

Population dynamics and genetic variability of *Escherichia coli* in a mixed land-use watershed

R. Udenika Wijesinghe, Yucheng Feng, C. Wesley Wood,
Donald M. Stoeckel and Joey N. Shaw

ABSTRACT

Better understanding of *Escherichia coli* population dynamics and genetic variability in the secondary habitat is essential to improve fecal contamination monitoring and contamination pathway characterization. In this study, water samples were collected monthly over a one-year period at eight locations in the Catoma Creek watershed, a mixed land-use watershed in Central Alabama. *E. coli* concentrations varied from 17 to 12,650 CFU/100 ml and were well correlated with stream flow rates. Repetitive sequence-based PCR DNA fingerprinting was used to generate 271 unique DNA fingerprint patterns from 502 *E. coli* isolated from water samples. Cluster analysis showed an overall similarity of 32.8% across all DNA fingerprints. Multivariate analysis of variance (MANOVA) showed that *E. coli* genotypes had a tendency to cluster according to season and stream flow rather than sampling sites. MANOVA of a subset of data within a given season and flow rate, however, revealed some geographical differentiation between urban and rural sampling sites. The results indicate that genetic diversity of *E. coli* populations was not only high in the secondary habitat but also varied with season, flow conditions and, to a lesser extent, sampling location. To our knowledge, this is the first report relating *E. coli* genotype to stream flow.

Key words | *E. coli*, fecal contamination, rep-PCR, secondary habitat, water quality

R. Udenika Wijesinghe
Yucheng Feng (corresponding author)
C. Wesley Wood
Joey N. Shaw
Department of Agronomy and Soils,
Auburn University,
Auburn, AL 36849,
USA
Tel.: 334-844-3967
Fax: 334-844-3945
E-mail: yfeng@auburn.edu

Donald M. Stoeckel
Environmental Sciences Graduate Program,
The Ohio State University,
Columbus, OH 43210,
USA

INTRODUCTION

Fecal contamination of surface water may pose health risks to humans and threaten the integrity of ecosystems. In the USA, the Clean Water Act was enacted in 1972 to achieve the goal of 'all water to be swimmable and fishable' (http://cfpub.epa.gov/npdes/cwa.cfm?program_id=0, accessed 16 September 2008). However, contamination of surface water is still a common problem nationwide. There are more than 1 million km of impaired rivers and streams in the USA and the most reported cause of impairment is high levels of fecal indicator bacteria (USEPA 2008).

Escherichia coli is a common type of fecal coliform bacteria that normally inhabits intestinal tracts of warm-blooded animals including humans. Despite its limitations,

E. coli has been used as an indicator of fecal contamination in fresh water (USEPA 1986). Potential sources of *E. coli* in the aquatic environment include runoff from land-applied animal wastes or animal feeding operations, urban runoff, inadequate or failing septic or sewer systems, and wildlife defecation. Environmental conditions, such as water chemistry, season and hydrologic processes, are known to influence *E. coli* densities and survival in surface waters (Gentry *et al.* 2006; Avery *et al.* 2008). Increased *E. coli* concentrations in water samples have been found during wet periods, especially after a storm, possibly because of transport of fecal material from the surrounding area and re-suspension of bacteria into water from disturbed

doi: 10.2166/wh.2009.105

sediments (Crabill *et al.* 1999; An *et al.* 2002). *E. coli* and fecal coliform concentrations often are higher in sediments or beach sand than in overlying water columns (Crabill *et al.* 1999; Bonilla *et al.* 2007). This has been attributed to better survival (Enzinger & Cooper 1976; An *et al.* 2002) and/or regrowth (Hartz *et al.* 2008) of fecal indicator bacteria owing to the presence of organics, a decreased level of UV radiation, and the presence of anaerobic regions in sediments/beach sand.

Recently, *E. coli* has also been used as a source identifier in bacterial source tracking studies (Scott *et al.* 2002; USEPA 2005). Thus, the diversity of *E. coli* populations present in fecal materials from humans and animals has been the subject of numerous studies (e.g. Carson *et al.* 2001; Gordon & Cowling 2003; McLellan *et al.* 2003; Johnson *et al.* 2004). However, limited information is available regarding the impact of season and flow conditions on the diversity of *E. coli* in contaminated surface waters. McLellan (2004) reported that there is less diversity in *E. coli* isolates obtained from river and beach water than in isolates from human and non-human sources. Kon *et al.* (2007) showed that *E. coli* isolates in interstitial water from the same sampling location on a Lake Huron beach have highly similar rep-PCR (repetitive sequence-based polymerase chain reaction) patterns, whereas those from different sampling locations were distinctly different. In contrast, Walk *et al.* (2007) reported lack of geographic differentiation among *E. coli* isolated from water and beach sand at different sampling sites on six other beaches of Lake Huron based on multilocus enzyme electrophoresis and multilocus sequence typing. Casarez *et al.* (2007) pointed out the importance of the discriminatory power of different techniques utilized to assess *E. coli* genetic diversity since contradictory conclusions may be reached as a result of differences in discriminatory power.

This study was conducted in the Catoma Creek watershed, a mixed land-use watershed of the Alabama River Basin, Alabama, USA. Catoma Creek was added to Alabama's 303(d) list of impaired water bodies in 1996 because of organic enrichment and low dissolved oxygen, and in 2002 because of high fecal coliform concentrations (ADEM 2008). We evaluated the population dynamics and genetic diversity of *E. coli* isolated from the Catoma Creek watershed and examined seasonal and flow impacts on these characteristics.

MATERIALS AND METHODS

Study site

The Catoma Creek watershed (Figure 1) is located in Montgomery County, Alabama, where the state capital Montgomery (32°23'N, 86°22'W, 76 m above mean sea level) lies at the northern end of the watershed. The watershed includes more than 50% of land in the county, most of which lies in the Blackland Prairie physiographic region. Catoma Creek, a tributary of the Alabama River, is approximately 68 km in linear length and drains an area of 932 km². The largest tributary of Catoma Creek is Ramer Creek, and other tributaries include Baldwin Slough, White Slough and Little Catoma Creek. Forest, pasture and urban land are the top three land use types representing 34%, 31% and 20%, respectively, of the watershed (personal communication, the Montgomery County Watershed Assessment Plan). The total population in Montgomery County is 223,510, of whom 88% reside in an urban area (US Census Bureau 2000). The average annual rainfall in the watershed is 1,391 mm (NOAA 2006). The present use classification of Catoma Creek is 'Fish and Wildlife'. A 37-km segment from Ramer Creek to the Alabama River is included on Alabama's 303(d) list of impaired water bodies because of organic enrichment and elevated fecal coliform concentrations (ADEM 2008).

The Catoma Creek watershed is located in the Sub-ecoregion 65, Southern Plains, within the Nutrient Ecoregion IX, Southeastern Temperate Forest Plains and Hills (USEPA 2000). Nutrient reference criteria for each Nutrient Ecoregion have been developed by USEPA to provide guidelines for each state to develop more refined nutrient criteria. Ecoregional nutrient criteria represent conditions of surface waters with minimal impacts from human activities (USEPA 2000).

