K functions for cases and
707 controls was measured.

708



709

710 **Figure 4.** Between farm contacts in between samplings as specified by farmers in
711 questionnaire. (Figure represents part of the island and is not to scale. The location of
712 farms have been shifted to avoid identification of individual farms and enable
713 presentation.

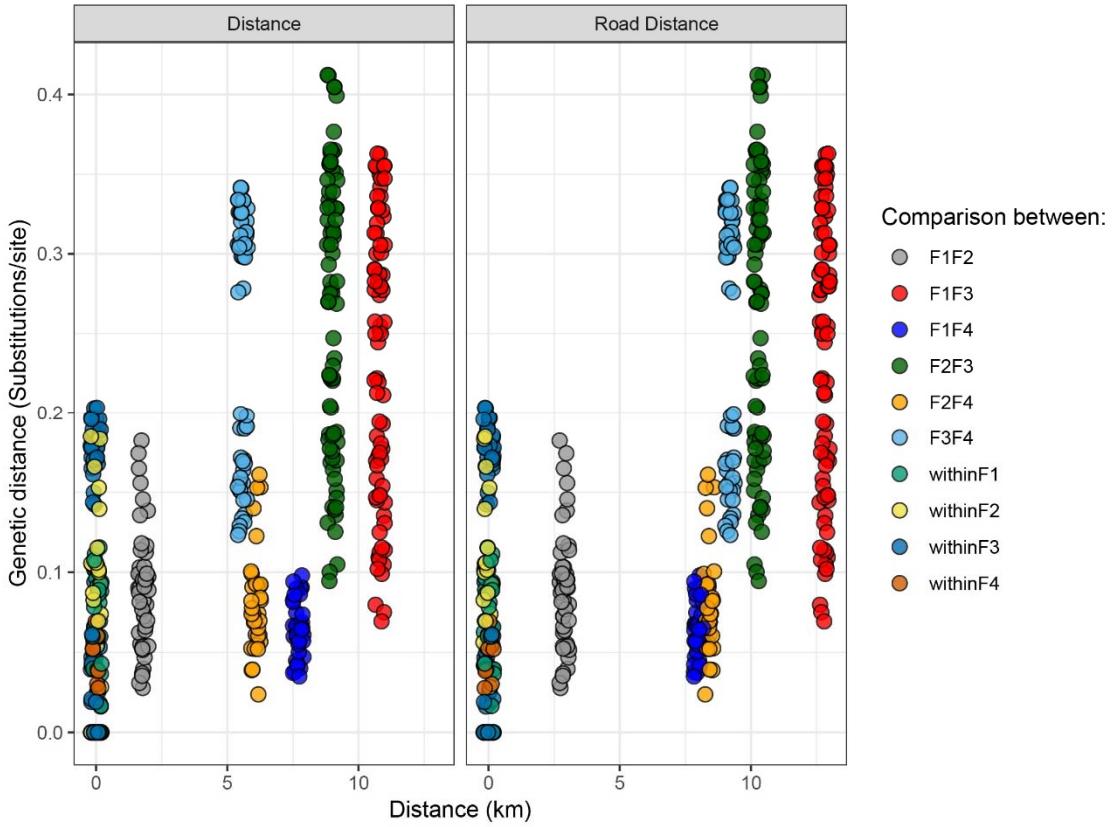


714

715 **Figure 5.** Phylogenetic tree based on core SNP-distance between isolates. Distance

716 indicates substitutions per site and date indicates day of sampling.

717



718

719 **Figure 6.** Genetic distance between all isolates (based on core SNP-distance derived
720 from the Maximum Likelihood phylogenetic tree) and the association with distance
721 between farms and road distance between farms.

722

723 **Supplementary material**

724 Table S1. Responses to farmer questionnaires sent out in fall 2014. (*NN= farm negative on*

725 *both sampling occasions, NP=negative in spring, positive fall, PN=positive spring, negative fall, PP =*

726 *positive on both occasions*

		NN	NP	PN	PP
Number of farms:		33	9	7	6
Farm characteristics:					
Type of farm	Beef	1	1	0	0
	Milk	18	6	5	1
	Combination	14	1	2	5
Dog	Yes	20	8	5	4
	No	13	1	2	2
Cat	Yes	6	3	2	3
	No	27	6	5	3
Sheep	Yes	7	2	1	2
	No	26	7	6	4
Horse	Yes	3	3	1	1
	No	30	6	6	5
Pig	Yes	2	0	0	0
	No	31	9	7	6
Poultry	Yes	5	0	2	0
	No	28	9	5	6
Using reproductive services	Yes	19	3	6	4
	No	14	6	1	2
Employees	Yes (without own animals)	14	4	4	4
	Yes (with own animals)	12	4	2	1
	No	6	1	1	1
	Missing	1	0	0	0
Collaborations and sharing of equipment					
Share agricultural machines	Yes	18	7	2	5
	No	15	2	5	1
Share claw treatment crush	Yes	5	3	1	0
	No	28	6	6	6
Share vehicles for animal transport	Yes	6	2	1	2
	No	27	7	6	4

Share manure spreader	Yes	21	8	4	4
	No	12	1	3	2
Pest problems:					
Wild game	Yes	2	2	1	0
	No	31	7	6	6
Birds	Yes	14	4	3	4
	No	19	5	4	2
Rodents	Yes	2	1	1	1
	No	31	8	6	5
Other	Yes	0	0	0	2
	No	33	9	7	4
Cleaning routines:					
Use high pressure for cleaning	Yes	17	3	5	5
	No	14	4	1	1
	Missing	2	2	1	0
Use hot water for cleaning	Yes	4	1	0	1
	No	27	6	6	5
	Missing	2	2	1	0
Only clean out bedding material from pens	Yes	2	1	1	1
	No	29	6	5	5
	Missing	2	2	1	0
Use slaked lime	Yes	9	1	2	3
	No	24	8	5	3
Between sampling events					
Contacts:					
Purchased animals	Yes	3	4	1	1
	No	30	5	6	5
Shared pasture during 2014	Yes	5	0	2	1
	No	28	9	5	5
Nose-nose contact on pasture	Yes	21	6	6	5
	No	12	3	1	1
Nose-nose contact with known positive	Yes	2	0	0	0
	No	31	9	7	6
Access to natural water resources on pasture	Yes	23	6	6	5
	No	10	3	1	1
Transport animals together with animals from other farm	Yes	2	1	0	0
	No	31	8	7	6

Transport with known positive farm	Yes	0	0	0	0
	No	33	9	7	6
Any known contact with positive	Yes	7	5	0	4
	No	26	4	7	2
Visits to other farms the passed 5 months	Yes	12	3	0	3
	No	19	6	7	3
	Missing	2	0	0	0
Cleaning					
Cleaning and disinfection of emptied stable during summer 2014	Yes	24	7	6	4
	No	7	2	1	2
	Missing	2	0	0	0
Continuous cleaning during summer	Yes	21	4	5	6
	No	11	3	1	0
	Missing	1	2	1	0
Change in cleaning routines the passed 5 months	Yes	1	0	1	0
	No	32	9	6	6

727