CENTER for PRODUCE SAFETY

CPS 2011 RFP
FINAL PROJECT REPORT

Project Title
The Role of Riparian Zones in Bacteria Dispersal to Produce Farms

Project Period
January 1, 2012 – December 31, 2013

Principal Investigator
Martin Wiedmann
Department of Food Science
Cornell University
607-254-2838, martin.wiedmann@cornell.edu

Co-Principal Investigator(s):

Objectives
Main Objective: Improve the prediction of produce contamination risk by modeling the rules of pathogen dispersal through riparian zones to produce fields.

> *Objective 1: Define candidate models for pathogen dispersal based on remotely-sensed and field-collected data to predict the dispersal efficiency of fecal bacteria across an agriculture landscape.*
> *Objective 2: Determine the genetic distribution of fecal bacteria from fecal and soil samples at participating farms in the landscape using multilocus sequence typing.*
> *Objective 3: Conduct goodness-of-fit tests comparing the fecal bacteria genetic distribution on the landscape with predictions of high efficiency dispersal pathways calculated from the models.*

Funding for this project provided by the Center for Produce Safety through:
CDFA SCBGP grant #SCB11072 and CPS Campaign for Research

FINAL REPORT

Abstract

Riparian buffer zones in agricultural landscapes provide essential ecosystem services while increasing species diversity and habitat connectivity for wildlife and plants. Unfortunately, improved wildlife habitat adjacent to crop production areas may increase risk for contamination of fruits and vegetables. Therefore, buffer zones could serve as transport pathways across landscapes for foodborne pathogens. Accurate information about the role of buffer zones in dispersing foodborne pathogens is essential for the produce industry to strike a balance between the use of buffer zones to enhance soil and water quality while controlling the risks associated with infiltration of fields by wildlife. In this project, we applied an innovative approach, using techniques that have been pioneered in the emerging field of landscape genetics, to systematically answer questions about pathogen movement, sources, and sinks within agricultural landscapes. We specifically selected two distinct produce growing areas in New York to (i) collect data on *E. coli* prevalence and genetic diversity in produce fields and adjacent riparian zones and to (ii) perform causal modeling of the effect of riparian zone and produce field qualities on the movement of fecal bacteria across agricultural landscapes. Competing models were formulated using remotely-sensed and field-collected data, e.g. vegetation density and type, buffer width, and soil habitat quality. These models were used to provide data to predict produce contamination risk in the landscape context and to produce maps of the most efficient paths for fecal bacteria across the landscape. The predictions from these maps were then compared to genetic data for the *E. coli* isolates collected in the field in order to identify the models that best fit our observed data. The set of models producing the best fits to the genetic data defines landscape attributes that promote pathogen dispersal. This approach has the advantage of analyzing fecal bacteria directly, as opposed to tracking host animals, and thus can account for unmeasurable events such as transmission of bacteria among host individuals or even among host animal species during dispersal. In addition, these models can also account for persistence across extrahost environments. Our specific analyses showed that models that included riparian corridors or water as a barrier to movement best fit our data, suggesting that terrestrial dispersal by animals was an important dispersal process for commensal *E. coli.*

Background

Riparian buffer strips are vegetative zones established between farm fields and streams that provide conservation (e.g. reducing soil erosion and leaching of chemical and fecal waste into ground water sources) and aesthetic benefits (habitat and connective pathways for wildlife) (Berges *et al.*, 2010, Borin *et al.*, 2010, Finder *et al.*, 1999). In contrast to the clear conservation value of providing wildlife habitat, the enhanced habitat value of riparian buffers may create unintended consequences for produce growers by providing wildlife with movement pathways under tree cover and near fresh water that is also adjacent to highly desirable food crops. In particular birds, deer, and feral swine have been implicated as sources of foodborne pathogens including *E. coli* O157:H7 and all three groups prefer forested corridors for movement.

Wildlife, which are colonized by commensal and pathogenic enteric bacteria, are thought to be important vehicles for transmission of foodborne pathogens. Enteric bacteria can be transmitted between hosts through contact and be deposited in new locations (e.g., produce fields) by defecation, which often occurs while animals are foraging for food (Avery *et al.*, 2004). Once present in a produce field, the fecal bacteria can be dispersed onto crops by insect species which may use both feces and plants as grazing surfaces. Storms can also act to wash fecal bacteria into contact with crops. Fecal deposits can also decompose and be integrated into soil, where foodborne pathogens can persist for months (Byappanahalli *et al.*, 2006, Islam *et al.*, 2005, van Elsas *et al.*, 2007). Since enteric bacteria can transiently inhabit new hosts, it is possible for wildlife to act as a link between domestic animal operations and produce farms in agricultural landscapes.

While a number of sources cite infiltration of production blocks by wild mammals and birds as a substantial risk to food safety, the evidence that wild animals are a significant source of produce contamination is less than clear. In the 2006 spinach outbreak, the outbreak strain of *E. coli* O157:H7 was isolated from cattle, wild hogs, soil and water. Numerous surveys targeting deer have reported prevalences of pathogenic *E. coli* in wild deer ranging from 15% (Bardiau *et al.*, 2010) to 23.9% (Sanchez *et al.*, 2009). Other studies have reported prevalence of specifically the O157:H7 group as between 0.5% to 2.3% (Henderson, 2008). Shiga toxin-producing *E. coli* were isolated from up to 5% of wild boars in one study (Sanchez *et al.*, 2010). While wildlife carriage of fecal pathogens has been firmly established, their potential to act as vehicles for produce contamination requires further investigation. Given the considerable controversy surrounding the role of wildlife in contamination of fresh fruit and vegetable crops, it is critical to study the role of riparian buffers in supporting the transmission of foodborne pathogens onto produce farms.

