APPENDIX C: COACHELLA COLIFORM DNA ANALYSIS SOURCE REPORT

A SUMMARY OF

FECAL CONTAMINATION SOURCE TRACKING BY RIBOTYPE

FINGERPRINTS OF ENVIRONMENTAL *E. COLI* FROM THE

COACHELLA VALLEY STORMWATER CHANNEL

Kitts, C., A. Shaffner, M. Samadpour, and I. Reyburn. 2004. Fecal Contamination Source Tracking by Ribotype Fingerprints of Environmental E. Coli from the Coachella Valley Stormwater Channel. Final Report. State Water Resources Control Board Contract Agreement # 02-118-257-1

EXECUTIVE SUMMARY

Because of high coliform counts in the Coachella Valley Stormwater Channel (CVSC), this study was undertaken to determine the distribution of fecal contamination sources and to assist in the formulation of a total maximum daily load plan for the area. Two hundred water samples were collected from three sites along the CVSC over a period of four months, from October 2003 through March 2004. These samples were sent to Dr. Mansour Samadpour's Institute for Environmental Health (IEH) in Seattle, Washington for isolation of *E. coli* followed by ribotype fingerprinting of the isolated bacterial strains. Over five hundred strains of *E. coli* were isolated, fingerprinted and their ribotypes compared to those in the IEH source library. Only 6% of the *E. coli* strains isolated in this study did not match fingerprints in the IEH source library. The two dominant sources of *E. coli* in the study were avian (40%), human (25%) and rodents plus other wild mammals (25%). Livestock sources accounted for less than 3% of the *E. coli* across the entire study, with a statistically higher proportion (5%) at Site 3, the most rural sampling site. The total contribution from human controlled sources (humans, livestock and domestic animals) across the entire study was 29%. Human sources were at a maximum of 29% at Site 2, down stream of the town of Coachella. Domestic animal sources accounted for less than 2% of the *E. coli* across the entire study, with a significantly higher proportion (5%) at Site 2. When the data were analyzed by sampling month, only livestock sources showed a significantly higher contribution (10%) in March. Significant differences in source contribution by site and sampling month may be artifacts of low number of strains isolated in this study (only 539 across three sites and five months). Fecal coliform counts were significantly higher at Site 1 and significantly higher at all three sites in January. Analysis of ribotype distributions across sampling sites indicated that avian and rodent *E. coli* contributions came consistently from the same or very similar host animals; although whether this is on an individual or a species level remains unclear.

INTRODUCTION

Because a TMDL plan was mandated for the area and previous studies showed that the fecal coliform counts were consistently exceeding water quality objectives, this study was undertaken to determine the distribution of fecal sources in the Coachella Valley Stormwater Channel (CVSC). Ribotyping of *E. coli* strains isolated from the CVSC was the method chosen for source determination. Dr. Mansour Samadpour's Institute for Environmental Health (IEH) was chosen as the subcontractor for this work since the IEH maintains a ribotype library of over 100,000 *E. coli* strains from known fecal sources. The size of the IEH library ensures that a minimal number of *E. coli* strains isolated in the study will not match an identified source. Although the IEH uses a direct match protocol for assigning sources to *E. coli* strains, the ribotype method has also been evaluated using a statistical approach (Parveen et. al. 2000). The statistical method was verified at over 84% accurate with a very limited library. The advantage of the direct match method employed by IEH is that poor matches are discarded as unknown.