Collection of water samples and field measurements

Water samples were collected monthly from eight locations in the Catoma Creek watershed from May 2003 to April 2004. Three locations were on the main stem of Catoma Creek and were designated as CO, CC and CW (Figure 1). Five locations were on tributaries of Catoma Creek:

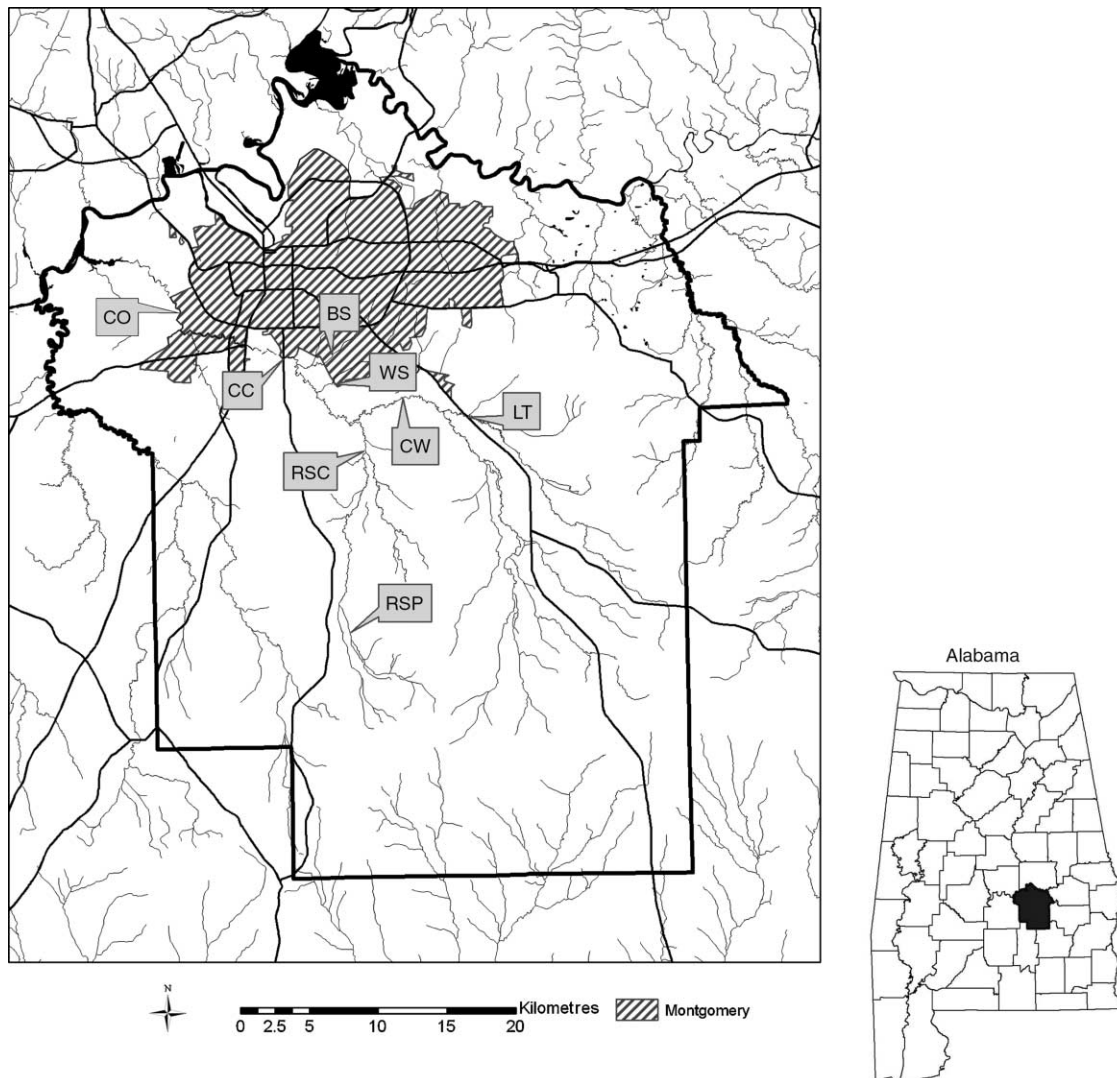


Figure 1 | Catoma Creek watershed showing sampling site designations (bold letters), the county boundary (heavy line), major roads (medium lines) and stream channels (thin lines).

Baldwin Slough (BS), White Slough (WS), Ramer Creek (RSP and RSC) and Little Catoma Creek (LT) (Figure 1). Duplicate grab samples were collected from each sampling site by submerging sterile polyethylene bottles while avoiding streambed disturbance. Samples were placed on ice, transported to the laboratory and processed within six hours of collection. One field blank sample was prepared at one randomly selected site each month. Stream water temperature and pH were measured on site. Flow measurements were performed at three sites where wading was possible: BS, WS and RSP. Flow velocities were measured across the streams at 1-m intervals using a Flo-Mate portable flowmeter (Marsh-McBirney Inc., Frederick, Maryland) each month. Water

level height with respect to each velocity and stream width was recorded and used to calculate mean flow rate (mean flow rate = average velocity \times stream width \times average depth of the flow). A US Geological Survey (USGS) gauging station is located at the CC site and flow data were obtained from the USGS National Water Information System website.

Enumeration and isolation of *E. coli*

E. coli concentration in water samples was analysed by membrane filtration followed by cultivation on modified mTEC agar (USEPA 2002). After enumeration, presumptive *E. coli* colonies (approximately five per sampling site) were

picked and streaked on MacConkey agar (Becton, Dickinson & Company, Sparks, Maryland) for confirmation and genetic testing. After overnight incubation at 37°C, a single colony of dark pink colour was selected from each plate. Half of that colony was streaked on CHROMagar ECC (CHROMagar Microbiology, France) and the other half was streaked on MacConkey agar. After overnight incubation at 37°C, colonies that were dark pink in colour on the MacConkey agar and blue on the CHROMagar ECC were selected to inoculate citrate agar (BBL, Cockeysville, Maryland), EC broth with 4-methylumbelliferyl-D-glucuronide (EC-MUG) (BD), 1% tryptone (Fisher Biotech, Fair Lawn, New Jersey), and methyl red–Voges-Proskauer (MR-VP) broth (BD). Isolates were confirmed as *E. coli* if they did not use citrate as a substrate, grew at 44.5°C, produced gas and fluorescence in EC-MUG broth, produced indole from tryptophan, and produced an acidic end product when grown in MR-VP broth. Confirmed *E. coli* isolates were suspended in tryptic soy broth (Remel, Lenexa, Kansas) with 50% glycerol and stored at –80°C prior to further analysis.