The dispersal of fecal bacteria and foodborne pathogens across an agrarian landscape is a complex process in which multiple sources, transport paths and sinks may play a role (Oliver *et al.*, 2009). Landscape ecological methods enable the quantitative description and prediction of pathogen transmission as it relates to landscape characteristics. The role of riparian buffers can then be quantified by modeling the interactions between land-uses against a multi-scale background generated by bacterial dispersal and persistence in soil and water environments within farms (Cushman *et al.*, 2006, Oliver *et al.*, 2009). These factors interact to describe a probability that a fecal bacterium deposited in one location will survive, disperse and/or reproduce. In this context, riparian zones potentially represent a “choke point” in the transmission of foodborne pathogens from various sources to produce fields.

To better understand the role of land-use in dispersal of pathogens from riparian zones onto fields and the role of riparian zones as corridors between domestic animal operations and produce fields, we employed a landscape level analysis that directly analyzes the interactions among land-uses at several spatial scales to test the dispersal of bacteria within and among fields. In this approach, GIS-based methods that combine layered information on landscape characteristics were used to develop numerous competing models that predict bacterial dispersal. These models were then tested and analyzed using genetic analysis of conserved loci in commensal *E. coli* isolates obtained from field sites, in order to identify the models that best fit the observed data on *E. coli* distribution and subtypes. This approach enabled us to create information on the movement of bacteria without the need for direct observation of dispersal.

Environmental sources or sinks of *E. coli* could also be important to the overall dispersal process. For example, the soil environment plays a role in the distribution of *E. coli* outside the host animal. In one study, the complexity of the soil microbial community was shown to influence persistence when the soil habitat was otherwise held constant (van Elsas *et al.*, 2007). In a further study, fecal bacteria population structure in soil at the field scale indicated that bacteria could disperse easily across 3-10 and 20-30 m distance scales under periods of little precipitation; this dispersal was probably caused by movement of animal hosts in the field (Bergholz *et al.*, 2011). This analysis and others have also determined that soil pH and organic matter are the “master” chemical variables in soil that determine the habitat quality for enteric bacteria (Semenov *et al.*, 2008). Near-neutral pH values with low pH variation were able to support a greater abundance and diversity of *E. coli* and were able to support slower population decline after inoculation into soil (P. Bergholz, unpublished data). Because soil characteristics can determine the persistence and diversity of fecal bacteria, certain soils may act as a source pools for dispersal of fecal bacteria across fields or from the riparian buffer strips to the field. One of the strengths of the modelling approach we used was that our models also accounted for the quality of environmental reservoirs of *E. coli* to estimate the load of *E. coli* arising from various locations.

Research Methods and Results

GIS data acquisition and landscape selection. In our first project objective, we defined models of bacterial dispersal in a landscape surrounding a riparian zone based on remotely-sensed and field-collected data. Two watershed landscapes, Flint Creek and Hoosic River, located in the central New York region, were selected for this study based on topography and land-cover composition using remotely-sensed GIS data. If wildlife are an important vehicle for dispersal of bacteria in these landscapes, land-cover configuration can substantially influence the behavior of wildlife and foodborne disease bacteria around different creeks. These streams represent a diversity of vegetable growing areas within New York State. Flint Creek is in an intensive vegetable and livestock production region that is sparsely forested, whereas the Hoosic River watershed is heavily forested areas interspersed between sporadic production regions. These two riparian areas represent an interesting comparison between a riparian area with a higher quality forest versus an area with a lower quality forest.

Dispersal model formulation for fecal bacteria moving among produce fields with wildlife. To determine the causes of fecal bacteria dispersal across agrarian landscapes, multiple alternative dispersal efficiency models were developed to describe landscape effects by integrating remotely-sensed and field-collected data into movement cost surfaces. Cost surfaces are maps indicating the difficulty of movement across a single unit of land (3 m x 3 m in this study) (Epps *et al.*, 2007, Cushman *et al.*, 2006). Higher cost indicates less efficient dispersal. These models provide data to predict produce contamination risk in the landscape context.

Many alternative hypotheses exist that compete against wild-ruminant driven dispersal of foodborne pathogens: i) a null model in which distance among sites is the only limitation to dispersal, ii) a model based on avian scavengers (e.g. crow or European starling), which tend toward cyclic movements over kilometers and in large flocks during the autumn, iii) a model based on migratory bird flocks (e.g., Canada geese) that move at scales larger than our study design, iv) water-driven dispersal, which would correspond best to flow through a hydrologic network, v) a model based on small terrestrial animals that migrate short distances over their lifetimes but may form transmission networks, and vi) a model based on large terrestrial nuisance wildlife such as whitetailed deer or feral swine. Each of these models was varied by adjusting the strengths of urban or hydrologic barriers (three levels) or by adjusting the spatial extent of the dispersal corridor around the riparian zone (three levels). Dispersal rates among reservoirs and croplands were determined by running a random walk model that resembles how animals make choices about movement from source reservoirs (e.g. livestock pastures) to cropland sink reservoirs. The random walk model uses the landscape resistance models and terrain models together to estimate dispersal cost. The random walk for avian vehicles did not take into account topography during the walk, but models for terrestrial vehicles did.

Quantitative models were calculated representing ideas about how *E. coli* might move across the landscape using an open source geographic information system (GIS).

Overall, a total of 144 landscape resistance models were formulated for each watershed (see examples in Figure 1). While the models belong to four major varieties based on the aforementioned four classes of wild animal vehicles carrying *E. coli* among sites, there are numerous subsets of models that vary in: i) how riparian corridors are represented, ii) the strength of barriers to movement (such as major roadways) and iii) the strength and type of interactions among forests, produce fields, pasturage, and urban areas. After application of various attraction models the total number of unique dispersal predictions was 576. A set of programs was developed in the GRASS GIS programming environment to automate landscape resistance model and dispersal rate calculation (Table 2). Predicted dispersal among sites was calculated according to the equation below:

$$D_{i,j} = \frac{L_i \times A_j}{C_{i,j}}$$

Where $D_{i,j}$ is the dispersal rate among sites i and j, L_i is the load from the source site (i.e., starting point), A_j is the attraction/gravity coefficient of the sink site (i.e., stopping point) and $C_{i,j}$ is the least-cost distance between sites i and j. For each dispersal model in each landscape, this generates and association matrix **D** containing predicted dispersal rates among all pairs of sites sampled in this study.