Literature Validation of *E. coli* as an Indicator Organism

This study relies upon the fingerprinting of *E. coli* strains as indicators for determining the sources of fecal contamination to the CVSC. Total coliforms and fecal coliforms have been the indicators traditionally used for bacterial water quality monitoring. As more data on the efficacy of these traditional indicators is amassed, their suitability is being questioned. In a recent review Leclerc et al. (2001) question the use of both total and fecal coliforms as indicators of fecal contamination because of the number of bacterial species that meet the culture requirements but are not of intestinal origin and grow commonly in the external environment. For example, many species of *Klebsiella* and *Citrobacter* meet all the functional criteria to be counted as fecal coliforms and yet have been commonly isolated from a variety of non-intestinal environments and shown to be indigenous to these environments. In contrast, *E. coli* is a permanent member of the intestinal microflora and is rarely if ever found growing in the external environment. Although several recent papers point out that *E. coli* will grow in the environment under special circumstances (Gauthier and Archibald 2001, Whitman et al. 2003, Solo-Gabrielle et al. 2000), it is still accepted as the best indicator organism to date because it is more exclusively intestinal in origin (Lang et al. 1999, Leclerc et al. 2001), it is a better predictor of the incidence of disease (Moe et al. 2001) and its decay in the environment better emulates some of the more prevalent pathogens of fecal origin (McLellan et al. 2001). As more work with specific pathogenic organisms is reported it has become clear that neither fecal coliforms nor *E. coli* are good indicators environmental contamination with human viruses and encysted parasites like *Giardia* and *Cryptosporidia* (Leclerc et al. 2001). However, *E. coli* is probably the best indicator available for pathogenic enterobacteria and as such remains a useful tool for water quality monitoring.

Summary of Recent Bacterial Monitoring

Regional Board staff collected bacteria, nitrate, and ammonia data for eight consecutive months beginning in February 2003. Although fecal coliforms and *E. coli* counts varied over the collection period, the general conclusion was that the entire length (approximately 16 miles) of the CVSC exceeds the Regional Board's Water Quality Objectives for bacteria to protect beneficial uses and that there are multiple sources of contamination.

Sampling Sites

Sampling Site 1 is located where Avenue 50 in the City of Coachella crosses the CVSC. Site 2 is located at the southern end of the City of Coachella, just upstream of the Airport Boulevard overpass. Sampling Sites 1 and 2 represent water influenced by urban

runoff, wastewater treatment facility discharges, and irrigated agriculture drainage. Site 3 is located where Avenue 66 crosses the storm channel west of the Town of Mecca and represents irrigated agriculture drainage but also includes urban runoff and potentially failing/leaking on-site sewage treatment facility discharge (Figure 1).

