Evaluation of Virulence Factor Profiling in the Characterization of Veterinary *Escherichia coli*

Donna E. David, Aaron M. Lynne, Jing Han, and Steven L. Foley

Marshfield Clinic Research Foundation, Marshfield, WI 54449

Department of Biological Sciences, Sam Houston State University, Huntsville, TX 77341

Division of Microbiology, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079

Address for correspondence:
Steven L. Foley, Ph.D.
National Center for Toxicological Research
3900 NCTR Road
Jefferson, AR 72079
Phone: (870) 543-7547
steven.foley@fda.hhs.gov

Running Title: Virulence Factor Profiling of Veterinary *E. coli*

Key Words: *Escherichia coli*, virulence factor PCR, BOXA1R PCR, Microbial Source Tracking

Disclaimer: Views presented in this manuscript do not necessarily reflect those of the U.S. Food and Drug Administration.
Escherichia coli have been used as indicator organisms for fecal contamination of water and other environments and are often commensal organisms in healthy animals, yet a number of strains can cause disease in young or immuno-compromised animals. In this study, 281 E. coli from bovine, porcine, chicken, canine, equine, feline and other veterinary sources were analyzed by BOXA1R PCR and by virulence factor profiling of 35 factors to determine whether they had utility in identifying the animal source of the isolates. The results of BOXA1R PCR analysis demonstrated a high degree of diversity, less than half of the isolates fell into one of 27 clusters with at least three isolates (based on 90% similarity). Nearly 60% of these clusters contained isolates from more than one animal source. Conversely, the results of virulence factor profiling demonstrated clustering by animal source. Three clusters, based on discriminant components analysis, named Bovine, Chicken and Porcine were represented by 90% or more of the respective isolates. A fourth group, termed Companion was most diverse, containing at least 84% of isolates from canine, feline, equine and other animal sources. Based on these results, it appears that virulence factor profiling may have utility, helping identify the likely animal host species sources of certain E. coli.
INTRODUCTION

Escherichia coli are common inhabitants of the intestinal tract of many animal species and are used as indicators that other pathogenic bacteria, such as Salmonella, may be present in a particular environment. Animal feces containing pathogenic organisms can enter watersheds through runoff or improper manure handling practices and enter the food supply through fecal contamination, serving as a potential source of human infections (2, 10, 17). Therefore, it is important to find improved ways of identifying sources of contamination to limit exposures to pathogenic microorganisms.

A common limitation of microbial source tracking (MST) methods, such as ribotyping and repetitive element-PCR is that these techniques derive genotypes based on the assumption that specific genetic sequences are present in different locations in the genomes of non-clonal strains. Thus these techniques provide little information beyond separating apparently non-clonal isolates (19). These methods also require a pure culture to be effective, thereby increasing the time to genotype the isolates (19).

An alternative approach to traditional genotyping methods for the initial identification of potential sources of fecal contamination is a virulence factor-based method that in addition to MST could provide information about potential pathogenicity in humans or other animal species. Virulence genes are likely under selective pressure and subject to natural selection processes allowing genes that are important in a specific niche to persist. Additionally, many virulence genes are known to be located on transmissible elements, such as plasmids, which can facilitate their dissemination through the bacteria in a specific environment (15). In poultry for example, E. coli from multiple different serogroups can share common plasmid-associated virulence factors, which
allows for avian pathogenic *E. coli* to be identified among a diverse population of strains (15, 21). Thus, isolates from different environments are likely to have different cohorts of genes that can potentially be exploited to distinguish bacteria from different sources.

There have been a large number of studies characterizing virulence factors from *E. coli* from a particular animal species (20-22, 28, 30). In each of these studies, the goals were to identify which genes are important for virulence in the respective animal species. Based on the results of these studies, it appeared that for the different animal species there were different sets of virulence factors that likely played a role in disease. Typically *E. coli*-associated disease occurs in very young or immuno-compromised animals, without being a major problem in adult animals, even though a percentage of adult animals harbored *E. coli* with identified virulence factors (20, 22, 30). Because a proportion of virulence factors persist in *E. coli* from adult animals and different virulence factors appear to be important in different animal species, they may aid in the determination of the animal source of bacteria. Therefore, this manuscript describes the potential utility of virulence factor profiling to distinguish among *E. coli* from different animal sources, which may have utility as part of a MST scheme.