Sample collection and isolation. Sampling was conducted in Flint Creek watershed and the Hoosic River watershed, the two New York watersheds for which dispersal models were developed. Samples (soil, drag, fecal, and water) were collected from seven New York farms comprising a total of twenty-three produce field sites and seventeen forest sites. Nine produce field sites and five forest sites were isolated by distance (greater than 1 km) from the main stream in the watershed; the remaining sites were classified as adjacent to the main stream, meaning the sites were only separated from the main water stream by a forested boundary.

In total, 571 samples were collected, including: 116 drag swabs, 157 fecal samples, 278 soil samples, and 20 surface water samples from the main stream. Among the two watersheds samples, 294 samples were collected from the Flint Creek watershed and 277 were collected from the Hoosic River watershed. Samples were returned to the lab and attempts were made to cultivate up to 10 generic *E. coli* isolates from each sample. In total, 74% of drag swab samples, 53% of soil samples, 76% of fecal samples and 100% of water samples were positive for generic *E. coli*, yielding a total of 2,871 *E. coli* isolates.

Environmental distribution of isolates. While the proportion of fecal samples yielding *E. coli* was somewhat similar at 74% in Flint Creek and 78% in the Hoosic River area, there was a pronounced difference in the proportion of positive soil samples between the two riparian areas: 35% of soils in the Flint Creek area and 72% of soils in the Hoosic River area were positive for *E. coli*. In general, soil samples obtained from riparian forest soils were more likely to yield *E. coli* than soil samples from produce field, with prevalence values of 70% and 48%, respectively. Moreover, samples collected from boundaries between fields and forests were somewhat more likely than samples from the produce fields proper (sampling sites located at least 25 m from the forest-field boundary) to yield *E. coli* isolates with 59% and 45% prevalence, respectively. These results indicate that forest habitats harbor *E. coli* in extra-host environments with greater prevalence, and likely abundance, than produce fields. Hence forest habitats may act as environmental sources for *E. coli* dispersal to produce fields. In particular, the observation that soil samples from the sparsely forested Flint Creek area showed considerably lower *E. coli* prevalence than soil samples from the more forested Hoosic River area conspicuously point toward the local density of forest cover (i.e., percent land cover of forest) as potentially correlating with the prevalence of *E. coli* in environmental samples. Likewise, the data are consistent with a possible role of riparian zones as a dispersal corridor for the transmission of enteric bacteria between and towards produce fields.

DNA sequence data collection. Objective 2 of our research project was to use a multi-locus typing scheme to enable us to estimate the rates and distance of movement of *E. coli* isolates among produce field and forest sites. The 2,871 isolates collected from our sampling effort were sequenced at two loci, *mdh* and *uidA*, to identify unique sequence types within each sample. This approach resulted in a subset of 1,426 isolates representing 301 unique sequence types. A single representative from each of these sequence types was selected for each sample and sequenced at five additional genetic loci (*aspC, clpX, icd, lysP, fadD*) in order to better identify different subtypes. In total, nearly 26,000 nucleotide sequence reads were collected during this project.

At completion of nucleotide sequence collection, 1,356 seven-gene multilocus nucleotide sequences were available to represent the genetic diversity of the original 2,871 isolates. There were 573 unique multilocus sequence types identified in the isolate collection belonging to eight subpopulations of *E. coli*. Phylogenetic analysis confirmed that these isolates belonged to *E. coli* ECOR groups (subpopulations) A, B1, B2, D, E and F and an intermediate clade called B1A (Figure 2). ECOR B1 is thought to be the dominant type in wild and domestic animals. ECOR A and B2 are the dominant types in human fecal samples. ECOR D has been reported to not be dominant in fecal or clinical samples, but comprises 10-20% of isolates from humans (Tenaillon *et al.*, 2010). However, its proportional representation in soils is consistently 20-30%, suggesting that this subpopulation may have enhanced capacity to persist in complex environments outside the host (Meric *et al.*, 2013 and Bergholz *et al.*, 2011). Shifts in proportional representation of ECOR groups were observed among sample types with ECOR D present in greater proportions in fecal and soil samples and ECOR B1 clades present in greater proportion in drag swabs and water isolate collections. The proportions of ECOR groups were similar when samples were subdivided by land cover into samples from riparian forests and samples from vegetable fields, though ECOR D was present in greater proportion in drag swab samples from forests than from fields (Table 1).

Tests of genetic divergence among isolates were performed to understand whether there was large-scale geographic structure (divergence between *E. coli* from the two riparian areas) or evidence of differential selection on populations in vegetable fields vs forests. We hypothesized that populations would be more genetically diverse in riparian forests than in vegetable fields because farm fields have less spatial heterogeneity than forests but greater fluctuations in key meteorological parameters like humidity, soil moisture, temperature and exposure to direct sunlight. We expected that nucleotide diversity in isolates from farm fields would be lower, on average, but that evidence for divergence between populations would be limited. ECOR B2 and ECOR D both exhibited evidence of modestly greater nucleotide diversity in forest populations vs those in vegetable fields. No fixed mutations were observed in any ECOR groups when comparing forest and field or Hoosic and Flint populations, suggesting there are not specific sequence types within an ECOR group that are predominant in a particular site. All together, we conclude the forests supported somewhat greater genetic diversity of *E. coli* types compared with vegetable fields but forest and field populations are not genetically isolated. There are two possible explanations for these observations: a) forests are a source habitat for *E. coli* types in vegetable fields, or b) both forests and fields are sink habitats receiving *E. coli* types from the same source, but forests support the persistence of more *E. coli* types. Little evidence of differentiation between the populations in the two sampled landscapes was observed, though ECOR B2 had greater nucleotide diversity in the Flint Creek landscape compared to the Hoosic River landscape.

