In Press: 2004. Applied and Environmental Microbiology

Frequency and Distribution of Tetracycline Resistance Genes in Genetically-Diverse, Non-Selected, and Non-Clinical *Escherichia coli* Isolated From Diverse Human and Animal Sources

Andrew Bryan[1†], Nir Shapir[1,2,3,4], and Michael J. Sadowsky[1,2,3,*]

[1]Department of Soil, Water, and Climate, [2]BioTechnology Institute, [3]Center for Microbial and Plant Genomics, and [4]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, St. Paul, MN 55108

Running Title: Tet Genes in Diverse Environmental *E. coli*

*Address of corresponding author: Department of Soil, Water, and Climate, University of Minnesota, 1991 Upper Buford Circle, 439 Borlaug Hall, St. Paul, MN 55108. Phone: (612) 624-2706, Fax: (612) 625-2208, E-mail: sadowsky@umn.edu

†Current Address: Department of Medical Microbiology and Immunology, University of Wisconsin, 1300 University Avenue, 481 MSC, Madison, WI 53706.

Non-selected and natural populations of *Escherichia coli* from 12 animal sources and humans were examined for the presence and types of 14 tetracycline resistance determinants. Of 1,263 unique *E. coli* isolates from humans, pigs, chickens, turkeys, sheep, cows, goats, cats, dogs, horses, geese, ducks, and deer, 31% were highly resistant to tetracycline. Over 78, 47, and 41% of the *E. coli* isolates from pigs, chickens, and turkeys were resistant or highly resistant to tetracycline, respectively, and 61, 29, and 29% of *E. coli* isolates from pig, chickens, and turkeys, respectively, had MIC values ≥233 μg tetracycline per ml. Muliplex PCR analyses indicated that 97% of these strains contained at least one of 14 tetracycline resistance determinants (*tetA, tetB, tetC, tetD, tetE, tetG, tetK, tetL, tetM, tetO, tetS, tetAP, tetQ,* and *tetX*) examined. While the most common determinants found in these isolates were *tetB* (63%) and *tetA* (35%), *tetC*, *tetD*, and *tetM* were also found. *E. coli* isolates from pigs and chickens were the only strains to have *tetM*. To our knowledge, this represents the first report of *tetM* in *E. coli*.

Problems associated with the presence of antibiotic resistant bacteria has reached epidemic proportions in recent years, with cost estimates exceeding $4 billion in the United States alone (6). The spread of antibiotics resistant bacteria in the environment is dependent on the presence and transfer of resistance genes among microorganisms, mutations, and selection pressure to keep these genes in a population, the later neatly provided by the approximately 50 million pounds of antibiotics that are produced and used each year in the United States (14). Only half of these antibiotics are used for humans, while the remainder are administered to animals or other organisms (8). The causes and effects of antibiotic overuse are varied. One of

the most controversial applications of antibiotics, however, is their use for growth promotion in livestock, and this has raised concerns that such use contributes to the presence of resistant bacteria in humans (1, 25).

Tetracyclines have become the drugs of choice to treat *Mycoplasma*- and *Chlamydia*-induced pneumonia (13), and have been used to treat other atypical pneumonias, rickettsial infections, Lyme disease, ehrlichiosis, and other diseases and cancers (23). The clinically useful tetracycline (Tet), chlortetracycline, was introduced in 1948 (24). Only a year later, it was shown that young chickens fed tetracyclines had enhanced growth characteristics (10). However, by 1953, it was reported that *Shigella dysenteriae* had developed resistance to tetracycline antibiotics and by 1955, a *Shigella* sp. strain had developed multidrug resistance (20). Because of that history and the broad clinical use of tetracycline, this antibiotic was chosen, along with commensal strains of *E. coli,* to provide a prototypical view of the use of antibiotics and their effects on bacterial populations (21). Tetracycline is a broad-spectrum antibiotic that inhibits bacterial protein synthesis by preventing aminoacyl-tRNA from binding to the bacterial ribosome (20). Resistance to the antibiotic is conferred by one or more of the 36 currently described *tet* genes, which encode for one of three mechanisms of resistance: an efflux pump, a method of ribosomal protection, or direct enzymatic inactivation of the drug (7). Efflux mechanisms appear to be more abundant among gram-negative microorganisms, while ribosomal protection mechanisms are more common among gram-positives (7). Generally speaking, the rapid spread of tetracycline resistance among bacteria is due to the localization of *tet* genes on plasmids, transposons, and integrons (7,15, 21).