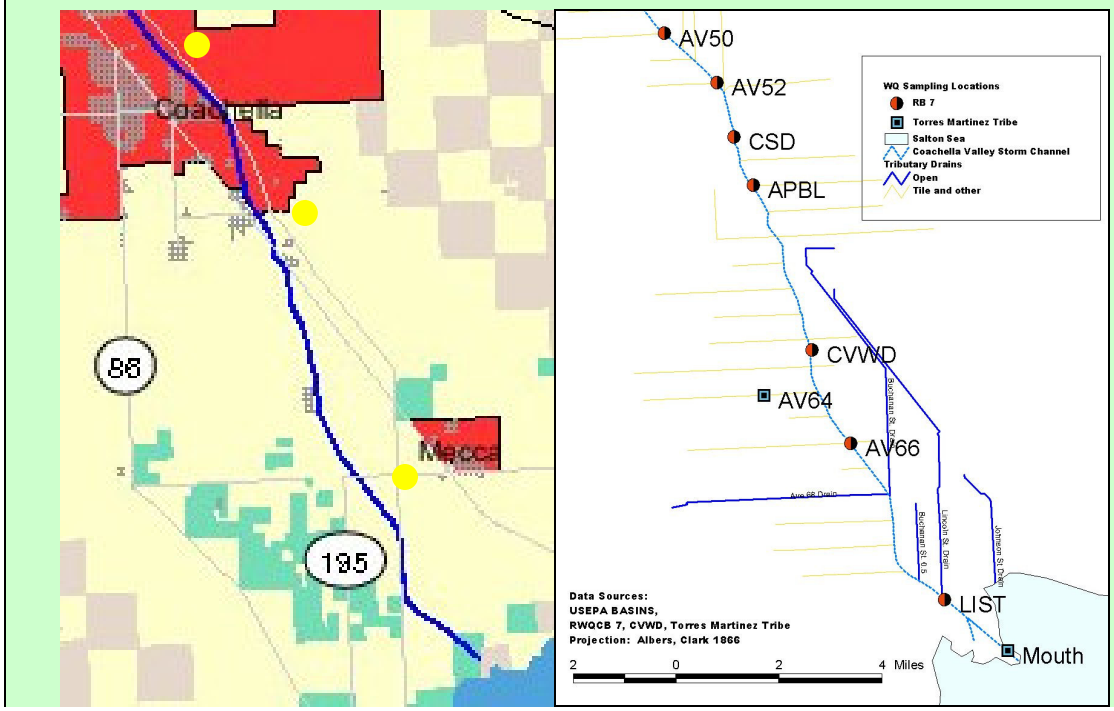

Figure 1. Maps of the sampling area — the Coachella Valley. The left panel shows land use while the right panel shows tributary drains. The CVSC is represented by a blue line in both panels. Site 1 (AV50), Site 2 (APBL) and Site 3 (AV66) are indicated by yellow circles in the left panel.

Possible Sources of Bacteria

Three wastewater treatment facilities and one fish culture facility are permitted point sources for fecal coliforms (and presumably *E. coli*) discharging into the CVSC. Non-point sources should reflect the land use in the area. Most of the land in the drainage area is wild desert shrub-land (57%) and very little is residential/industrial (6%) so the majority of *E. coli* sources are expected to be wild animals and birds. Since the sampling in this study was undertaken during months of high bird populations (over-wintering migratory birds) it is expected that birds will be a large source of *E. coli*. Failing septic tanks are another possible non-point source for *E. coli* in the CVSC and combined with the wastewater treatment discharge this makes humans likely to be another large source of *E. coli*.

Sampling Plan

Sampling took place over a six-month period from October 2003 to March 2004. Replicates were collected to provide a total of 200 total samples (Table 1). It was

anticipated that IEH would isolate a minimum of two *E. coli* strains per sample for a minimum of 400 strains to be fingerprinted in the study. There was some variation from the original sampling plan due to uncertainty in funding that resulted in a stop-work after the first week of sampling in November. The funding issue was resolved in late November and sampling resumed in the first week of December, creating an offset in the sampling schedule.

Table 1. Sampling for this study.

Month 2003/04	Site	Samples Week 1	Samples Week 2	Samples Week 3	Samples Week 4	Samples Week 5	Samples / Month / Site	Monthly Samples
October	1	2	3	2	3		10	40
	2	4	3	4	4		15	
	3	4	3	4	4		15	
November	1	2					2	10
	2	4					4	
	3	4					4	
December	1	3	2	3			8	30
	2	4	4	3			11	
	3	4	4	3			11	
January	1	2	2	3	3	3	13	49
	2	4	4	3	4	3	18	
	3	4	4	3	4	3	18	
February	1	2	3	2	3		10	42
	2	4	4	4	4		16	
	3	4	4	4	4		16	
March	1	2	2	3			7	29
	2	4	4	3			11	
	3	4	4	3			11	
Grand Total	1						50	200 Samples
	2						75	
	3						75	

RESULTS AND DISCUSSION

A total of 200 samples were processed by IEH. Unfortunately, four samples taken in the first week of February lost their labels in transport to IEH. Two were from

Site 1 and two from Site 2 so a determination could not be made for the origin of each sample. Fecal coliforms were isolated using the membrane filtration method (Table 2). On average, membrane filter (MF) fecal coliform counts were significantly higher at Site 1 (ANOVA of $\log_{10}$ transformed data, $p=0.032$) and significantly higher during the month of January ($p<0.001$).

Table 2. Membrane filter fecal coliform counts (per 100 mL) in Coachella samples.

