The genetic relatedness of *E. coli* associated with post-collection drinking water contamination in rural households

M du Preez[1]*, W le Roux[1], N Potgieter[2] and SN Venter[3]

[1]*Natural Resources and the Environment, CSIR, PO Box 395, Pretoria 0001, South Africa*

[2]*Department of Microbiology, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa*

[3]*Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0002, South Africa*

Abstract

Rural households are often dependent on rivers, springs, boreholes or standpipes some distance from their homes for their daily water requirements. Water for drinking and domestic use is consequently stored in containers in-house which are prone to post-collection contamination. The objective of the study was to determine the most likely origin or place of introduction of *E. coli* associated with post-collection contamination in rural households, by assessing the degree of genetic relatedness of *E. coli* present in the stored water and other environmental samples. *E. coli* isolates were obtained using either mFC agar with confirmation of indole production (44 isolates) or Colilert®-18 (52 isolates). Amplified fragment length polymorphism (AFLP) fingerprinting was applied to determine the genetic relatedness of *E. coli* isolated from in-house storage containers, drinking cups, hand-swab samples, cattle dung and from the source water (spring water). DNA fingerprints of *E. coli* produced a number of clusters (>85% similarity scores calculated with the cosine coefficient). Identical *E. coli* genetic patterns were observed at closely linked points within the domestic pathway of water handling, such as between hand-swab and drinking-cup samples, between storage container and source isolates, and between drinking cups, source water and storage containers. The results indicated that AFLP fingerprinting could be applied to determine the genetic relatedness of *E. coli* isolated from closely linked points within the domestic pathway of water use within a household. However, the high genetic diversity observed for *E. coli* bacteria isolated from the different water and environmental samples tested in this study, hampered the identification of post collection points of contamination.

Keywords: typing, fingerprinting, amplified fragment length polymorphism, *E. coli,* water quality, genetic relatedness, AFLP

Introduction

Many people in developing countries are still dependent on untreated rivers, springs or boreholes for drinking and domestic purposes. To improve access to better quality water in rural areas in South Africa much effort has gone into providing people with protected boreholes and standpipes at some distance from their homes. In these areas water for drinking and domestic purposes is mostly stored in-house in containers. However, many studies have shown that the microbial quality of the water stored in-house, deteriorates considerably between point-of-collection and point-of-use (Wright et al., 2004; Jensen et al., 2002; Sobsey, 2002, Maraj et al., 2006).

Quantifying faecal coliforms by culturing is the most widely used method for measuring faecal pollution (*Standard Methods*, 1995). This method, however, does not identify the source of the contamination. During the last decade, many methods directed at the DNA of bacteria have been developed to 'fingerprint' genetic characteristics of micro-organisms (Chasseignaux et al., 2001; Wang et al., 1999). Through application of these methods, the genotypic characteristics of the standard water quality indicator bacteria, *E. coli* have been applied to understand the origin and sources of faecal pollution (Guan et al., 2002; Stoeckel et al., 2004). Consequently several methods described as bacterial and microbial source-tracking methods have been developed to distinguish between various sources of faecal contamination in water sources (Meays et al., 2004; Hagedorn et al., 1999; Harwood et al., 2000). Researchers employed ribotyping and repetitive extragenic palindromic-polymerase chain reaction (Myoda, 2003; Carson et al., 2003) polymerase chain reaction restriction fragment length polymorphism (PCR-RFLP; Aslam et al*.,* 2003), pulsed field gel electrophoresis (PFGE; Liebana et al., 2003), fluorescent amplified fragment length polymorphism (FAFLP; Smith et al., 2000) and others.

Various comparative studies based on the discriminatory powers and resolution of these different molecular methods have been conducted (Scott et al., 2003; Radu et al., 2001; Vos et al., 1995; Clerc et al., 1998). Fluorescent AFLP has subsequently been shown to have high powers of discrimination, universal applicability and good reproducibility for microbial fingerprinting. The method is based on selective amplification of a subset of DNA fragments from a digest of total genomic DNA (Vos et al., 1995) and has the ability to discriminate between strains of a bacterial species (Arnold et al., 1999; Speijer et al., 1999).

Accordingly, our study determined the genetic relatedness and potential source of *E. coli* bacteria present in stored drinking water in selected rural households using AFLP analysis. Information about the most likely point of post-collection contamination could provide valuable insight when hygiene education, provision of drinking water and management of the in-house quality is considered.