While several studies have examined Tet resistance among bacteria, most have employed clinically-isolated bacteria (4,11,17) or populations specifically isolated for their ability to grow

in the presence of tetracyclines (5,22). These studies, while useful, do not give an unbiased appraisal of the presence and types of *tet* genes that are present in natural (non-clinical), non-selected, populations of bacteria in the environment.

Only a limited number of studies have examined tetracycline resistance determinants in bacteria isolated from a large variety of animal species with different exposure histories to tetracyclines, or in environmental samples (11). While Sengeløv and coworkers (22) examined 100 *E. coli* isolates for the presence of five *tet* resistance determinants and Blake *et al.* (5) used PCR to examine 200 Tet resistant *E. coli* for seven *tet* genes, few have examined a large number of *tet* determinants in non-clinical *E. coli* isolated from a variety of animal species. To better understand the distribution of resistance genes in the environment and to provide insight into selection pressures involved with the use of antibiotics in animal feed, we investigated Tet resistance among natural and unselected populations of *Escherichia coli* from 12 animal sources and humans and determined which resistance genes were present in this population.

Isolates and determination of minimum inhibitory concentration. In order to characterize tetracycline resistance in natural, non-clinical *E. coli* strains from both human and animal sources, 1263 unique isolates were obtained from humans, cat, cow, deer, turkey, duck, sheep, goose, dog, pig, horse, chicken, and goat (Table 1). Fecal materials were collected by swabbing the rectal or cloacal region of individual wild and domesticated animals located throughout Minnesota and western Wisconsin as previously described (9). Fecal samples were kept at 4°C and analyzed within 6 hr of swabbing. Fecal material was streaked onto mFC agar plates (Difco, BD Diagnostic Systems, Sparks, MD), incubated at 44.5°C for 24 hours, and six blue colonies from mFC plates were picked and evaluated using selective and differential growth media as previously described (9). Only isolates giving growth and color responses, on

all media, that were typical for *E. coli* were used in these studies. Three *E. coli* colonies from each individual fecal sample were used for DNA fingerprinting. All isolates were DNA fingerprinted using rep-PCR and BOXA1R primers (9) and identical clones from the same animal were eliminated from analyses. Unique isolates were grown overnight in 150 µl of Luria-Bertani liquid medium in microtiter plates and were spot inoculated, using a multiple inoculator, onto Tryptic Soy Agar (Difco Laboratories, Detroit, MI) supplemented with 0, 5, 10, 20, 40, 70, 93, 117, 175, and 233 µg tetracycline per ml (Sigma Chemicals, St. Louis, MO). Plates were incubated at 37°C overnight and visually examined for growth. Minimum inhibitory concentrations (MIC) were determined from growth patterns and average values are shown in Figure 1. MIC values of <5 µg tetracycline per ml were considered sensitive to the antibiotic, while those having MIC values of 10 – 70, or >90 µg tetracycline per ml were considered resistant or highly resistant, respectively. A MIC of >233 µg/ml was considered as MIC of 233 µg/ml for statistical analysis.