**MATERIALS AND METHODS**

**Bacterial Strains.** In this study, 281 *E. coli* isolated from bovine (n=57), canine (n=46), chicken (n=63), equine (n=9), feline (n=8), swine (n=85), and other (zoo, sheep, rodents; n=13) clinical/diagnostic samples from Arkansas, Georgia, Missouri, North Dakota, and Oklahoma were evaluated (Figure 1).
BOXA1R PCR. Isolates were genotyped using BOXA1R PCR as described by Dombek (9). The banding pattern results were analyzed with BioNumerics software (Applied-Maths, Kortrijk, Belgium) using Dice coefficients with 1% tolerance and 1.5% optimization coefficients to determine isolate similarity and generate a dendrogram (Figure 1).

Virulence Factor Profiling. For virulence factor profiling using PCR, supernantant from boiled preps was combined with specific virulence factor primers, H₂O and 2X Master Mix (Promega, Madison, WI) and amplified using optimized conditions for each primer set (Supplemental Table 1). The PCR was carried out as previously described using the optimized annealing temperature shown in Supplemental Table 1 (18). PCR products were analyzed using 2% E-gels (Invitrogen, Carlsbad, CA). Positive control strains were included with each set of reactions and each set was repeated to confirm the results. PCR data were imported into BioNumerics and analyzed based on the presence or absence of virulence genes.

Statistical Analysis. Data was summarized as percent positive by animal host (Supplemental Table 2) and analyzed with BioNumerics using discriminant analysis with variance, to determine whether isolates from the defined animal source groups could be separated from each other based upon their virulence gene profiles (16). Additionally, Jackknife analysis was carried out using BioNumerics on the results of both BOXA1R PCR and virulence factor profiling using the maximum similarities setting to determine
RESULTS AND DISCUSSION:

The results of BOXA1R PCR analysis demonstrated a high degree of diversity. Based on 90% similarity, only 125 of 281 (44%) isolates fell into one of 27 clusters with at least three isolates. Of those clusters, 16 (59%) contained isolates from more than one animal source, indicating a lack of source and common BOXA1R PCR profile agreement in most instances. To gain a better understanding of how well the method worked to separate isolates by their particular source, Jackknife analysis was carried out. With the Jackknife analysis, a series of analyses are done in which each individual isolate profile is removed from the phylogenetic analysis, the analysis is repeated without the isolate and the isolate profile is compared to the results of the analysis of the different species groups. The analysis determined whether the removed isolate was correctly grouped with its original animal source group or misclassified as another animal source. The process was repeated for all of the isolates in the original analysis and a cumulative percentage of classification results were determined. The results of the Jackknife analysis for the BOXA1R PCR are shown in Table 1. Overall, the correct classification percentage ranges from 0% for feline isolates to 81.2% for porcine isolates. The feline isolates were most often classified as canine isolates (62.5% of the time), which is likely due to the similarity of the overall similarity of the isolates from the two sources coupled with the much larger sample size of canine isolates.
Overall, the results of virulence factor genotyping appeared to demonstrate clustering based on the original animal source. Distinct clusters with isolates from predominant sources were identified (Figure 2). Three clusters, referred to as Bovine, Chicken and Porcine, were represented by 90% or more of isolates from their respective sources. The Companion group was most diverse, containing at least 84% of isolates from canine and feline sources as well as those from equine and “other” animal sources. The result of the Jackknife analysis for the virulence factor genotyping is shown in Table 2. Overall, the correct classification percentage ranged from 37.5% for feline isolates to 96.5% for porcine isolates. Again, the primary misclassification for the feline isolates was as canine isolates. When the overall virulence factor profiling results were compared to the BOXA1R PCR results, virulence factor typing correctly classified the isolates by their respective sources more often than did BOXA1R PCR (Tables 1 and 2).