rep-PCR DNA fingerprinting

Repetitive sequence-based PCR (rep-PCR) DNA fingerprints of the *E. coli* isolates were obtained using the BOX A1R primer (5'-CTA CGG CAA GGC GAC GCT GAC G-3') (Rademaker & de Bruijn 1997) and *E. coli* whole cells as the templates for PCR. PCR was performed according to a protocol modified after Rademaker & de Bruijn (1997) and Dombek *et al.* (2000). Briefly, *E. coli* isolates were grown on Plate Count agar (BD) for 18 hours. A portion of a single colony was then removed using a 1 µl sterile inoculation loop and suspended in 100 µl of PCR grade water in a microcentrifuge tube. The 25 µl PCR mixture contained 2 µl of whole cell suspension as template, 1X PCR buffer without MgCl₂ (Promega, Madison, Wisconsin), 2 units of Taq DNA polymerase (Promega), 0.2 mM (each) dNTP (Promega), 0.4 µM of BOX A1R primer (Invitrogen, Carlsbad, California), 3 mM of MgCl₂ (Promega), and 0.16 µg µl⁻¹ of bovine serum albumin (Sigma Life Science, St Louis, Missouri). PCR was performed using a TGradient thermocycler (Biometra, Germany) using the following conditions: initial denaturation at 95°C for 5 min, 35 cycles of denaturation at

94°C for 1 min, annealing at 60°C for 1 min, and extension at 72°C for 1 min, then a final extension at 72°C for 10 min. A negative control containing sterile water and a positive control containing *E. coli* ATCC 25,922 were included in each PCR run. PCR products were mixed with 5 µl of 6X loading dye (Promega) and 10 µl of each mixture was resolved using 1.5% agarose (25 cm × 20 cm) prepared with 0.5X Tris-borate-EDTA (TBE) buffer. One kb Plus DNA ladder (0.66 µg/well; Invitrogen) was added to the 1st, 10th, 19th, 28th and 36th lanes to allow for variable migration rates across the gel slab; a repeated-measure positive control was added to the 2nd lane and a negative control was added to the 35th lane. The gels underwent electrophoresis at room temperature for 7 h at 130 V with constant buffer circulation and then stained for 1 h in 0.5X TBE buffer containing 0.5 µg of ethidium bromide per ml. Gel images were captured using a Gel Logic 200 imaging system (Eastman Kodak Co., Rochester, New York).

Chemical analysis

Every three months, two additional water samples from each site were collected and analysed for total N, NH₄-N, NO₃-N, total P and metal ion concentrations. Ammonium-N and NO₃-N were determined using a modified indophenol method adapted to microplate format (Sims *et al.* 1995). Total N was measured using the Kjeldahl method described by Bremner & Mulvaney (1982), and total P and metal ions (Ca, K, Mg, Al, Zn, Cu, Mn and Fe) were determined using a SPECTRO CIROS inductively coupled plasma atomic emission spectrometer (SPECTRO Analytical Instruments, Germany). Electrical conductivity (EC) was determined monthly for each site in the laboratory with a conductivity bridge.

Data analysis

Correlations of *E. coli* concentrations with chemical and flow parameters as well as precipitation were determined using raw data without transformation. BioNumerics version 5.0 software (Applied Maths, Belgium) was used to analyse the rep-PCR DNA fingerprints of *E. coli* isolates and to perform multivariate analysis of variance (MANOVA). Each gel was normalized by using the 1 kb Plus DNA ladders, in

the range from 200 to 4,000 bp, as an external reference standard to allow comparison of multiple gels. DNA fingerprints were compared using a densitometric curve-based method with the Cosine coefficient with 0.5% optimization and 1% position tolerance. Dendrograms were developed using the unweighted pair group method with arithmetic averages (UPGMA). A similarity score of 90% was used as the cut-off for the same strain types. Isolates with similarity scores below this value were considered unique. MANOVA was used to determine fingerprint distribution based on the variability among *E. coli* isolates from different seasons, stream flow and sampling sites.

RESULTS AND DISCUSSION

Physiochemical parameters

During the 12-month sampling period, pH, water temperature and EC were monitored monthly for all eight sampling sites in the Catoma Creek watershed (Table 1). The mean water pH values ranged from 6.72 to 7.21. The geometric mean water temperature ranged from 16.9°C to 19.5°C and EC from 0.11 to 0.28 mmhos cm⁻¹. Nutrient analyses were performed seasonally and the results are summarized in Table 2. Ammonium-N concentrations varied from 0.003 mg l⁻¹ to 0.783 mg l⁻¹; none of the samples exceeded the EPA criterion for fresh water aquatic life at corresponding pH and temperature (USEPA 1999). The level of NO₃-N ranged from 0.003 to 1.177 mg l⁻¹; 78% of the samples were

above the Ecoregion IX reference level of 0.125 mg l⁻¹ for NO₂ + NO₃-N (USEPA 2000). Water samples collected in the winter showed highest NO₃-N concentrations at seven of the eight sites. Total N concentrations were found at levels up to 4.704 mg l⁻¹; 44% of the samples exceeded the Ecoregion IX criterion for total nitrogen (0.692 mg l⁻¹) (USEPA 2000).

The Ecoregion IX nutrient criterion for total P is 0.036 mg l⁻¹ (USEPA 2000). Measured concentrations of total P at all sites were consistently higher than this criterion (Table 2). Another often cited total P critical concentration for stream eutrophication is 0.1 mg l⁻¹ (USEPA 1986). Again, all the water samples from the watershed exceeded this level. The LT site had the highest geometric mean concentration of 0.304 mg l⁻¹. Thus, there is a risk of eutrophication in the Catoma Creek watershed. Concentrations of nutrients found in this study are within the ranges reported in a USGS study for the Mobile River Basin (McPherson *et al.* 2003), in which the Catoma Creek is located. Elevated total P, total N and NO₃-N levels in the watershed may influence persistence of *E. coli* in the stream water.

E. coli concentrations

From May 2003 to April 2004, 96 water samples were collected from eight locations in the Catoma Creek watershed and all samples were positive for *E. coli*. *E. coli* concentrations ranged from 17 CFU/100 ml in March at

Table 1 | Ranges and means for pH, water temperature, electrical conductivity (EC) and stream flow

Site name	pH		Temperature (°C)		EC (mmhos cm ⁻¹)		Flow (m ³ s ⁻¹)	
	Range	Mean	Range	GM	Range	GM	Range	GM
CO	5.49–7.58	6.86	10.0–27.0	17.5	0.10–0.20	0.16	NA	NA
CC	5.50–7.75	6.84	10.0–27.0	17.6	0.11–0.20	0.16	3.3–253	36.9
BS	5.61–7.91	7.21	10.0–31.0	19.5	0.12–0.49	0.28	0–0.54	0.02
WS	5.72–7.70	7.12	7.0–28.0	16.9	0.14–0.37	0.25	0.01–0.63	0.06
CW	5.76–7.59	6.77	9.0–28.0	17.7	0.06–0.15	0.11	NA	NA
RSC	5.95–7.39	6.89	9.5–27.0	17.8	0.13–0.24	0.19	NA	NA
RSP	5.77–7.61	6.94	11.0–27.0	18.2	0.12–0.18	0.15	0.01–0.60	0.10
LT	5.80–7.56	6.72	11.5–29.0	19.4	0.07–0.24	0.16	NA	NA

GM: geometric mean.