Methodology

Isolates were collected from households in two rural villages and were named by household and type (Table 1 and 2). In

* To whom all correspondence should be addressed.
☎ +27 12 841-3950; fax: +27 12 841-2560;
e-mail: mdupreez@csir.co.za
Received 10 July 2007; accepted in revised form 13 November 2007.

TABLE 1
Source reference, household numbers and the origin of samples from which *E. coli* isolates were obtained for AFLP analysis (Village 1)

Source reference	Household	Origin of water sample
1st	1	Storage container
3st	3	Storage container
3c	3	Cup
3h	3	Hand swab
Dung	3	Dung smeared floor
10st	10	Storage container
10c	10	Cup
10h	10	Hand swab

TABLE 2
Source reference, household numbers and origin of samples from which *E. coli* were obtained for AFLP analysis (Village 2)

Source reference	Household	Origin of water samples
2st	2	Storage container
2c	2	Cup
2h	2	Hand swab
4h	4	Hand swab
4st	4	Storage container
6st	6	Storage container
6c	6	Cup
6h	6	Hand swab
8st	8	Storage container
10st	10	Storage container
Spring	-	Water source for households 2, 6, 8, 10

Village 1 (Households 1, 3, and 10) the samples consisted of traditional plastic storage containers, drinking cups, and hand swabs. The drinking cup sample consisted of the water poured into the cup or scooped from the storage container. The hands of mothers or caregivers and children in a household were swabbed. Dung from a dung-smeared floor (Household 3) was also collected and analysed for *E. coli*. All the households collected treated water from communal standpipes and the water supplied did not contain any *E. coli*. In Village 2 the same type of samples were collected from 5 households (Households 2, 4, 6, 8 and 10, see Table 2). All households except for Household 4 used untreated water obtained from a spring. *E. coli* isolated from the spring was included in the analysis. No *E. coli* could be isolated from the borehole used by Household 4.

All samples collected in Village 1 were analysed by membrane filtration (*Standard Methods*, 1995) using mFC agar. To ensure the optimal recovery of *E. coli,* between 200 and 500 mℓ of all water samples were filtered. The tip of each hand swab was rinsed in sterile saline which was then filtered and analysed for the presence of *E. coli.* The dung sample was analysed by suspending about 1 g in sterile saline and tenfold dilutions of the suspension were analysed on mFC agar. The samples collected from Village 2 were analysed using the Colilert®-18 system. For the water samples 100 mℓ were analysed. The hand swabs were thoroughly rinsed in 100 mℓ sterile water which was then analysed in the same manner as the water samples. *E. coli* positive wells in each Colilert tray were clearly marked. Each tray's upper surface was wiped clean with 70% alcohol. Wells were punctured with a sterile scalpel. A drop of the growth medium was transferred to an mFC agar plate and streaked for single colonies. Plates were incubated at 44.5 ± 0.5°C. All isolates obtained from both villages were confirmed by testing for indole production. Isolates were stored on nutrient agar slopes at 4 to 8°C.

The AFLP method as described by Vos et al. (1995) was followed with some modifications as described by Brady et al. (2007). *E. coli* cells were collected from freshly streaked nutrient agar plates incubated at 37ºC and genomic DNA was extracted from the cell pellets with a DNeasy Tissue Kit (QIAGEN, Hilden, Germany). About 100 ng of genomic DNA was digested with *Eco*RI (Roche, Basel, Switzerland) and *Mse*I (Roche) restriction enzymes. *Eco*RI and *Mse*I adapters (Applied Biosystems, Foster City, California) were ligated to the generated DNA fragments according to the specifications of the manufacturer. Pre-amplification and selective amplification was carried out as described by Brady et al. (2007). Sequences of the primers used are given in Table 3.

TABLE 3
Primer sequences for pre-amplification and selective-amplification during AFLP

Amplification step	Primer	Sequence
Pre-amplification	Eco-O	5' – GACTGCGTACCAATTC – 3'
	Mse-O	5' – GATGAGTCCTGAGTAA – 3'
Selective-amplification	Eco-G	5' – GACTGCGTACCAATTCG – 3'
	Mse-T	5' – GATGAGTCCTGAGTAAT – 3'

The amplified DNA products were separated using an automated PAGE gel system (Licor Global IR2 DNA analyzer, Licor Inc. Nebraska, USA). A sizing standard 50 to 700 bases (Li-Cor 4200-60[700]) was included in each run as a reference. The Automated Li-Cor system generated digitised fingerprints (16 Bit TIFF images) of the gel run, which were used in analysis with GelCompar II software (Applied Maths, Kortrijk, Belgium). Images were normalised by alignment to molecular-size standards loaded on each gel. Curve-based dendograms were generated using the Cosine correlation coefficient.