This initial study focused on determining whether virulence factor-based methods would have potential utility in a MST scheme for veterinary E. coli. Because disease pathogenesis varies among animal species, E. coli strains originating from different sources may acquire a unique cohort of genes for survival (5). E. coli infections in livestock and poultry are generally associated with immuno-naïve young animals in cattle and swine and with pre-existing disease or immuno-compromised animals due to environmental stress in poultry (1, 3). Thus, potential pathogens may persist in immuno-competent animals and in production environments even in the absence of disease. In poultry for example, nearly 20% of E. coli isolates from apparently healthy birds contained genes associated with avian E. coli virulence (21). Likewise in swine, Rosengren et al (2009) found that in a study of 151 isolates from healthy animals during
growth and finishing stages that over half contained at least 2 factors associated with porcine virulence (22). Additionally, virulence factors associated with disease in calves can be detected in healthy adult cows (6, 25, 29, 30). Based on these studies of E. coli from different healthy animal species from various geographical locations, it appeared that virulence factors were found in healthy animals regardless of geographical origin. Therefore, the selection of isolates from clinical/diagnostic sources likely provided a good starting point for the evaluation and determination of virulence factors that would have the most potential in a source tracking scheme.

Virulence associated with E. coli from different animal species likely varies due to differences in pathogenesis and host systems. Poultry disease is typified by septicemia following respiratory exposure to pathogenic E. coli, while mammalian infections are generally associated with diarrhea or extra-intestinal infections following gastrointestinal colonization. Thus, virulence factors associated with poultry disease include those for respiratory colonization, serum resistance (iss) and iron acquisition (iut and iuc), while those associated with mammalian hosts include factors for attachment to the gut of the different animal species (K88, K99, F18, F41) and production of diarrheal toxins (estA and estB) (1, 20). Indeed, iss along with tsh and cvaC were preferentially found in chicken isolates, while K99, F41, espP, and afuE_8 were associated with bovine isolates and K88, F18 and estB with swine isolates. There were a few individual factors that were distinctly associated with companion animals, equine or other isolates; it was not until the cumulative profiles were examined that these isolates (Companion cluster) largely separated from the food animal isolates (Figure 2). Therefore, future work will be necessary to identify whether there are specific differences associated with the
Companion animal cluster isolates, and if there are differences, how we will distinguish them from isolates of other various sources. Certain factors do not appear overly useful in a MST scheme, for example 987P and katP were not detected in any of the isolates, while fimA and ompT were widely distributed among the different sources. There was also an attempt to minimize the potential impact of spatial and temporal differences in the results. For each of the animal species, there were isolates collected in Arkansas, and there were cattle, swine and chicken isolates from North Dakota. Overall, these isolates continued to cluster with the isolates from the other states based on source, rather than geography (data not shown).

It is important to note that there are potential limitations in this study, including the number of isolates used in the study was somewhat limited. Because of available resources, the isolates were chosen based on availability of clinical samples. As described above, the focus of this work was on veterinary clinical isolates because they are more likely to have divergent virulence profiles, yet a percentage of these “virulent” organisms may remain in the healthy population as commensals. The number of clinical isolates available was especially limited from equine and feline sources, for which there were fewer than 10 isolates each. This limited our ability to discriminate them from the canine isolates, as indicated in the Companion cluster of Figure 2. Future studies will need to be done to collect a larger set of isolates from these sources to better identify discriminatory factors that will separate the isolates from different sources in the Companion cluster. Likewise, additional isolates, such as those from humans, turkeys and wildlife sources will need to be tested to determine whether the current panel is effective in discriminating these isolates based on their sources, or whether additional
host specific factors need to be identified. If additional factors are required, they could potentially be genes that are from other enteric organisms. For human fecal contamination, organisms like *Bacteroides* spp. and *Bifidobacterium* spp. have been evaluated as marker organisms for human fecal contamination (28); thus specific gene targets from these organisms could potentially be utilized to identify human fecal contamination.