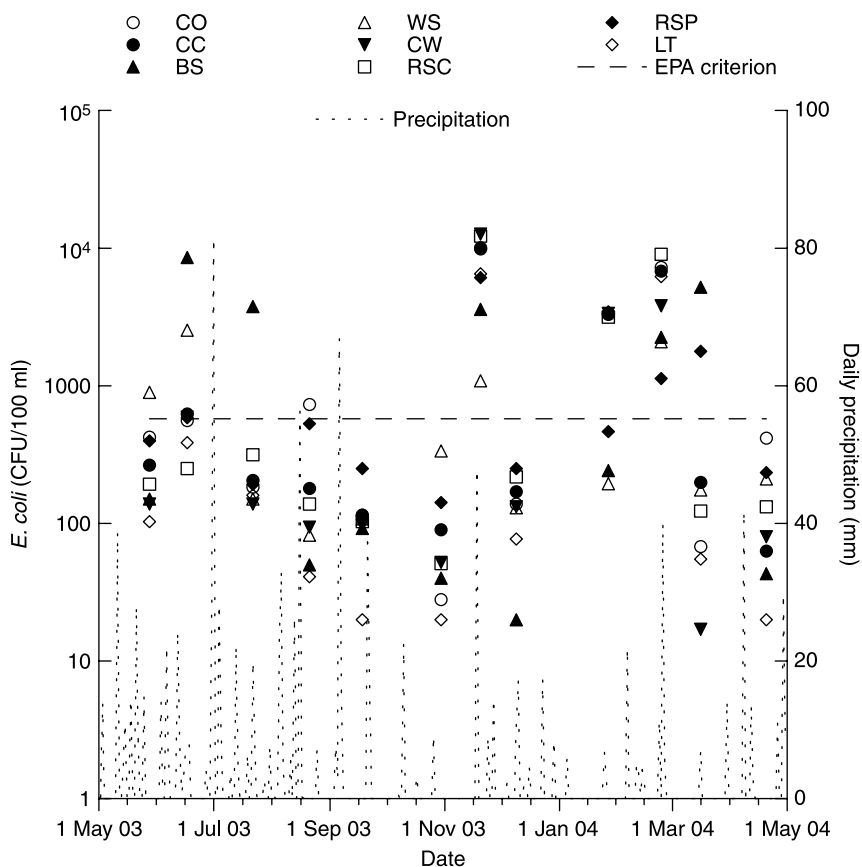
NA: not available.

Table 2 | Ranges and geometric means (GM) for *E. coli* and nutrient concentrations

Site name	<i>E. coli</i> (CFU/100 ml)		NH ₄ -N (mg l ⁻¹)		NO ₃ -N (mg l ⁻¹)		Total N (mg l ⁻¹)		Total P (mg l ⁻¹)	
	Range	GM	Range	GM	Range	GM	Range	GM	Range	GM
CO	28–10,050	462	0.003–0.404	0.078	0.003–0.284	0.074	0.075–4.704	0.404	0.195–0.294	0.235
CC	63–9,850	419	0.003–0.385	0.033	0.092–0.266	0.169	0.075–2.274	0.385	0.157–0.234	0.194
BS	19–8,517	373	0.003–0.417	0.031	0.197–0.530	0.324	0.075–3.763	0.960	0.137–0.280	0.223
WS	82–2,533	337	0.026–0.335	0.128	0.013–1.177	0.181	0.075–1.563	0.160	0.147–0.270	0.201
CW	17–12,650	271	0.104–0.328	0.195	0.003–0.318	0.061	0.075–2.352	0.617	0.150–0.254	0.207
RSC	51–12,200	393	0.003–0.349	0.028	0.003–0.326	0.075	0.075–1.421	0.156	0.130–0.347	0.218
RSP	142–6,100	506	0.003–0.413	0.077	0.003–0.536	0.091	0.075–2.038	0.563	0.141–0.518	0.233
LT	20–6,500	181	0.006–0.783	0.087	0.204–0.651	0.351	0.075–3.410	0.195	0.215–0.656	0.304

the CW site to 12,650 CFU/100 ml in November at the same site (Figure 2). The highest *E. coli* concentrations recorded at the CO, CC, BS, WS, RSC, RSP and LT sites were 10,050, 9,850, 8,517, 2,533, 12,200, 6,100 and 6,500 CFU/100 ml, respectively. Water samples collected from the RSP site had

the highest geometric mean *E. coli* concentration (Table 2). The other sampling site (RSC) on Ramer Creek downstream from the RSP site had a much lower geometric mean *E. coli* concentration. Samples collected at the LT site had the lowest geometric mean for *E. coli* although water at this site

**Figure 2** | *E. coli* concentrations at Catoma Creek sampling sites from May 2003 to April 2004. The USEPA criterion shown is the single sample maximum of 576 CFU/100 ml.

was visibly turbid and stagnant during the entire study period. Among three sampling sites on the main stem of Catoma Creek, the geometric mean *E. coli* concentrations gradually increased from the upstream CW site to the downstream CO site. Both CO and CC sites were located in the urban area.

According to USEPA recommendations for freshwater, the steady state geometric mean *E. coli* concentration allowable for swimming water is 126 CFU/100 ml; single sample maximum *E. coli* concentration criteria range from 235 CFU/100 ml for designated bathing beach to 576 CFU/100 ml for infrequently used swimming water (USEPA 1986). Of the 96 water samples collected in the Catoma watershed, 30% exceeded the single sample maximum value of 576 CFU/100 ml. High concentrations of fecal indicator organisms have frequently been found in surface waters. For example, a study in Michigan reported that five out of nine sampling sites showed *E. coli* concentrations above allowable state limits for recreational water (Tam *et al.* 2005). Davis-Colley *et al.* (2004) reported *E. coli* concentrations of 50,000 CFU/100 ml at a cow-crossing stream in New Zealand, while Howell *et al.* (1995) found that fecal coliforms in Western Kentucky streams exceeded water quality standards between 87 and 100% of the time.

***E. coli* occurrence and nutrient concentrations**

Correlations between *E. coli* and nutrient concentrations were not found during the study period in the Catoma Creek watershed as indicated by the low correlation coefficients ($r = 0.05$ for $\text{NH}_4\text{-N}$, $r = 0.10$ for $\text{NO}_3\text{-N}$, $r = 0.20$ for total N, and $r = 0.03$ for total P). A positive correlation between *E. coli* and nutrient concentrations would indicate that fecal contamination contributes to nutrient enrichment in the Catoma Creek watershed. *E. coli* did not correlate with any measured nutrient parameters, an indication that direct contamination of the watershed with fresh fecal material was not the primary source of nutrient loading during the study period. The origin of nutrients could not otherwise be elucidated; other pathways of fecal-origin nutrients, such as septic leachate, fertilizers and composted manure, may have contributed nutrients (without *E. coli*) to the water bodies studied.