Statistical goodness-of-fit tests for dispersal models

Actual dispersal of *E. coli* among sampled sites

There are two means of measuring dispersal: i) assignment methods that quantify shared sequence types of *E. coli* among sampled sites, and ii) phylogenetic methods that measure genetic relationships among *E. coli* from different sites. We have estimated dispersal among sites using an assignment method that compares the number of shared and unshared types among sampled locations using a Sørensen index of similarity normalized to the total number of types in each pair of sites. In this analysis, an index of 1.0 means that all types are shared among a pair of sites. An index of 0 means that no types are shared. It was also important to examine the level of population structure imposed by the environment, because strong environmental selection acting on the *E. coli* population could obscure dispersal patterns. While assignment methods are most appropriate to examine recent dispersal events, phylogenetic methods can better reveal patterns of environmental selection acting on *E. coli* across the landscape. The phylogenetic relatedness of *E. coli* isolates from soil samples was estimated using ClonalFrame. This genetic distance (difference) output for each *E. coli* subgroup was tested for correlation with a matrix describing differences in the soil environments from which isolates were obtained.

Low levels of sequence type dispersal among sites were observed in both sampled landscapes. The frequency of shared sequence types between sampled sites was lower in Flint Creek compared with the Hoosic River. However, there was greater dispersal among forest sites than among field sites in the Flint Creek landscape. On average forested sites in the Flint Creek landscape shared sequence types with 5.5 other sites (min=1, max=9, n=15) and vegetable fields shared sequence types with 3.6 other sites (min=0, max=8, n=15). The Hoosic River landscape, by comparison, displayed a much greater degree of linkage as measured by shared sequence types. Fields in the Hoosic River landscape linked to an average of 14 other sites (min=8, max=17, n=19) and forests linked to an average of 11 other sites (min=8, max=15, n=19). Furthermore, the increased dispersal observed among sites exists in spite of the greater geographic distance among sites of 22 km total distance in Hoosic compared with 14.5 km total distance among sites in the Flint Creek landscape. It may be that dispersal in the Hoosic River landscape is less confined to riparian corridors because the dominant land cover type throughout the watershed is forest - providing a greater variety of pathways for the dispersal of *E. coli* with wildlife. In general, the extent of *E. coli* dispersal in the agrarian landscape seems to be influenced by the degree of forest coverage with higher forest coverage supporting increased dispersal of *E. coli*. Importantly, our data indicate that the transmission of fecal bacteria is not uniform across the agricultural landscape, rather it may be shaped by local landscape features.

Correlations between actual dispersal and hypothetical dispersal models.

Mantel tests were used to estimate the correlation between dispersal rate models and actual (measured) dispersal among sampled sites (Table 2). Statistical significance of model fits was estimated by 100,000 permutations of the test using randomized (by sample site) versions of the actual *E. coli* dispersal data. So, a P-value of 0.05 means that only 5% of randomized data fit the hypothetical model better than the real data did. In the Hoosic landscape, dispersal patterns for all *E. coli* combined had best correlation with models of dispersal via terrestrial, “deer-like” (class LT), model animals (Table 3). Of 29 dispersal models with significant correlation to sequence type mixture data, 28 modeled the main river as an absolute barrier to dispersal. Correlation coefficients were weak but statistically significant (Mantel's R range = 0.17 to 0.18, P < 0.01; Table 3). *E. coli* from group B1A produced stronger fits to the dispersal models (Table 4). Those statistics indicate that models with riparian corridors produced the best fit (Mantel's R = 0.31, P-value < 0.001). While adding other landscape resistance factors associated with model animal behaviors did not improve the model fits, the best fitting of the more complex models all had riparian effects and either assumed "deer-like" vehicles with weak proximity effects or "starling-like" vehicles with strong proximity effects. The B1 and B1A subgroups may be desirable for future modeling for other reasons: a) they have been demonstrated by at least two separate lab groups to have the best phenotypes for survival outside the host (Meric *et al.*, 2013 and P. Bergholz, unpublished data), b) they are the dominant isolates obtained from wild and domestic animals, c) they are closely related to many pathogens of concern, including *E. coli* O157:H7, and d) the two groups B1 and B1A consistently constitute 30-40% of *E. coli* isolates from our environmental samples. No other subgroups exhibited significant fits against dispersal models. In the case of group B2, this may be because this group exhibits poor survival outside the host. This may have limited our ability to observe long-range dispersal of group B2 among sites (Meric *et al.*, 2013 and P. Bergholz, unpublished data). Dispersal was observed to be limited mainly to forested sites in the Flint Creek landscape, and this dynamic was so different from our hypotheses that no significant correlations were observed between actual dispersal and predictions. The result does suggest that riparian forests were very important to the movement of *E. coli* in the Flint Creek landscape, as very little dispersal was detectable outside of the forested sites.