Of the 1263 *E. coli* isolates examined, 31% were resistant to tetracycline (MIC values >10 µg/ml); consisting of 42% from livestock, 21% from humans, 17% from companion animals (cats, dogs, horses), and 4% from wild animals. Over 78, 47, and 41% of the *E. coli* isolates from pigs, chickens, and turkeys were resistant or highly resistant to tetracycline, respectively. Together these resistant isolates represent about 20% of the 1263 isolates examined. In contrast, about 22, 30, 3, 3, 21, 33, 7, 23, 6, and 12.2% of the *E. coli* isolates from cats, cows, deer, duck, humans, sheep, geese, dogs, horses, and goats were resistant or highly resistant to tetracycline, respectively. Moreover, 61, 29, and 29% of *E. coli* isolates from pig, chickens, and turkeys, respectively, had MIC values ≥233 µg tetracycline per ml. In contrast, goats, horses, ducks, geese, and deer had the least number of *E. coli* strains showing resistance or a high level of

resistance to tetracycline. Our results may be explained by the potential exposure of livestock, humans, and companion- and wild-animals to tetracyclines. Tetracycline is often continuously fed to livestock at sub-therapeutic levels for the purpose of growth promotion. For example, up to 70% of U.S. cattle and pig operations use feeds supplemented with antibiotics for growth promotion and the majority are tetracyclines (2). In contrast, humans and companion animals are most often treated therapeutically, for a limited time, for bacterial infections, perhaps reflecting the intermediate level (MIC 10- 70 μg/ml) of resistance to tetracycline. This may be changing, however, as other uses of antibiotics become more common, such as treatment of parasitic and non-infectious diseases (21). The low level of occurrence of tetracycline resistance among isolates from wild animals is presumably due to their low exposure to these antibiotics. Most isolates either had a high level of resistance or none at all, suggesting that the acquisition of a mobile genetic element accounts for resistance.

Epidemiology of *tet* genes. All isolates (325) with a tetracycline MIC of ≥93 μg/ml (which we considered to constitute a high level of resistance) were examined further using multiplex PCR for the presence of *tetA, tetB, tetC, tetD, tetE, tetG, tetK, tetL, tetM, tetO, tetS, tetAP, tetQ,* and *tetX* genes (18). Single-colony isolates were streaked onto Plate Count Agar (Difco, Detroit, MI), picked using disposable 10 μl sterile loops, and were suspended in 50 μl sterile H_2O. One μl of the standardized cell suspensions served as template DNA for colony-based multiplex PCR. The primers used for PCR amplification of the 14 tetracycline resistance genes were as described by Ng *et al.* (18). The primers were aliquoted into four groups: Group I contained primers for *tetB*, *tetC* and *tetD*; Group II contained *tetA*, *tetE* and *tetG*; Group III contained *tetK*, *tetL*, *tetM*, *tetO*, and *tetS*; and Group IV contained primers for *tetA(P)*, *tetQ* and *tetX*. PCR was done in 96 well plates using a MJ Model PTC100 Thermocycler (Waltham, MA), using the following

conditions as described (18): 5 min initial denature at 94°C, followed by 35 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1.5 min. PCR products were separated by gel electrophoresis in 1% (w/v) agarose gels in 1×TAE buffer, stained with ethidium bromide, and visualized under U.V. illumination. The validity of multiplex PCR reactions and product sizes was ascertained by using the following positive control plasmids: pSL18, pRT11,pBR322, pSL106, pSL1504, pJA8122, pAT102, pVB.A15, pJ13, pUOA1, pAT451, pJIR39, pNFD13-2, and pBS5, for *tet* genes A, B, C, D, E, G, K, L, M, O, S, A(P), Q, and X, respectively (18). Sizes of PCR products were determined by comparison to the migration of 100-bp ladder (Gibco, BRL). The identity of all *tet* genes in a representative sample of non-clinical isolates was ascertained by DNA sequencing of PCR products, following extraction from agarose gels. A representative agrose gel of PCR products obtained using primer Group I, amplifying *tetB*, *tetC,* and *tetD*, is shown in Figure 2.