***E. coli* occurrence with precipitation**

The daily rainfall recorded at the CC site is plotted together with *E. coli* concentration in Figure 2. Correlation between the *E. coli* concentration in the stream water and monthly rainfall or 7-day antecedent rainfall was poor ($r < 0.28$). This is probably due to increased hydrological connectivity between uplands and streams following rainfall events. However, the timing and intensity of a rainfall event appeared to be important. Water samples were collected two days after 48 mm of rainfall in November and one day after 11 mm of rainfall in February. High *E. coli* concentrations were found at all sampling sites for both months (Figure 2). During October this watershed experienced mostly dry weather, and the November 48 mm event occurred after a 22-day dry spell. This event resulted in *E. coli* concentration increases of approximately threefold to approximately 360-fold from October to November over the sampling sites. This suggests that rainfall can be a significant factor owing to surface runoff of fecal matter into streams. It is also important to take into account factors that influence infiltration and runoff in terrestrial settings including antecedent soil moisture content and rainfall amount, duration and intensity.

Rodgers *et al.* (2003) studied the effect of hydrological events on fecal coliform concentrations in stream water. After the first significant rainfall event, fecal coliform concentrations in the watershed increased 4- to 100-fold over previous samples. They reported that fecal coliform concentrations decreased with increasing number of rainfall events, suggesting depletion of fecal matter on the land. Another study in New Zealand examined the correlation between runoff and the concentration of fecal bacteria in streams (Collins *et al.* 2005). Their results suggested that overland flow delivered fecal bacteria directly into streams. Crabill *et al.* (1999) reported a dramatic increase of fecal coliforms in stream water after a storm event. They suggested that runoff water transports fecal material that accumulates in watersheds. In addition, increased hydraulic velocity disturbs the bottom sediments, which can be a reservoir of fecal coliforms. Gentry *et al.* (2006) reported that *E. coli* loading rather than concentration at or near base flow is strongly correlated with the 7-day antecedent precipitation. In our study, neither *E. coli* load nor concentrations correlated with the 7-day antecedent precipitation in the Catoma Creek watershed.

***E. coli* occurrence with stream flow**

Stream flow rates at CC, WS, BS and RSP sites are summarized in Table 1. The correlation coefficients (r) between *E. coli* concentration and flow rate were 0.84, 0.81, 0.80 and 0.56, respectively ($P \leq 0.001$). These data demonstrate a positive correlation between flow rates and *E. coli* concentrations despite the lack of statistical correlation with antecedent rainfall. This correlation appeared to be site specific, as correlation was not present when data from all four sites were pooled.

Inconsistent results have been reported regarding the relationship between fecal indicator bacteria and stream flow. Jagals (1997) and Crowther *et al.* (2002) showed a positive relationship between fecal indicator bacteria and stream flow. Donnison *et al.* (2004), however, did not find an obvious correlation between *E. coli* concentration and stream flow or rainfall after monitoring *E. coli* concentrations at 14 sampling sites over a two-year period in a New Zealand rural watershed. In addition to surface runoff delivering fecal bacteria to streams, high *E. coli* concentrations observed under high flow conditions may be attributed to disturbance of sediments, since stream sediments may harbour *E. coli*. Muirhead *et al.* (2004) used a series of three artificial flood events during dry weather to quantify *E. coli* associated with stream sediments. Artificial floods increased the *E. coli* concentration in the water column by two orders of magnitude. The peak *E. coli* concentration coincided with the flood peak and was reduced on each successive flood event. *E. coli* stored in the streambed was estimated to be 10^8 CFU m^{-2} . Although we did not determine *E. coli* concentrations in Catoma Creek sediments, it is possible that *E. coli* harboured by the sediments contributed to the high counts under high flow conditions.

Analysis of DNA fingerprint patterns obtained from water isolates

The 502 *E. coli* isolates obtained from water samples collected from the Catoma Creek watershed were used to generate DNA fingerprints using the BOX A1R primer. DNA fingerprint patterns showed high variability. Total numbers of PCR product bands varied between 5 and 28;

271 unique fingerprint patterns were found. Cluster analysis results showed that overall similarity among DNA fingerprints from the Catoma Creek watershed was 32.8%, suggesting a high degree of genetic diversity. Limited information is available regarding diversity of *E. coli* in stream water. Casarez *et al.* (2007) examined occurrence of *E. coli* genotypes in water samples using pulsed-field gel electrophoresis (PFGE) and enterobacterial repetitive intergenic consensus polymerase chain reaction (ERIC-PCR) methods. Although fewer genotypes were revealed by ERIC-PCR than PFGE, both methods indicate the presence of diverse *E. coli* populations in two Texas lakes and nearby streams.

We also performed multivariate analysis of variance on the rep-PCR DNA fingerprint data to determine whether *E. coli* diversity was correlated with season, flow or sampling location. Figure 3(a) shows the MANOVA results of *E. coli* DNA fingerprint patterns by season. The first three discriminants accounted for 44%, 32% and 25% of the total variance, respectively, with P values ≤ 0.00003 . Although there was mixing of data points to some extent, the low P value indicated there was a tendency of *E. coli* fingerprints to cluster according to season. These results suggest that sources of fecal pollution in this watershed exhibited some seasonality, which can be attributed to temporal change in agricultural land use, ground vegetation cover, wildlife behaviour and/or precipitation and stream flow. Survival of *E. coli* is also influenced by season. Seasonal changes in fecal contamination sources have been reported by Ishii *et al.* (2007). They found that in the Duluth-Superior harbour of Lake Superior, *E. coli* in spring samples originated mainly from treated wastewater effluent whereas the percentage of *E. coli* from waterfowl increased from summer to autumn.

Stream flow at the CC site was obtained from the on-site USGS gauging station and divided into three categories: low flow ($<10 \text{ m}^3 \text{ s}^{-1}$), medium flow ($10\text{--}100 \text{ m}^3 \text{ s}^{-1}$) and high flow ($>100 \text{ m}^3 \text{ s}^{-1}$). Since the CC, CO and CW sites are all on the main stem of Catoma Creek, we assumed that these three sites had similar flow patterns. A portable flow meter was used to measure the flow rates at the BS, WS and RSP sites. Stream flows at these sites were also divided into three categories: low flow ($<0.01 \text{ m}^3 \text{ s}^{-1}$), medium flow ($0.01\text{--}0.1 \text{ m}^3 \text{ s}^{-1}$) and high flow ($>0.1 \text{ m}^3 \text{ s}^{-1}$).