Results

E. coli strains were considered to be identical or of the same type when AFLP banding patterns were more than 85% similar. Figure 1a shows the genetic relatedness of the 44 *E. coli* strains from Village 1. Twenty-two AFLP types were identified. Nine of these types contained more than one isolate. Of the 9 types only 5 (55%) consisted of *E. coli* isolated from a single sample type (AFLP types B,D,E,J,O) leaving 4 types that showed similar genetic patterns for *E. coli* isolates that originated from different sample types from different points in a household or households. Type *A* showed that *E. coli* identical to those isolated from the hands of a person in Household 3 were also present in the cup sample of the same household. Type F showed that *E. coli,* identical to isolates from a cup sample (Household 10), was also found in the storage container of the same household. Type N consisted of identical *E. coli* that were isolated from storage containers in different households (1 and 3). *E. coli* isolated from the dung sample, collected at Household 3 were identical to *E. coli* isolated from the hands of a person in the same household (Type T). Eight genetically different strains were observed in the storage container of household 1 (AFLP types *K, L, M, N, O, P, Q R).*

Available on website http://www.wrc.org.za
ISSN 0378-4738 = Water SA Vol. 34 No. 1 January 2008
ISSN 1816-7950 = Water SA (on-line)

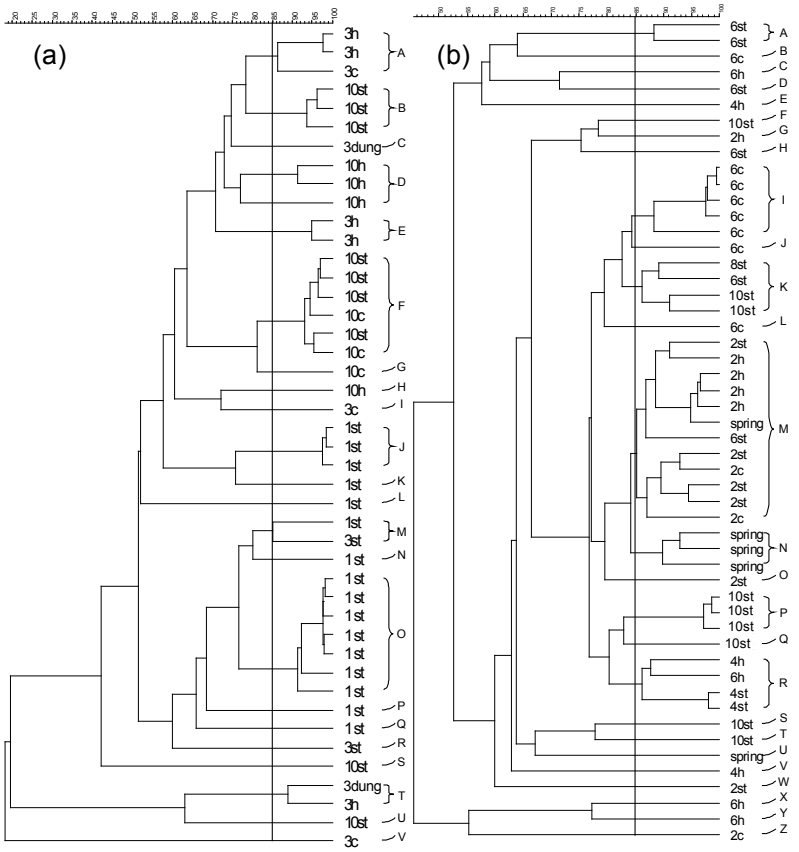

Figure 1
Dendograms based on AFLP patterns of E. coli isolates obtained from Village 1 (Fig. 1a), and Village 2 (Fig. 1b). AFLP types in Fig. 1a are designated A to V and A to Z in Fig. 1b. Household numbers and types of water samples appear in parenthesis. st=storage container; h=hand swab sample; c= cup.