Outcomes and Accomplishments

The observed model correlations were weak, probably because many small scale temporal and spatial contingencies interfere with the signal of dispersal. For example, temperature or precipitation prior to sampling do not necessarily effect dispersal, but may impact our ability to isolate *E. coli* from farm environments. In spite of the fact that model fit coefficients were somewhat weak, there are some consistent patterns in the data that can serve as the basis for more precise models and hypotheses about the movements of enteric bacteria, including pathogens, associated with wild animals. **First**, different landscapes have quantitatively different dispersal dynamics for enteric bacteria, and both the dispersal pathways and persistence of *E. coli* in the environment may be related to the extent of forest cover. **Second**, the best fitting dispersal models contained either absolute or porous barrier effects, which indicates that terrestrial transport, likely by wild animals, is important (as major roadways and open water block dispersal of the bacteria). Dispersal by human activities would be able to easily traverse these barriers and would quickly distort or destroy any signal associated with barriers to terrestrial movement. Dispersal by avian or waterborne pathways would also be expected to eliminate the ability for a river to act as a barrier. **Third**, and perhaps most importantly from the perspective of this project, our data suggest that the role of riparian corridors in dispersal of *E. coli* across a landscape depends on the landscape. In contrast to the Flint Creek area, which is characterized by sparse forestation and where dispersal was detectable almost exclusively in the riparian area, the Hoosic River landscape (which is more forested) exhibited widespread dispersal which correlated with riparian effect models only slightly better than with non-riparian models. This could be the result of selection for differing *E. coli* populations in forests than in agricultural fields, which the Hoosic area would not display as strongly. Alternatively, it could be that riparian corridors are quite a bit more important to dispersal in less-forested landscapes. Overall, this project developed new data indicating that, in the areas sampled in New York, density of forest cover correlates with the percent of *E. coli* positive environmental samples and riparian zones may play a role as dispersal corridor for the transmission of enteric bacteria between and towards produce fields, especially where forest cover is very limited. Specifically, our modelling data indicate that dispersal of *E. coli* follows a pattern that suggests dispersal by terrestrial animals (with our models indicating "large nuisance wildlife species "such as white tail deer and wild pigs as most likely dispersal vectors). Importantly, our project developed the methodology and tools that can be used to perform similar analyses in other landscapes, which will further enhance our understanding of the roles of riparian zones in bacterial dispersal to produce fields.

Summary of Findings and Recommendations

Key findings

- The dispersal dynamics of *E. coli* varied by landscape (with the landscapes studied differing by the proportion of forest land) and forested areas may act as a reservoir of extrahost *E. coli*.
- Different dispersal behaviors were exhibited by different *E. coli* subgroups. Our ability to distinguish these groups was crucial to finding the strongest possible signals of overland dispersal.
- *E. coli* groups D and F were frequently isolated from environmental samples and their populations were structured by extra-host environmental factors. When these groups were included in dispersal analysis of all *E. coli,* the signal of environmental selection on these groups distorted the pattern of dispersal associated with other *E. coli*.
- *E. coli* group B1A was one of the most frequent isolates, and this group exhibited a relatively strong signal of dispersal with some preference for riparian corridors.
- *E. coli* group B1A may be a good target for examining the potential for dispersal of *E. coli* pathogens with wildlife for numerous reasons.
- Riparian forests played a role in the overland movement of *E. coli*. Our dispersal observations indicate that in less forested landscapes, detectable dispersal is largely limited to riparian forests. However, a different set of models may be needed to test the role of riparian zones in less-forested, more agriculturally intensive landscapes.

Recommendations based on these findings

- Our results indicated that future attempts to understand *E. coli* movements on farms might do better if they focus on *E. coli* groups B1, B1A and E which are closely related to the *E. coli* pathogens of greatest concern and exhibit the best correlation to wildlife-based dispersal models.
- Groups D and F are less relevant from a risk perspective (though they do contain some pathogens) and exhibit strong signals of environmental selection, suggesting that dispersal of these groups is harder to track. If understanding dispersal and/or risk of STEC contamination is the goal, it may be better to not include *E. coli* isolates representing groups D and F in the data analyses.
- While the dispersal and persistence of *E. coli* is complex, our analyses indicate that terrestrial wildlife likely played a role in the movements of *E. coli* among produce fields. This suggests that future projects should further evaluate the effects, in *E. coli* dispersal, of barriers that reduce terrestrial wildlife movement.
- The observation that dispersal dynamics of *E. coli* varied by landscape suggests that generalization from studies in a specific landscape to other landscapes may provide potentially misleading information on *E. coli* and pathogen dispersal.

Literature Cited/Bibliography

Avery, S. M., A. Moore & M. L. Hutchison, (2004) Fate of *Escherichia coli* originating from livestock faeces deposited directly onto pasture. *Lett Appl Microbiol* **38**: 355-359.

Bardiau, M., F. Gregoire, A. Muylaert, A. Nahayo, J. N. Duprez, J. Mainil & A. Linden, (2010) Enteropathogenic (EPEC), enterohaemorragic (EHEC) and verotoxigenic (VTEC) *Escherichia coli* in wild cervids. *J Appl Microbiol* **109**: 2214-2222.

Berges, S. A., L. A. S. Moore, T. M. Isenhart & R. C. Schultz, (2010) Bird species diversity in riparian buffers, row crop fields, and grazed pastures within agriculturally dominated watersheds. *Agroforest Syst* **79**: 97-110.

Bergholz, P. W., J. D. Noar & D. H. Buckley, (2011) Environmental patterns are imposed on the population structure of *Escherichia coli* after fecal deposition. *Appl Environ Microbiol* **77**: 211-219.

Borin, M., M. Passoni, M. Thiene & T. Tempesta, (2010) Multiple functions of buffer strips in farming areas. *Eur J Agron* **32**: 103-111.

Byappanahalli, M. N., R. L. Whitman, D. A. Shively, M. J. Sadowsky & S. Ishii, (2006) Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ Microbiol* **8**: 504-513.

Cushman, S. A., K. S. McKelvey, J. Hayden & M. K. Schwartz, (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *Am Nat* **168**: 486-499.

Epps, C. W., J. D. Wehausen, V. C. Bleich, S. G. Torres & J. S. Brashares, (2007) Optimizing dispersal and corridor models using landscape genetics. *J Appl Ecol* **44**: 714-724.

Finder, R. A., J. L. Roseberry & A. Woolf, (1999) Site and landscape conditions at white-tailed deer vehicle collision locations in Illinois. *Landscape Urban Plan* **44**: 77-85.

Henderson, H., (2008) Direct and indirect zoonotic transmission of Shiga toxin-producing *Escherichia coli*. *Javma-J Am Vet Med A* **232**: 848-859.

Islam, M., M. P. Doyle, S. C. Phatak, P. Millner & X. P. Jiang, (2005) Survival of *Escherichia coli* O157 : H7 in soil and on carrots and onions grown in fields treated with contaminated manure composts or irrigation water. *Food Microbiology* **22**: 63-70.