Site	October	November	December	January	February	March	Average by Site
#1	185	303	257	1716	1100	530	615
#2	203	519	212	1052		107	396
#3	244	244	201	819	150	503	372
LFO[a]		78					78
Average by Month	214	313	220	1182	467	383	438

[a] Label Fell Off, 2 samples from Site 1 and 2 samples from Site 2.

Candidate bacterial colonies were confirmed as *E. coli* by growth on MacConkey agar and by biochemical tests with the API20E kit. Confirmed *E. coli* strains were catalogued and DNA was extracted to produce ribotype fingerprints. IEH provided Cal Poly with the ribotypes and library matches to fecal sources for the 539 strains of *E. coli* isolated in this study (Table 3). The complete data set is attached in Appendix A. A total of 162 strains were isolated from Site 1, 167 from Site 2 and 202 strains from Site 3. The number of strains isolated was highest in January and lowest in November and March (Table 3). This was probably due to changes in fecal coliform counts in the samples as well as the number of samples collected in each month.

Table 3. Distribution *E. coli* strains isolated by month.

Month in 2003	Site	Week 1	Week 2	Week 3	Week 4	Week 5	Month Total
	1	10	15	11	2		38
October	**2**	22	7		3		32
	3	18	16		11		45
	1	4					4
November	**2**						
	3	6					6
	1	30	4	9			43
December	**2**	17	12	9			38
	3	7	12	9			28
	1		6	9	9	10	34
January	**2**	5	12	8	12	9	46
	3	9	12	8	12	10	51
	1	4	9	6	10		29

February	**2**	3	13	13	5	34
	3	12	9	8	16	45
	LFO[a]	8				8
	1		5	9		14
March	**2**	1	8	8		17
	3	11	12	4		27
Grand Total						**539**

[a] Label Fell Off, 2 samples from Site 1 and 2 samples from Site 2.

Determination of Fecal Sources

Ribotypes from the 531 *E. coli* strains in the Coachella Valley samples matched to 20 different sources in the IEH library (Table 4). The four samples that lost their labels produced a total of eight *E. coli* strains that were not included in these analyses. A total of 33strains (6.2%) did not produce ribotypes that matched anything in the IEH source library. This is an excellent result that may reflect a lower diversity of sources at the Coachella Valley site. To facilitate statistical analysis, the 20 sources were placed into six groups.

Table 4. IEH library matches for sources of *E. coli* found in this study. Row headers (bold) are the groupings used for later analyses. Column numbers are either the total number of strains isolated in a category or the percent of the total for a site.

Group	Source	All Sites	**(%)**	**Site 1**	**(%)**	**Site 2**	**(%)**	**Site 3**	**(%)**
Avian	*avian*	207	39.0	62	38.3	58	34.7	87	43.1
	duck	5	0.9	0	0.0	2	1.2	3	1.5
	waterfowl	1	0.2	0	0.0	0	0.0	1	0.5
Domestic	*dog*	7	1.3	0	0.0	6	3.6	1	0.5
	feline	3	0.6	2	1.2	1	0.6	0	0.0
Human	*human*	106	20.0	33	20.4	41	24.6	32	15.8
	sewage	20	3.8	5	3.1	8	4.8	7	3.5
	WW effluent	3	0.6	3	1.9	0	0.0	0	0.0
	WWTP sludge	1	0.2	0	0.0	0	0.0	1	0.5
Livestock	*bovine*	9	1.7	0	0.0	3	1.8	6	3.0
	horse	4	0.8	0	0.0	0	0.0	4	2.0
	sheep	1	0.2	0	0.0	0	0.0	1	0.5
Rodent	*muskrat*	2	0.4	0	0.0	2	1.2	0	0.0
	rabbit	2	0.4	1	0.6	0	0.0	1	0.5
	rodent	73	13.7	25	15.4	20	12.0	28	13.9
	squirrel	1	0.2	0	0.0	0	0.0	1	0.5