Figure 1b shows the genetic relationship of the 52 *E. coli* strains isolated from Village 2. Results include a spring shared by Households 2, 6, 8, and 10. Twenty-six AFLP types were identified of which 7 types were represented by more than one isolate. Of these 7 types 4 (57%) consisted of groups of identical *E. coli* isolated from a single sample type (AFLP types A, I, N, P) leaving 3 types that showed similar genetic patterns for *E. coli* isolates between different sampling points within a household or between different households. Type K included identical *E. coli* found in the storage container of Households 6, 8 and 10. In Type M, 10 out of 12 *E. coli* isolates originated from the storage container, the hands and the cup of Household 2. The remaining two isolates originated from the spring (source of Households 2 and 6) and the storage container of Household 6. *E. coli* isolated from the hands of a person in Household 4, the storage vessel of Household 4, and the hands of a person from Household 6 were also identical and belonged to type R.

Discussion

Our results document the genetic relatedness of *E. coli* isolated from storage containers and potential in-house contamination sources such as drinking cups and hands. Isolates were also obtained from contaminated source water used by some of the houses. For one of the households, cow dung used in the house was also sampled. A total of 96 *E. coli* isolates were obtained and typed using fluorescent AFLP. Overall the results showed a considerable degree of genetic diversity among the *E. coli* isolated from the 8 rural households. All but one of the samples yielded *E. coli* belonging to more than one of the AFLP types. The highest diversity was noted for the strains isolated from the storage vessel from Household 1 (Village 1), which showed 8 different AFLP types. The same level of diversity was observed amongst the strains isolated by means of either the membrane method using mFC agar or the Colilert®-18 system.

Isolates that originated from different types of samples collected within a household and which shared the same genetic profile provided a good indication of possible routes of post-collection contamination. In this study such isolates mostly came from closely linked points within the pathway of water use in a single household. For example isolates from the hands, cup and/or storage water could be genetically linked (Fig. 1a, AFLP types: A, and F; Fig 1b, AFLP types: K, M and R). These links are of special importance in households which were using water of good microbial quality. The reason is that the primary route for introduction of the *E. coli* detected in the stored water, almost certainly can be attributed to post-storage contamination. When source water of poor quality was used by households, the AFLP types observed suggested that the *E. coli* detected in these households may have originated from the source water (Fig. 1b, AFLP type M).

In this study only a small number of *E. coli* isolates from a limited number of households were analysed but a high genetic diversity was observed. High genetic diversity levels in *E. coli* studies have also been reported by other researchers (Lu et al., 2004; McLellan et al., 2003). The high genetic diversity observed could be the result of multiple faecal contamination incidents but could also be explained by the persistence of diverse *E. coli* strains in the environment. Milkman (1997) and LeClerc et al. (1996) have noted that recombination is an important and frequent process in *E. coli*. It is thus a feasible assumption that new types with changed genetic patterns are generated during regrowth in the environmental reservoirs (Gordon et al., 2002). Using a target organism such as *E. coli* with potentially high clonal diversity could therefore complicate the genetic analysis of isolates found in secondary habitats (Simpson et al., 2002; Gordon 2001).

Due to the high diversity of the strains many of the *E. coli* isolated from the stored water could not be linked to any of the potential in-house contamination sources investigated. The isolation of corresponding isolates, from environmental samples, might have been obscured by the presence of the high diversity of the other types within the same sample. To overcome this problem extensive isolation and screening of isolates obtained from the environmental samples will have to be performed. Other alternatives included the PCR-based detection of virulence factors (Gordon, 2001), the isolation of other species of enteric bacteria such as *Enterococci* (Hassan et al., 2007), alternative fingerprinting methods such as metabolic fingerprints (Ahmed et al., 2005), and the detection of polymorphisms within a single gene (Soule et al., 2006).

Conclusion

Overall a high degree of genetic diversity for the *E. coli* isolates was observed which may be indicative of multiple sources of *E. coli* organisms in the environment and which could not always be clearly linked with the way water was handled and used in these rural homes. However, in spite of the high genetic diversity, linkages between isolates were observed between closely linked points within the domestic pathway of water handling, e.g. the storage container, hand-swab sample, and the cup sample within a household. The high degree of genetic diversity observed made it difficult to make specific inferences about the exact point of introduction or the exact origin of the *E. coli* contamination observed in the stored drinking water of these households. It confirmed recent conclusions (Gordon, 2001) that it is not feasible to use the relatedness of genetic patterns of commensal *E. coli* to determine the origin of faecal contamination of water.