Meric G., E.K. Kemsley, D.Falush, E.J.Saggers, S. Lucchini, (2013) Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. *Environ Microbiol* **15**:487-501.

Oliver, D. M., A. L. Heathwaite, R. D. Fish, D. R. Chadwick, C. J. Hodgson, M. Winter & A. J. Butler, (2009) Scale appropriate modelling of diffuse microbial pollution from agriculture. *Prog Phys Geog* **33**: 358-377.

Sanchez, S., A. Garcia-Sanchez, R. Martinez, J. Blanco, J. E. Blanco, M. Blanco, G. Dahbi, A. Mora, J. H. de Mendoza, J. M. Alonso & J. Rey, (2009) Detection and characterisation of Shiga toxin-producing *Escherichia coli* other than *Escherichia coli* O157:H7 in wild ruminants. *Vet J* **180**: 384-388.

Sanchez, S., R. Martinez, A. Garcia, D. Vidal, J. Blanco, M. Blanco, J. E. Blanco, A. Mora, S. Herrera-Leon, A. Echeita, J. M. Alonso & J. Rey, (2010) Detection and characterisation of O157:H7 and non-O157 Shiga toxin-producing *Escherichia coli* in wild boars. *Vet Microbiol* **143**: 420-423.

Semenov, A. V., E. Franz, L. van Overbeek, A. J. Termorshuizen & A. H. van Bruggen, (2008) Estimating the stability of *Escherichia coli* O157:H7 survival in manure-amended soils with different management histories. *Environ Microbiol* **10**: 1450-1459.

Tenaillon, O., D. Skurnik, B. Picard & E. Denamur, (2010) The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* **8**: 207-217.

van Elsas, J. D., P. Hill, A. Chronakova, M. Grekova, Y. Topalova, D. Elhottova & V. Kristufek, (2007) Survival of genetically marked *Escherichia coli* O157 : H7 in soil as affected by soil microbial community shifts. *Isme Journal* **1**: 204-214.

APPENDICES

Publications and Presentations (required)

Bergholz, P.W., T. K. Chapin, R.C. Pfuntner, L. K. Strawn, and M. Wiedmann. 2012. Quantifying environmental reservoir quality and landscape connectivity for foodborne pathogen transmission to produce fields. 112th General Meeting American Society for Microbiology. June 16-19, 2012, San Francisco, CA

Ryan, G., S. Warchocki, L. Strawn, M. Wiedmann, and P. Bergholz. 2013. Impacts of riparian forests on the prevalence of non-pathogenic *Escherichia coli* contamination in produce fields. International Association of Food Protection Annual Meeting, Charlotte, NC, July 18-31, 2013

Bergholz, P. W., G. T. Ryan, L. K. Strawn, S. Warchocki, and M. Wiedmann. 2013. The potential role of riparian corridors in overland dispersal bacteria among vegetable farms. Ecological Society of America (ESA) Annual Meeting Minneapolis, Minnesota, August 4 - 9, 2013

Wiedmann, M. 2013. The role of riparian corridors in bacteria dispersal to produce farms. Center for Produce Safety (CPS) 2013 Produce Research Symposium, Rochester, NY, June 25 – 26, 2013

Budget Summary (required)

<u>Account balances</u>

University of California – Davis; Agreement #201121642-01 (Cornell account 1438377)

Salaries	$ 67,950.50
Fringe	$ 34,316.19
Supplies	$ 37,825.29
Services	$ 15,912.58
Travel	$ 1,367.40
Indirect Costs	$ 4,573.04
Total Costs	**$161,945.00**

Univ. of California – Davis flowthrough from Produce Marketing Assoc.; Agreement # 201121882-CORNELL (figures below through 3/10/14) (Cornell account 1438378)

Salaries	$ 37,520.62
Fringe	$ 20,621.48
Supplies	$ 45,694.99
Services	$ 23,285.97
Travel	$ 1,572.63
Total Costs	**$128,695.69**

Tables and Figures (optional)