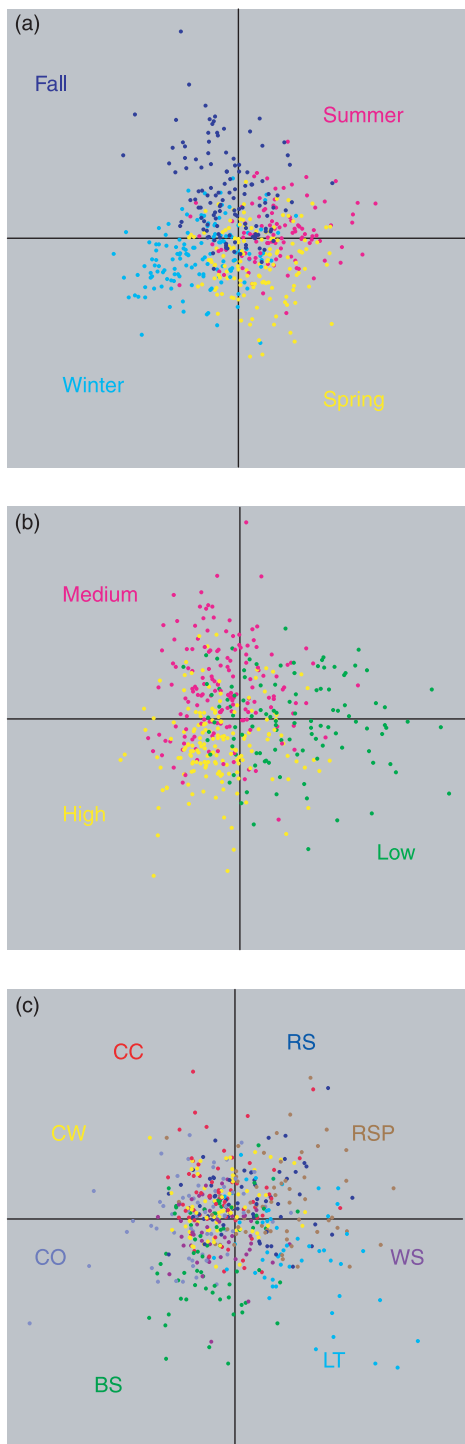


Figure 3 | MANOVA plots of rep-PCR DNA fingerprint patterns of *E. coli* isolated from the Catoma Creek watershed by (a) seasons, (b) stream flow rates and (c) sampling sites. Spring included April, May and June; summer included July, August and September; autumn included October, November and December; and winter included January, February and March.

Since RSP and RSC are two locations on the same creek, we again assumed flow patterns at both sites were similar. LT site had stagnant water during the entire sampling period; therefore we did not use LT site data for this analysis. The total number of *E. coli* isolates used in the data analysis was 450 including 168 isolates each for high and medium flow conditions, and 114 isolates for the low flow condition. MANOVA results showed that DNA fingerprint patterns clustered according to stream flow (Figure 3(b)). The total variance explained by the first two discriminants was 64% ($P = 0.00001$) and 36% ($P = 0.008$), respectively. Thus, *E. coli* DNA fingerprints seemed to vary with the flow rate. High stream flow conditions are often associated with heavy rainfall events. *E. coli* populations in the stream waters probably came from runoff, combined sewer overflows and stream sediments. Under low flow conditions (near base flow), *E. coli* populations in the water may represent those that came from continuing sources, such as leaking septic tanks and wildlife defecation.

Figure 3(c) shows the MANOVA plot of *E. coli* DNA fingerprint patterns based on sampling sites. Data points are intermixed among the sampling sites. The first three discriminants explained 34%, 17% and 14% of the total variance. However, P values of these discriminants were 0.00084, 0.62 and 0.93, respectively, indicating an insignificant relationship between sampling sites and DNA fingerprints. To further evaluate the relationship between *E. coli* genotype and sampling location, the MANOVA was performed within a given season (Figure 4). The sum of the first three discriminants accounted for 63%, 72%, 74% and 66% of the total variance for spring ($0.07 \leq P \leq 0.00006$), summer ($0.59 \leq P \leq 0.0026$), autumn ($0.55 \leq P \leq 0.00041$) and winter ($0.99 \leq P \leq 0.57$), respectively. In the spring, most *E. coli* isolates from the CW site scattered to the lower left of the plot whereas those from the RSP site scattered to the lower right (Figure 4(a)). Land uses surrounding these two sites are dominated by pasture, upland mixed forest and forested wetland. Data points from the six remaining sites were intermixed (Figure 4(a)). In summer, *E. coli* isolates from the RSP site were separated from all other sites (Figure 4(b)). Within the large cluster of summer data, most isolates from the CW site formed a smaller and tighter sub-cluster on the upper right side of the plot. Data obtained in the autumn showed two major

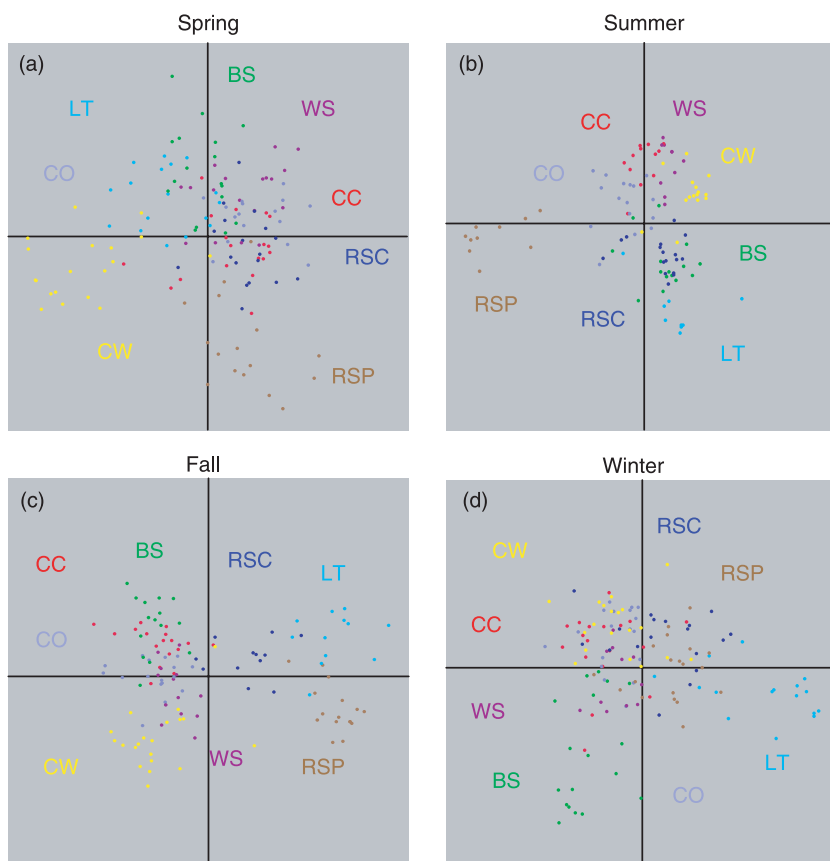


Figure 4 | MANOVA plots of rep-PCR DNA fingerprint patterns of *E. coli* isolated from the Catoma Creek watershed within a given season: (a) spring, (b) summer, (c) autumn and (d) winter.

clusters, with isolates from three rural sites (RSP, RSC and LT) grouped together on the right side of the plot (Figure 4(c)). Isolates from the CW site appeared on the lower part of the cluster on the left side of the plot whereas those from urban sites (CO, CC, BS and WS) were intermixed (Figure 4(c)). The data set for winter samples had high P values and the majority of data points were mixed (Figure 4(d)). Taken together, MANOVA of *E. coli* isolates within a given season revealed *E. coli* genetic variability by sampling location, which were not shown when the entire data set was used. Rural sites, especially RSP, seemed to have *E. coli* genotypes different from urban sites for most of the year. Unlike the other two sites on the main stem of Catoma Creek, the CW site is located in a rural area. *E. coli* isolates from this site in summer and autumn formed a sub-cluster of their own within a larger cluster, supporting the previous observation that *E. coli* isolated from rural sites were different from those from urban sites.