A potential advantage of virulence factor PCR genotyping approach is that it could potentially be deployed in a culture independent fashion, since a pure culture is not necessarily required to identify source-associated virulence signatures. The PCR-based approach will likely allow for the detection of virulence factors in a population of *E. coli*. Numerous studies have found *E. coli* isolates from healthy animals contain virulence genes, however the prevalence is typically lower than in diseased animals (20, 21, 30). Therefore the ability to screen populations of *E. coli* for virulence genes will likely be beneficial when strains containing the source-associated virulence factors are present in the animal population at a relatively low frequency. When collecting samples for genotyping, if high bacterial loads are detected in the water, the bacteria could be concentrated and their DNA extracted for use as template DNA for PCR screening of virulence genes, without the need for enrichment. Conversely, if the levels are too low for direct sample analysis, the concentrated bacteria could be inoculated into a non-specific enrichment broth to allow the bacteria numbers to increase prior to DNA isolation and PCR testing.

The PCR results could potentially be interpreted in several different ways to identify the likely animal source. These include the use of a library matching algorithm.
in which the composite virulence factor profile is compared to an existing library of profiles from *E. coli* of known sources to identify its likely origin (23). Alternatively, a decision tree can be developed, such that isolates from different sources are separated in a stepwise approach based on the presence or absence of specific virulence factors (24). For example, *iss* could be used to separate chicken isolates from the majority of isolates from mammalian sources, then *afaE_8* could separate many cattle isolates from the remaining mammalian isolates and so on until an optimum decision tree is built. The ability to rapidly detect potentially host specific signatures could allow investigators to more rapidly focus on potential sources of contamination associated with a particular animal source prior to the isolation of the contaminating *E. coli*. Then as investigators hone in on the source of contamination, they can then use more definitive genotyping tools, such as pulsed field gel electrophoresis, ribotyping or sequencing-based methods, which are more time consuming, expensive and require a pure culture, to link the contamination source to the contamination event.

Overall, associating bacteria causing contamination to a specific animal source should allow regulators to focus their search to farms producing that particular animal species and mitigate circumstances which led to the contamination. Based on the results of this study, it appears that the use of a virulence factor profiling may be a useful part of a MST scheme for *E. coli* of veterinary origin. The identification of a cohort of genes associated with a particular animal source would potentially allow investigators to narrow the search for the contamination source allowing for a more efficient use of resources.
ACKNOWLEDGEMENTS

The authors would like to thank Dr. David White and the Arkansas Livestock and Poultry Commission for providing isolates for the study, Drs. Rajesh Nayak and Huizhong Chen for their assistance in reviewing the manuscript, and the Marshfield Clinic Research Foundation and U.S. Food and Drug Administration for financial support of the research. Dr. Jing Han is supported through Oak Ridge Institute for Science and Education.
REFERENCES


Escherichia coli strains from patients with urosepsis in relation to phylogeny and 

2002. Location of increased serum survival gene and selected virulence traits on a 
conjugative R plasmid in an avian Escherichia coli isolate. Avian Dis. 46:342-
352.


1999. Molecular typing methods to investigate transmission of Escherichia coli 

Characterization of antimicrobial resistance in Salmonella enterica serotype 
Heidelberg isolated from food animals. Foodborne Pathog Dis 6:207-15.


traits among Escherichia coli isolates from apparently healthy birds and birds 


FIGURE LEGENDS

Figure 1: Results of BOXA1R and virulence factor PCR for 281 *E. coli* isolates used in this study. The dendrogram at the left is based on the BOXA1R results. The red bars in the gel images indicate the bands present in each of the gels. For virulence factor PCR results, a blue box indicates the detection of the virulence gene and golden box the absence of a detectable gene. The columns to the right of the virulence genes indicate the animal source, State isolated from and year isolated, respectively. For the animal source, isolates marked in green are porcine, red are bovine, yellow are chicken, light blue are canine, pink are feline, brown are equine and purple are from other animal sources, respectively. For the States, isolates marked in green are from Arkansas, red from Georgia, dark blue from Missouri, yellow from North Dakota and turquoise from Oklahoma. For the year of isolation, those marked in green are from 1992, red from 1996, pink from 1997, yellow from 1998, turquoise from 1999, and purple from 2004.