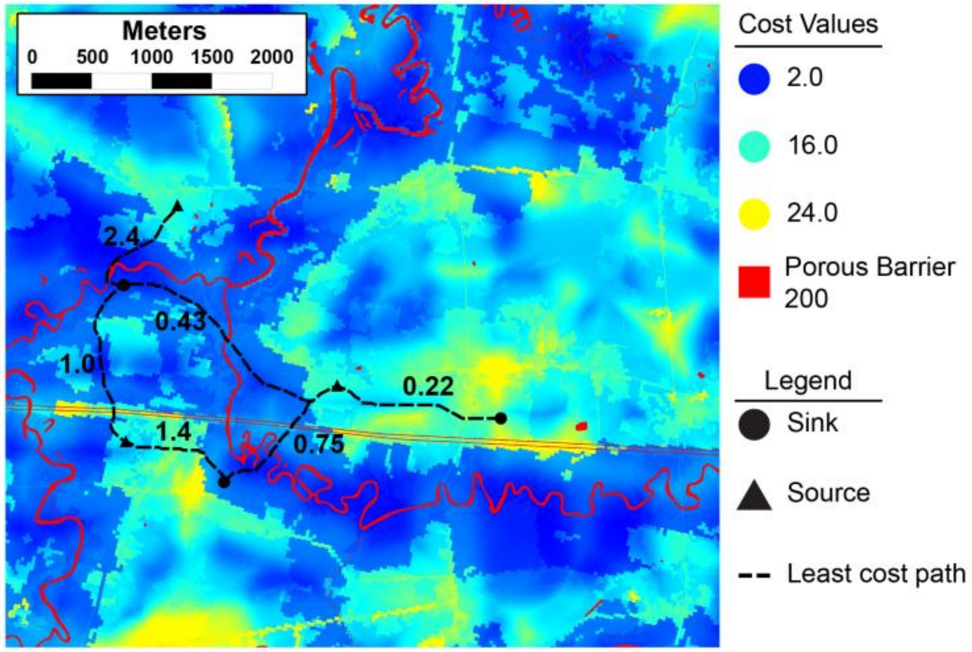

Figure 1. Wild ruminant-based landscape resistance (i.e., dispersal cost) model. This model was calculated as described in the text. The color surface represents the predicted difficulty of dispersal across a 1 m^2 area. Red areas are somewhat porous barriers (interstates and higher order creeks). The riparian zone is defined in this model as land areas adjacent to higher order creeks. A subset of three potential sources (livestock pastures) and three potential sinks (croplands) was selected to represent a connectivity network for display purposes. Dashed lines are the paths of least resistance from sources to sinks. Linkages with movement cost greater than the equivalent of 25 km (15 mi) walking distance are not displayed. Numeric line labels indicate the instantaneous probability of dispersal multiplied

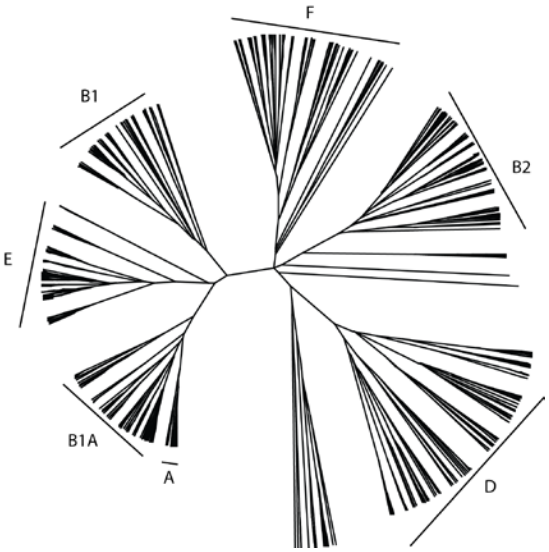

Figure 2. Clonal Phylogeny of the 573 unique seven-gene multilocus sequence types from this study. ECOR group assignments are indicated with labels.

Table 1. Proportional ECOR Group Membership in Soil and Drag Samples by Landcover Type

	Sample Type	A	B1A	B1	B2	D	E	F
Forest	Drag (n=100)	0.110	0.200	0.180	0.060	0.220	0.140	0.090
	Soil (n=215)	0.005	0.163	0.112	0.186	0.242	0.130	0.163
Field	**Drag (n=362)**	**0.011**	**0.235**	**0.157**	**0.224**	**0.166**	**0.133**	**0.075**
	Soil (n=394)	**0.005**	**0.114**	**0.104**	**0.211**	**0.266**	**0.188**	**0.112**

Table 2. Basic Model Characteristics

Dispersal Vehicle[a]	
Class	Description
LT:Large Nuisance Wildlife Species (e.g. "deer" or "feral swine")	Forests, scrublands, grasslands, wooded wetlands and cultivated croplands impose low movement costs, roads impose increased movement costs based on census category
ST:Small mammals (e.g. "shrews", "voles" and "mice")	Light urban development, grassland and scrublands impose low movement costs, roads impose increased movement costs based on census category
LB:Small flocking insectivore/granivores ("starlings")	Open water and heavy urban development are higher cost, all other movement costs are low
MB:Migratory bird flocks (e.g. "Canada goose")	Heavy urban development imposes higher cost, all other movement costs are low.

Riparian Corridor (movement costs are reduced by half)	
Class	Description
Adjacency	All land parcels that overlap a 100 m zone around the main river/creek
Distance	Land within 100 m of the main river/creek
Biological	Land below the 50 yr flood height for the main river/creek and adjacent wetlands
None	No riparian corridor effect

Dispersal Barriers	
Absolute	Major roads and waterways are absolute barriers (movement cost 40,000 per pixel)
Porous	Major roads and waterways are porous barriers (movement cost 200 per pixel)
None	No barrier effect

Proximity Effects (Specifics vary by vehicle)	
Strong	Nearness to high quality habitat substantially reduces movement cost
Weak	Nearness to high quality habitat somewhat reduces movement cost
None	No benefit of proximity to good cover

Attraction (Gravity) Coefficients (Specifics vary by vehicle)	
Habitat Quality	Proximity, interspersion and area of high quality habitat increase the chances that *E. coli* will be deposited
Reduced Habitat Effect	Effect of high quality habitat is reduced by half
Area Independent	As habitat quality model, but area of high quality habitat does not impact the result
None	Attraction does not influence dispersal

Load (Source) Coefficient	
E. coli load estimation	Areas close to forest or pasture-class landcover are higher load.

Simple Landscape Models (Alternatives to Complex Landscape Resistance)	
Isolation by distance	The Euclidean distance among sites is the only limitation
Biological Riparian Only	Only the biological riparian effect (above) influences dispersal
Distance Riparian Only	Only the distance buffered riparian effect (above) influences dispersal
Adjacency Riparian Only	Only the land cover adjacent riparian effect (above) influences dispersal
Absolute Barrier, River Only	Only the listed barrier influences dispersal
Porous Barrier, River Only	Only the listed barrier influences dispersal
Absolute Barrier, River and Major Roads	Only the listed barrier influences dispersal
Porous Barrier, River and Major Roads	Only the listed barrier influences dispersal

[a] Conceptual models are based on published literature and interviews with wildlife ecologists. They should be understood as concepts of bacteria dispersal with wildlife, not definitive predictions.