An additional MANOVA was also performed under a given flow condition with the LT site excluded from the analysis (Figure 5). The sum of the first three discriminants accounted for 92%, 68% and 69% of the total variance for low ($0.036 \leq P \leq 0.00001$), medium ($0.2 \leq P \leq 0.02$) and high flow rates ($0.65 \leq P \leq 0.1$), respectively. Under medium and high flow conditions, data points from all seven sampling sites were mostly intermixed (Figure 5(b) and (c)). Under low flow conditions, there were fewer *E. coli* isolates, none of which was obtained from the WS site (Figure 5(a)). These isolates seemed to cluster better by sampling location. Although there were only seven isolates from Ramer Creek sites (four isolates from RSP and three from RSC), they formed distinct clusters by location and were well separated from the four remaining sites (Figure 5(a)). Within the cluster formed by isolates from the BS, CO, CC and CW sites, *E. coli* from the CC and CW sites formed two sub-clusters at the two ends whereas those

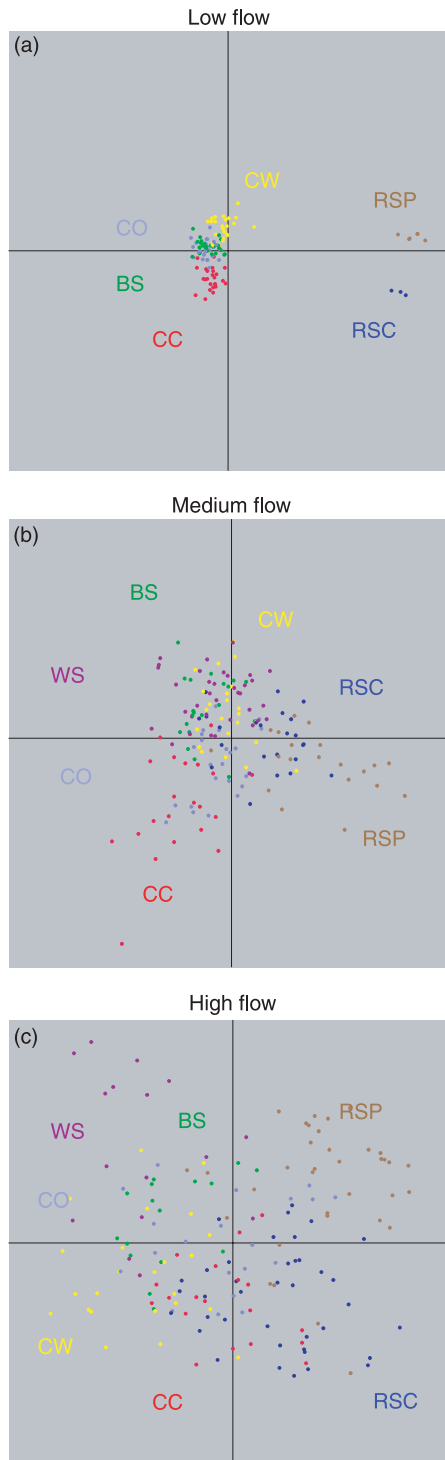


Figure 5 | MANOVA plots of rep-PCR DNA fingerprint patterns of *E. coli* isolated from the Catoma Creek watershed within a given flow condition: (a) low flow, (b) medium flow and (c) high flow.

from the BS and CO sites were intermixed in the middle (Figure 5(a)). These results indicate that genetic variability of *E. coli* can be confounded by fecal source, season and hydrological variability. Spatial variability of *E. coli* seems to be more prominent under base flow and dry conditions.

CONCLUSIONS

This study showed that stream flow and *E. coli* concentration were positively correlated ($0.56 < r < 0.84$). There was no overall statistical correlation between *E. coli* concentration and rainfall. However, rain after a long dry period increased the *E. coli* concentration by up to 360-fold relative to the previous dry month. Therefore, rainfall appeared to be an important contributor to fecal loading into the streams. There were no correlations between *E. coli* and nutrient concentrations during the study period. Analysis of rep-PCR DNA fingerprint profile revealed a high degree of genetic diversity of *E. coli* in the secondary habitat. Cluster formation of *E. coli* isolates by season and flow indicates that some *E. coli* were carried to the stream by overland flow, and that sources of *E. coli* contamination varied by season. Geographic differentiation among sampling locations was not observed when the entire data set was used in the MANOVA. Under low flow conditions and in certain seasons, differentiation of *E. coli* genotypes by sampling location appeared to be more prominent.

ACKNOWLEDGEMENTS

The authors acknowledge Brenda Wood and John Owen for technical assistance, and the Montgomery Water Works and Sanitary Sewer Board for equipment support. Gratitude is also extended to April Jones, Sabra Sutton and Brian Atkins for their assistance with selection of sampling sites. This project would not be possible without the moral support of Dr Wil Smith, MD, of Montgomery, Alabama. This work was supported, in part, by grants from the Alabama Water Resource Research Institute, the Montgomery County Soil and Water Conservation District and the Alabama Agricultural Experiment Station.