Wild Mammal	*canine*	41	7.7	11	6.8	9	5.4	21	10.4
	deer	2	0.4	0	0.0	0	0.0	2	1.0
	deer/elk	1	0.2	0	0.0	0	0.0	1	0.5
	raccoon	9	1.7	4	2.5	2	1.2	3	1.5
Unknown	*no match*	33	6.2	16	9.9	15	9.0	2	1.0
	Site Total	531		162		167		202	

The dominant group of fecal sources in the study was clearly avian with an overall contribution of 213 strains (40.1%). The next most common source group was human (including sewage, wastewater effluent and wastewater treatment plant sources) with 130 strains (24.6%). When rodent (including muskrat, rabbit and squirrel) and wild mammal (canine, deer, elk and raccoon) sources were added together they contributed a total of 131 strains (24.7%). Canine sources could belong to wild (coyotes) or domestic (dog) canines and were arbitrarily grouped with wild mammals based on the rural nature of the area. Rodent sources alone produced 78 strains (14.7%). Livestock sources (including horse, bovine and sheep) contributed 14 strains (2.7%). Domestic mammals (dog and feline) contributed 10 strains (1.9%).

Distribution of Sources by Site

When the six source groupings, plus unknowns, were analyzed for site distribution, a statistical difference was detected in the composition of sources at each site (Pearson Chi-Square, $p<0.001$). This was mostly due to changes in the contributions from livestock at Site 3 (Pearson Chi-Square, $p=0.001$), and domestic animal sources at site 2 ($p=0.031$). Differences in the avian ($p=0.175$), human ($p=0.099$), rodent ($p=0.758$), and wild mammal ($p=0.088$) contributions by site were not significant. There was also a significantly lower number of unknown strains at Site 3 ($p<0.001$). Because these differences were only significant in the low abundance source groups (< 6% total contribution) their statistical significance might be an artifact of the low number of strains isolated for this study. IEH recommends that a minimum of 200 strains be isolated per site to obtain an accurate estimate of source contribution and this was not achieved for Sites 1 and 2.