References

AHMED W, NELLER R and KATOULI M (2005) Host species-specific metabolic fingerprint data base for *Enterococci* and *Escherichia coli* and its application to identify sources of fecal contamination in surface waters. *Appl. Environ. Microbiol.* **71** 4461-4468.

ARNOLD C, METHERELL L, WILLSHAW G, MAGGS A and STANLEY J (1999) Predictive fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: High-resolution typing method with phylogenetic significance. *J. Clin. Microbiol.* **37** 1274-1279.

ASLAM M, NATTRESS F, GREER G, YOST C, GILL C and McMULLEN L (2003) Origin of contamination and genetic diversity of *Escherichia coli* in beef cattle. *Appl. Env. Microbiol.* **69** 2794-2799.

BRADY C, VENTER S, CLEENWERCK I, VANCANNEYT M, SWINGS J and COUTINHO T (2007) A FAFLP system for the improved identification of plant-pathogenic and plant associated species of the genus *Pantoea*. *Syst. Appl. Microbiol.* **30** 413-417.

CARSON CA, SHEAR BL, ELLERSIECK MR and SCHNELL J (2003) Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. *Appl. Env. Microbiol.* **69** 1836-1839.

CHASSEIGNAUX E, TOQUIN MT, RAGIMBEAU C, SALVAT G, COLIN P and ERMEL G (2001) Molecular epidemiology of *Listeria monocytogenes* isolates collected from the environment, raw meat and raw products in two poultry- and pork processing plants. *J. Appl. Microbiol.* **91** 888-899.

CLERC A, MANCEAU C and NESME X (1998) Comparison of randomly amplified polymorphic DNA with amplified fragment length polymorphism to assess genetic diversity and genetic relatedness within genospecies III of *Pseudomonas siringae. Appl. Environ. Microbiol.* **64** 1180-1187.

GUAN S, XU R, CHEN S, ODUMERU J and GYLES C (2002) Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources. *Appl. Environ. Microbiol.* **68** 2690-2698.

GORDON DM (2001) Geographical structure and host specificity in bacteria and implication for tracing the source of coliform contamination. *Microbiol.* **147** 1079-1085.

GORDON DM, BAUER S and JOHNSON R (2002) The genetic populations of *Escherichia coli* populations in primary and secondary habitats. *Microbiol.* **148** 1513-1522.

HAGEDORN C, ROBINSON SL, FILTZ JR, GRUBBS SM, ANGIER TA and RENEAU RB (Jr) (1999) Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Appl. Environ. Microbiol.* **65** 5522-5531.

HARWOOD VJ, WHITLOCK J and WITHINGTON V (2000) Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: Use in predicting the source of fecal contamination in subtropical waters. *Appl. Environ. Microbiol.* **66** 3698-3704.

HASSAN WM, ELLENDER RD and WANG SY (2007) Fidelity of bacterial source tracking: *Escherichia coli* vs. *Enterococcus* spp. and minimizing assignment of isolates from nonlibrary sources. *J. Appl. Microbiol.* **102** 591-598.

JENSEN PK, ENSINK JHJ, JAYASINGHE G, VAN DER HOEK W, CAIRNCROSS S and DALSGAARD A (2002) Domestic transmission routes of pathogens: the problem of in-house contamination of drinking water during storage in developing countries. *Trop. Med. Int. Health.* **7** 604-609.

LECLERC JE, LI B, PAYNE WL and CEBULA TA (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Sci.* **274** 1208-1211.

LIEBANA E, SMITH RP, LINDSAY E, MCLAREN I, CASSAR C, CLIFTON-HADLEY FA and PAIBA GA (2003) Genetic diversity among *Escherichia coli* O157:H7 isolates from bovines living on farms in England and Wales. *J. Clin. Microbiol.* **41** 3857-3860.

LU L, HUME ME, STERNES KL and PILLAI D (2004) Genetic diversity of *Escherichia coli* isolates in irrigation water and associated sediments: implications for source tracking. *Water Res.* **38** 3899-3908.

LU Z, LAPEN D, SCOTT A, DANG A and TOPP E (2005) Identifying host sources of fecal pollution: Diversity of *Escherichia coli* in confined dairy and swine production systems. *Appl. Environ. Microbiol.* **71** 5992-5998.