Figure 2: The cumulative virulence factor PCR data were also analyzed by discriminant component analysis with variance in BioNumerics to determine grouping of isolates based on their cumulative virulence gene profiles. The isolates are displayed by a symbol corresponding to animal source group: Bovine (★), Chicken (⊙), Swine (●), Canine (II), Feline (☑), Equine (☑) and Other (☑). Four major clusters were identified, delineated by ovals on the figure and defined by the predominant source (Bovine, Chicken, Swine and Companion). Adjacent to the ovals are the numbers of isolates from the different animal species present in the respective clusters.
Chicken: 61
Canine: 3
Porcine: 1
Bovine: 1
Other: 2

Bovine: 47
Canine: 1
Porcine: 6
Equine: 8
Feline: 7
Bovine: 6
Porcine: 75
Equine: 1
Other: 11
Table 1: Jackknife similarity results for BOXA1R PCR

<table>
<thead>
<tr>
<th></th>
<th>Porcine</th>
<th>Bovine</th>
<th>Chicken</th>
<th>Equine</th>
<th>Canine</th>
<th>Other</th>
<th>Feline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porcine</td>
<td>81.2</td>
<td>3.5</td>
<td>11.1</td>
<td>33.3</td>
<td>19.6</td>
<td>30.8</td>
<td>25</td>
</tr>
<tr>
<td>Bovine</td>
<td>4.7</td>
<td><strong>75.4</strong></td>
<td>6.4</td>
<td>11.1</td>
<td>10.9</td>
<td>15.4</td>
<td>12.5</td>
</tr>
<tr>
<td>Chicken</td>
<td>9.4</td>
<td>14.0</td>
<td><strong>71.4</strong></td>
<td>22.2</td>
<td>10.9</td>
<td>7.7</td>
<td>0</td>
</tr>
<tr>
<td>Equine</td>
<td>0</td>
<td>0</td>
<td>3.2</td>
<td><strong>22.2</strong></td>
<td>4.4</td>
<td>7.7</td>
<td>0</td>
</tr>
<tr>
<td>Canine</td>
<td>2.4</td>
<td>5.3</td>
<td>6.4</td>
<td>11.1</td>
<td><strong>39.1</strong></td>
<td>7.7</td>
<td>62.5</td>
</tr>
<tr>
<td>Other</td>
<td>2.4</td>
<td>1.8</td>
<td>0</td>
<td>0</td>
<td>2.2</td>
<td><strong>30.8</strong></td>
<td>0</td>
</tr>
<tr>
<td>Feline</td>
<td>0</td>
<td>0</td>
<td>1.6</td>
<td>0</td>
<td>13.0</td>
<td>0</td>
<td><strong>0</strong></td>
</tr>
</tbody>
</table>
Table 2: Jackknife similarity results for Virulence Factor Profiling

<table>
<thead>
<tr>
<th></th>
<th>Porcine</th>
<th>Bovine</th>
<th>Chicken</th>
<th>Equine</th>
<th>Canine</th>
<th>Other</th>
<th>Feline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porcine</td>
<td>96.5</td>
<td>0</td>
<td>0</td>
<td>11.1</td>
<td>2.2</td>
<td>7.7</td>
<td>12.5</td>
</tr>
<tr>
<td>Bovine</td>
<td>1.2</td>
<td>89.5</td>
<td>0</td>
<td>22.2</td>
<td>4.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chicken</td>
<td>2.4</td>
<td>3.5</td>
<td>95.2</td>
<td>0</td>
<td>4.4</td>
<td>15.4</td>
<td>0</td>
</tr>
<tr>
<td>Equine</td>
<td>0</td>
<td>3.5</td>
<td>44.4</td>
<td>2.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Canine</td>
<td>0</td>
<td>3.5</td>
<td>4.8</td>
<td>22.2</td>
<td>82.6</td>
<td>15.4</td>
<td>37.5</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.2</td>
<td>61.5</td>
<td>12.5</td>
</tr>
<tr>
<td>Feline</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.2</td>
<td>0</td>
<td>37.5</td>
</tr>
</tbody>
</table>