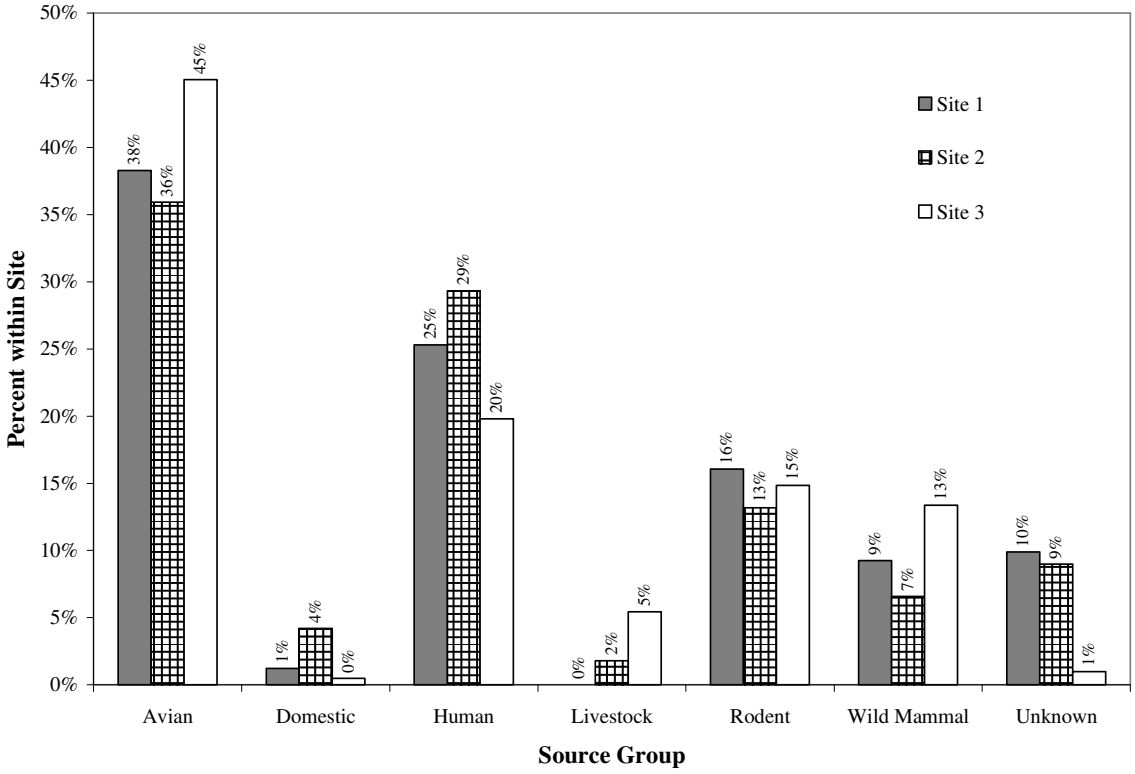

Figure 2. Source group distribution by site. Each bar represents the number of *E. coli* from a specified source isolated at a particular site as a percentage of the total *E. coli* isolated at that site.

Distribution of Sources by Sampling Month

When the six source groupings, plus unknowns, were analyzed for distribution by month of collection (Figure 3), a statistical difference was detected in the source composition (Pearson Chi-Square, p=0.006). This was mostly due to a significant increase in livestock contributions for March (Pearson Chi-Square, p=0.004) and a significantly larger unknown contribution in December (p=0.007). All other groups did not have significantly different contributions across sampling month. The total number of strains isolated declined in March (Figure3, Table 3) due in part to an offset in the sampling schedule created by the stop-work in November. The difference in livestock contributions for March may be an artifact because of the lower number of strains isolated in March and the overall low number of strains isolated from livestock sources.