MARAJ S, RODDA N, JACKSON S, BUCKLEY C and MACLEOD N (2006) Microbial deterioration of stored water for users supplied by stand-pipes and ground-tanks in a peri-urban community *Water SA* **32** (5) 693-699.

MEAYS CL, BROERMA K, NORDIN R and MAJUMDER A (2004) Source tracking fecal bacteria in water: A critical review of current methods. *J. Environ. Manage.* **73** 71-79.

MYODA SP, CARSON CA, FUHRMANN JJ, HAHM BK, HARTEL PG, YANPARA_LQUISE H, JOHNSON L, KUNTZ RL, NAKATSU CH, SADOWSKY MJ and SAMADPOUR M (2003) Comparison of genotypic-based microbiological source tracking methods requiring a host origin database. *J. Water Health* 4167-180.

McLELLAN L, DANIELS AD and SALMORE AK (2003) Genetic characterization of *Escherichia coli* from host sources of faecal pollution by using DNA fingerprinting. *Appl. Environ. Microbiol.* **5** 2587-2594.

MILKMAN R (1997) Recombination and population structure in *Escherichia coli*. *Genetics* **146** 745-750.

RADU S, LING OW, RUSUL G, KARIM MIA and NISHIBUCHI M (2001) Detection of *Escherichia coli* O157H:7 by multiplex PCR and their characterization by plasmid profiling, antimicrobial resistance, RAPD and PFGE analyses. *J. Microbiol. Meth.* **46** 131-139.

SCOTT T, PARVEEN S, PORTIER K, ROSE J, TAMPLIN M, FARRAH S, KOO A and LUKASIK J (2003) Geographical variation in ribotype profiles of *Escherichia coli* Isolates from humans, swine, poultry, beef, and dairy cattle in Florida. *Appl. Env. Microbiol.* **69** 1089-1092.

SIMPSON JM, SANTO DOMINGO JW and REASONER DJ (2002) Microbial source tracking: State of the science. *Environ. Sci. Technol.* **36** 5279-5288.

Available on website http://www.wrc.org.za
ISSN 0378-4738 = Water SA Vol. 34 No. 1 January 2008
ISSN 1816-7950 = Water SA (on-line)

SMITH D, WILLSHAW G, STANLEY J and ARNOLD C (2000) Genotyping of verocytotoxin-producing *Escherichia coli* O157: Comparison of isolates of a prevalent phage type by fluorescent amplified-fragment length polymorphism and pulsed-field gel electrophoresis analyses. *J. Clin. Microbiol.* **38** 4616-4620.

SOUL M, KUHN E, LOGE F, GAY J and CALL DR (2006) Using DNA microarrays to identify library-independent markers for bacterial source tracking. *Appl. Environ. Microbiol*. **72** 1843-1851.

SOBSEY MD (2002) Managing water in the home: Accelerated health gains from improved water supply. WHO/WSH/02.07. WHO, Geneva.

SPEIJER H, SAVELKOUL PH, BONTEN MJ, STOBBERINGH EE and TJHIE JH (1999) Application of different genotyping methods for *Pseudomonas aeruginosa* in a setting of endemicity in an intensive care unit. *J. Clin. Microbiol*. **37** 3654-3661.

STANDARD METHODS (1995) *Standard Methods for the Examination of Water and Wastewater* (19th edn.). American Public Health Association, Washington DC.

STOECKEL DM, MATHES MV, HYER KE, HAGEDORN C, KATOR H, LUKASIK J, O"BRIEN TL, FENGER TW, SAMADPOUR M, STRICKLER KM and WIGGENS BA (2004) Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. *Environ. Sci. Technol*. **38** 6109-6117.

VOS P, HOGERS R, BLEEKER M, REIJANS M, VAN DALEE T, HORNES M, FRIJTERS A, POT J, PELEMAN J, KULPER M and ZABEAU M (1995) AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res*. **23** 4407-4414.

WANG SM, DEIGHTON MA, CAPSTICK JA and GERRATY N (1999) Epidemiological typing of bovine streptococci by pulsed field gel electrophoresis. *Epidemiol. Infect*. **123** 317-324.

WRIGHT JA, GUNDRY SW and CONROY R (2004) Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use. *Trop. Med. Int. Hlth*. **9** 106-117.

Available on website http://www.wrc.org.za
ISSN 0378-4738 = Water SA Vol. 34 No. 1 January 2008
ISSN 1816-7950 = Water SA (on-line)