REFERENCES

- ADEM (Alabama Department of Environmental Management) 2008 303 (d) Information and Map. Available at <http://www.adem.state.al.us/WaterDivision/Wquality/303d/WQ303d.htm> (accessed 16 September 2008).
- An, Y. J., Kampbell, D. H. & Breidenbach, G. P. 2002 *Escherichia coli* and total coliform in water and sediments at lake marinas. *Environ. Pollut.* **120**(3), 771–778.
- Avery, L. M., Williams, A. P., Killham, K. & Jones, D. L. 2008 Survival of *Escherichia coli* O157:H7 in waters from lakes, rivers, puddles and animal-drinking troughs. *Sci. Total Environ.* **389**(2–3), 378–385.
- Bonilla, T. D., Nowosielski, K., Cuvelier, M., Hartz, A., Green, M., Esiobu, N., McCorquodale, D., Fleishner, J. & Rogerson, A. 2007 Prevalence and distribution of fecal indicator organisms in South Florida beach sand and preliminary assessment of health effects associated with beach sand exposure. *Mar. Pollut. Bull.* **54**, 1472–1482.
- Bremner, J. M. & Mulvaney, C. S. 1982 Nitrogen–total. In *Methods of Soil Analysis, Part 2. Chemical and Microbiological Properties* (ed. A. L. Page, R. H. Miller & D. R. Keeney), pp. 595–624. American Society of Agronomy, Inc, Madison, Wisconsin.
- Carson, C. A., Shear, B. L., Ellersieck, M. R. & Asfaw, A. 2001 Identification of fecal *Escherichia coli* from humans and animals by ribotyping. *Appl. Environ. Microbiol.* **67**(4), 1503–1507.
- Casarez, E. A., Pillai, S. D. & Di Giovanni, G. D. 2007 Genotype diversity of *Escherichia coli* isolates in natural waters determined by PFGE and ERIC-PCR. *Water Res.* **41**(16), 3643–3648.
- Collins, R., Elliott, S. & Adams, R. 2005 Overland flow delivery of fecal bacteria to headwater pastoral stream. *J. Appl. Microbiol.* **99**(1), 126–132.
- Crabill, C., Donald, R., Snelling, J., Foust, R. & Southam, G. 1999 The impact of sediment fecal coliform reservoirs on seasonal water quality in Oak Creek, Arizona. *Water Res.* **33**(9), 2163–2171.
- Crowther, J., Kay, D. & Wyer, M. D. 2002 Faecal-indicator concentrations in waters draining lowland pastoral catchments in the UK: relationships with land use and farming practices. *Water Res.* **36**(7), 1725–1734.
- Davis-Colley, R. J., Nagels, J. W., Smith, R. A., Young, R. G. & Phillips, C. J. 2004 Water quality impact of a dairy cow herd crossing a stream. *New Zeal. J. Mar. Fresh. Res.* **38**(4), 569–576.
- Dombek, P. E., Johnson, L. K., Zimmerley, S. T. & Sadowsky, M. J. 2000 Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Appl. Environ. Microbiol.* **66**(6), 2572–2577.
- Donnison, A., Ross, C. & Thorrold, B. 2004 Impact of land use on the fecal microbial quality of hill country streams. *New Zeal. J. Mar. Fresh. Res.* **38**(5), 845–855.
- Enzinger, R. M. & Cooper, R. C. 1976 Role of bacteria and protozoa in the removal of *Escherichia coli* from estuarine waters. *Appl. Environ. Microbiol.* **31**(5), 758–763.
- Gentry, R. W., McCarthy, J., Layton, A., McKay, L. D., Williams, D., Koirala, S. R. & Saylor, G. S. 2006 *Escherichia coli* loading at or near base flow in a mixed-use watershed. *J. Environ. Qual.* **35**(6), 2244–2249.
- Gordon, D. M. & Cowling, A. 2003 The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology-Sgm* **149**, 3575–3586.
- Hartz, A., Cuvelier, M., Nowosielski, K., Bonilla, T. D., Green, M., Esiobu, N., McCorquodale, D. S. & Rogerson, A. 2008 Survival potential of *Escherichia coli* and enterococci in subtropical beach sand: implications for water quality managers. *J. Environ. Qual.* **37**(3), 898–905.
- Howell, J. M., Coyne, M. S. & Cornelius, P. 1995 Fecal bacteria in agricultural waters of the Bluegrass Region of Kentucky. *J. Environ. Qual.* **24**(3), 411–419.
- Ishii, S., Hansen, D. L., Hicks, R. E. & Sadowsky, M. J. 2007 Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior. *Environ. Sci. Technol.* **41**(7), 2203–2209.
- Jagals, P. 1997 Stormwater runoff from typical developed and developing South African urban development: definitely not for swimming. *Wat. Sci. Technol.* **35**(11–12), 133–140.
- Johnson, L. K., Brown, M. B., Carruthers, E. A., Ferguson, J. A., Dombek, P. E. & Sadowsky, M. J. 2004 Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. *Appl. Environ. Microbiol.* **70**(8), 4478–4485.
- Kon, T., Weir, S. C., Howell, E. T., Lee, H. & Trevors, J. T. 2007 Genetic relatedness of *Escherichia coli* isolates in interstitial water from a Lake Huron (Canada) beach. *Appl. Environ. Microbiol.* **73**(6), 1961–1967.
- McLellan, S. L. 2004 Genetic diversity of *Escherichia coli* isolated from urban rivers and beach water. *Appl. Environ. Microbiol.* **70**(8), 4658–4665.
- McLellan, S. L., Daniels, A. D. & Salmore, A. K. 2003 Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution by using DNA fingerprinting. *Appl. Environ. Microbiol.* **69**(5), 2587–2594.
- McPherson, A. K., Moreland, R. S. & Atkins, J. B. 2003 *Occurrence and Distribution of Nutrients, Suspended Sediment, and Pesticides in the Mobile River Basin, Alabama, Georgia, Mississippi, and Tennessee, 1999–2001*. USGS Water Resources Investigations Report 03–4203. Montgomery, Alabama.
- Muirhead, R. W., Davies-Colley, R. J., Donnison, A. M. & Nagels, J. W. 2004 Fecal bacteria yields in artificial flood events: quantifying in-stream stores. *Water Res.* **38**(5), 1215–1224.
- NOAA 2006 *Monthly Station Normals of Temperature, Precipitation, and Heating and Cooling Degree Days 1971–2000*. Climatology of the United States No. 81. National Oceanic and Atmospheric Administration, Asheville, North Carolina.
- Rademaker, J. L. W. & de Bruijn, F. J. 1997 Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer assisted pattern analysis. In *DNA*

- Markers: Protocols, Applications, and Overviews* (ed. G. Caetano-Anolles & P. M. Gresshoff), pp. 151–171. Wiley-VCH, New York.
- Rodgers, P., Soulsby, C., Hunter, C. & Petry, J. 2003 **Spatial and temporal bacterial quality of a lowland agricultural stream in northeast Scotland**. *Sci. Total Environ.* **314–316**, 289–302.
- Scott, T. M., Rose, J. B., Jenkins, T. M., Farrah, S. R. & Lukasik, J. 2002 **Microbial source tracking: current methodology and future directions**. *Appl. Environ. Microbiol.* **68**(12), 5796–5803.
- Sims, G. K. C., Ellsworth, T. R. & Mulvaney, R. L. 1995 **Microscale determination of inorganic nitrogen in water and soil extracts**. *Commun. Soil Sci. Plant Anal.* **26**(1–2), 303–316.
- Tam, J., Scott, T. M., Morgan, M. R. & Rose, J. B. 2005 Occurrence of alternative fecal indicators and enteric viruses in Michigan Rivers. *J. Great Lakes Res.* **31**(1), 22–31.
- US Census Bureau 2000 State & County QuickFacts. Available at <http://quickfacts.census.gov/qfd/states/01/01101.html> (accessed 16 September 2008).
- USEPA 1986 *Ambient Water Quality Criteria for Bacteria*. EPA440/5–84–002 US Environmental Protection Agency, Washington, DC.
- USEPA 1999 *Water Quality Criteria for Ammonia*. EPA-822-R-99–014 US Environmental Protection Agency, Washington, DC.
- USEPA 2000 Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health. Available at <http://epa.gov/waterscience/criteria/humanhealth/method> (accessed 16 September 2008).
- USEPA 2002 *Method 1603: E. coli in water by membrane filtration using modified membrane-thermotolerant Escherichia coli Agar (modified mTEC)*. EPA-R-02–023 US Environmental Protection Agency, Washington, DC.
- USEPA 2005 *Microbial Source Tracking Guide Document*. EPA/600/R-05/064 US Environmental Protection Agency, Cincinnati, Ohio.
- USEPA 2008 *National summary of state information*. Available at: http://iaspub.epa.gov/waters10/attains_nation_cy.control (accessed 20 March 2009).
- Walk, S. T., Alm, E. W., Calhoun, L. M., Mladonicky, J. M. & Whittam, T. S. 2007 **Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches**. *Environ. Microbiol.* **9**(9), 2274–2288.

First received 18 September 2008; accepted in revised form 6 December 2008. Available online May 2009