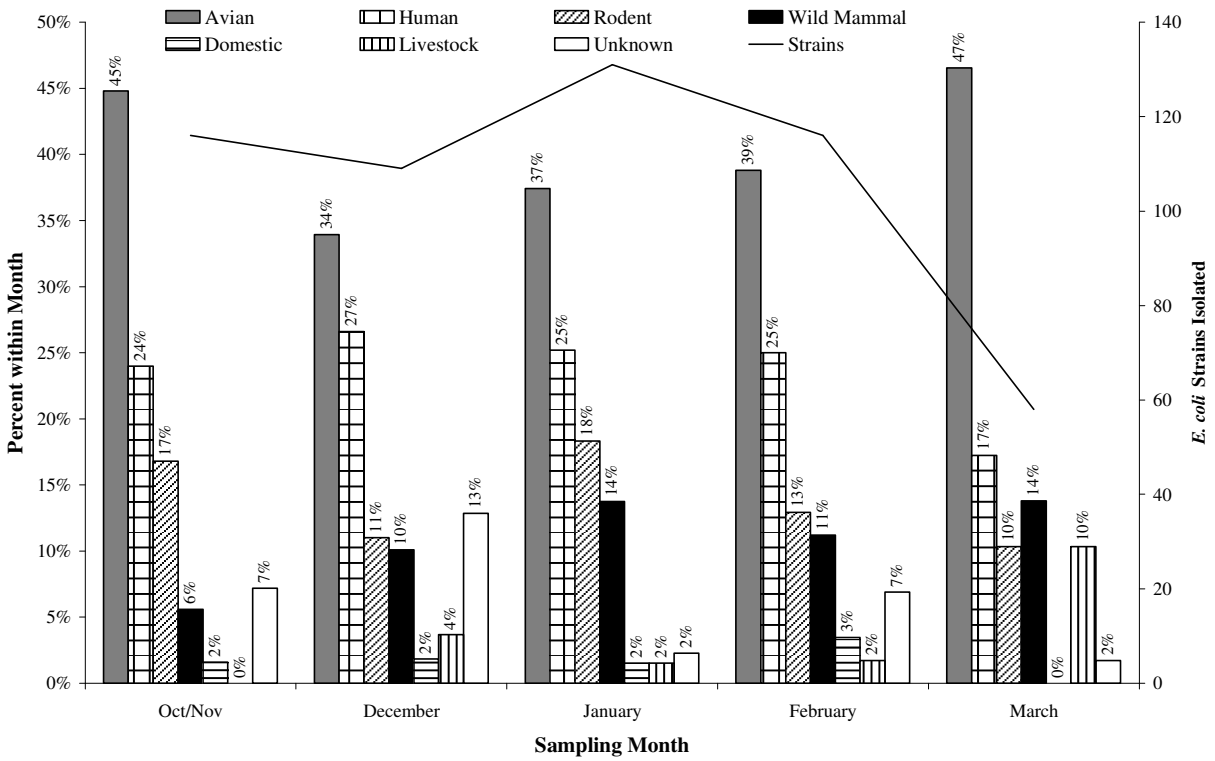

Figure 3. Source group distribution by sampling month. Each bar represents the number of *E. coli* from a specified source isolated in a particular month as a percentage of the total *E. coli* isolated in that month (scale on the left). The line represents the total number of *E. coli* strains isolated in each month (scale on right).

Distribution of Ribotypes

A total of 141 different ribotype fingerprints were produced from the 539 strains isolated in this study (Table 5). Most ribotypes (110) were found only at a single site. Only 15 ribotypes were found at all three sites and all of these ribotypes were from the most abundant source groups: avian, human or rodent/wild mammal. Avian and rodent source groups showed a ratio of ribotypes per strain of about 1:5, avian with 39 ribotypes per 214 strains and rodent with 18 ribotypes per 78 strains. Conversely, the other source groups showed ratios greater than 1:3. There are two possible explanations for this difference. First, the fecal input from avian and rodent sources may be restricted to consistent input from fewer host animals or fewer species, producing a limited number of ribotypes for the large number of strains isolated. Alternatively, there may be generally less variation in *E. coli* strains from avian and rodent sources. However, the second possibility is unlikely since the avian source category in particular covers many species while the human source category covers only one host species but still showed a higher ratio (~1:3). However, the differences in ribotype to strain ratio between source groups may not be significant considering the low number of strains isolated in this study.

Table 5. *E. coli* ribotypes seen in this study, as distributed by source group and sample site.

Seen at	Avian	Domestic	Human	Livestock	Rodent	Wild	Unknown	Total
Site 1	12	1	12	0	8	5	7	45

1 only	6	1	6	0	4	2	6	25
Site 2	16	6	23	2	9	7	7	70
2 only	8	6	12	1	4	3	6	40
Site 3	24	1	19	7	10	9	1	71
3 only	17	1	9	6	4	7	1	45
1 & 2	1	0	1	0	0	2	1	5
1 & 3	0	0	0	0	1	0	0	1
2 & 3	2	0	5	1	2	0	0	10
All 3 Sites	5	0	5	0	3	2	0	15
Sum	39	8	38	8	18	16	14	141

CONCLUSIONS

Dominant Fecal Sources

- Avian and human sources clearly dominated all three sites.
- Rodent and wild mammal sources together were as abundant as human sources.
- Domestic animal and livestock contributions were minimal.

Source Distribution by Site and Month

- Differences in contribution by site were only significant for the low abundance sources: domestic animals and livestock. These differences might not be significant if more strains were isolated.
- Differences in contribution by collection month were only significant for livestock (a very low abundance source) and so may also be artifactual.

Ribotype Distributions

- Avian and rodent sources were possibly contributed consistently from the same or very similar host animals. It is unclear if this means the same individual animals or just animals of the same species.