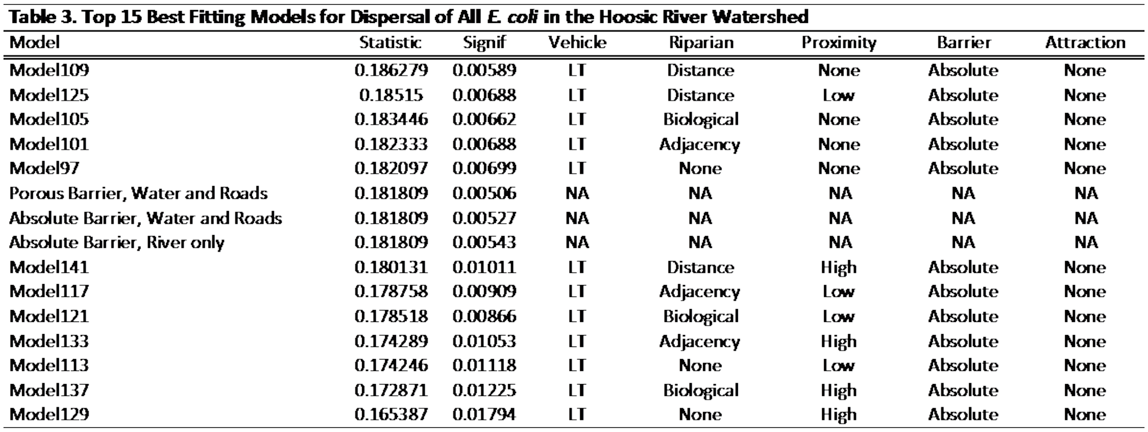

Table 3. Top 15 Best Fitting Models for Dispersal of All *E. coli* in the Hoosic River Watershed

Model	Statistic	Signif	Vehicle	Riparian	Proximity	Barrier	Attraction
Model109	0.186279	0.00589	LT	Distance	None	Absolute	None
Model125	0.18515	0.00688	LT	Distance	Low	Absolute	None
Model105	0.183446	0.00662	LT	Biological	None	Absolute	None
Model101	0.182333	0.00688	LT	Adjacency	None	Absolute	None
Model97	0.182097	0.00699	LT	None	None	Absolute	None
Porous Barrier, Water and Roads	0.181809	0.00506	NA	NA	NA	NA	NA
Absolute Barrier, Water and Roads	0.181809	0.00527	NA	NA	NA	NA	NA
Absolute Barrier, River only	0.181809	0.00543	NA	NA	NA	NA	NA
Model141	0.180131	0.01011	LT	Distance	High	Absolute	None
Model117	0.178758	0.00909	LT	Adjacency	Low	Absolute	None
Model121	0.178518	0.00866	LT	Biological	Low	Absolute	None
Model133	0.174289	0.01053	LT	Adjacency	High	Absolute	None
Model113	0.174246	0.01118	LT	None	Low	Absolute	None
Model137	0.172871	0.01225	LT	Biological	High	Absolute	None
Model129	0.165387	0.01794	LT	None	High	Absolute	None

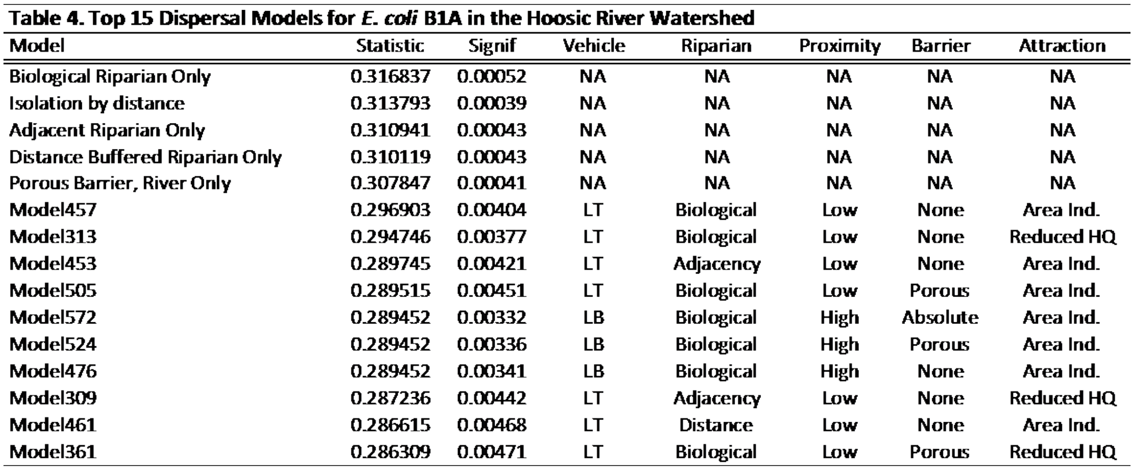

Table 4. Top 15 Dispersal Models for *E. coli* B1A in the Hoosic River Watershed

Model	Statistic	Signif	Vehicle	Riparian	Proximity	Barrier	Attraction
Biological Riparian Only	0.316837	0.00052	NA	NA	NA	NA	NA
Isolation by distance	0.313793	0.00039	NA	NA	NA	NA	NA
Adjacent Riparian Only	0.310941	0.00043	NA	NA	NA	NA	NA
Distance Buffered Riparian Only	0.310119	0.00043	NA	NA	NA	NA	NA
Porous Barrier, River Only	0.307847	0.00041	NA	NA	NA	NA	NA
Model457	0.296903	0.00404	LT	Biological	Low	None	Area Ind.
Model313	0.294746	0.00377	LT	Biological	Low	None	Reduced HQ
Model453	0.289745	0.00421	LT	Adjacency	Low	None	Area Ind.
Model505	0.289515	0.00451	LT	Biological	Low	Porous	Area Ind.
Model572	0.289452	0.00332	LB	Biological	High	Absolute	Area Ind.
Model524	0.289452	0.00336	LB	Biological	High	Porous	Area Ind.
Model476	0.289452	0.00341	LB	Biological	High	None	Area Ind.
Model309	0.287236	0.00442	LT	Adjacency	Low	None	Reduced HQ
Model461	0.286615	0.00468	LT	Distance	Low	None	Area Ind.
Model361	0.286309	0.00471	LT	Biological	Low	Porous	Reduced HQ

Suggestions to CPS (optional)