Of the 325 strains analyzed by PCR, 97% contained at least one of 14 (*tetA, tetB, tetC, tetD, tetE, tetG, tetK, tetL, tetM, tetO, tetS, tetAP, tetQ,* and *tetX*) tetracycline resistance determinants. The most common determinants were *tet*(B) (63% of isolates) and *tet*(A) (35% of isolates) (Figure 3). However, *tet*(C), *tet*(D), and *tet*(M) were also found with varying frequencies. The frequencies of *tetA, tetB, tetC,* and *tetD* in the tested isolates (Figure 3) were consistent with those previously reported for lactose-fermenting coliforms using colony hybridization (11). In contrast, Sengeløv and coworkers (22) reported that 71% and 25% of 100 isolates from diseased and healthy pigs, cattle and chickens they tested for five tetracycline resistance determinants contained *tetA* and *tetB*, respectively. None of the tested strains contained genes for *tetE, tetG, tetK, tetL, tetO, tetS, tetAP, tetQ,* or *tetX*. Because in our studies only highly resistant isolates were analyzed by PCR, it is possible that additional resistant genes

were present in the *E. coli* populations, but were non-functional or only provided intermediate or a low-level of resistance.

Isolates from pigs and chickens were the only strains to contain *tetM*, and commonly had more than one tetracycline resistance determinant per strain (Figure 4). *E. coli* from these animals had the greatest number of strains with high MIC values. Over 30% of *E. coli* isolates from turkeys, pigs, and horses contained two Tet resistance determinants, and 4.5% of the pig isolates contained three *tet* genes. However, the presence of more than one resistance determinant did not lead to noticeably higher MIC values. It is possible that strong selection pressures provided by environments containing elevated levels of tetracycline leads to the acquisition of more than one tetracycline gene in a given strain due their prevalence in the environment, rather than a selective advantage. Results of our studies also showed that 22.2% and 1.9% of the isolates contained two and three *tet* genes, respectively. This is in contrast to results from previous studies, in which only 3.5% (16) and 5.4% (22) of isolates had two genes, perhaps due to our use of a larger number and variety of isolates, and the greater number of genes examined.

To our knowledge, this is the first report documenting the presence of the *tetM* gene in *E. coli* (7). Due to the uniqueness of these results, the presence of *tetM* in one of our *E. coli* isolates from pigs was verified by sequencing the PCR product produced using *tetM*-specific primers. Blast analysis (3) indicated that of the 386 bp of high-quality and continuous sequence examined, there was 98% nucleotide sequence identity to the *tetM* gene from *Enterococcus faecalis* (accession number M85225). The *tetM*, which imparts resistance to tetracyclines by encoding a ribosomal protection mechanism, commonly occurs in transposons Tn*916* and Tn*1545*, and is widely dispersed among various gram-positive organisms, but has only rarely

been documented in gram-negative bacteria (19, 21). The presence of *tetM in E. coli* is most likely due to genetic transfer from *Enterococcus*, a common carrier of *tetM* (8). Evidence for this possibility is provided by studies of Poyart, *et al.* (19) who demonstrated the *in vitro* transfer of Tn*916* from *E. faecalis* to *E. coli* (16).

In summary, by examining the frequency and distribution of tetracycline resistance among diverse natural *E. coli* populations present in different animal species, a picture of the selection pressures in the various host animals can be inferred. Those animal hosts that presumably had continuous exposure to tetracycline not only had a greater percentage of tetracycline resistant *E. coli* isolates, but those isolates carried a greater diversity of resistance genes. Moreover, these isolates often had more than one *tet* resistance determinant, and contained a *tet* gene previously thought not to be present in *E. coli*. This suggests that human activity provides environments that select for resistant strains and encourages the transfer of genetic information from unrelated bacterial species. Although this study examined only non-clinical *E. coli* isolates, the prevalence of *tet* resistant genes among these unrelated bacteria, and circumstantial and direct evidence of horizontal gene transfer, suggests that these same resistance determinants may also be present in animal and human pathogens.

ACKNOWLEDGEMENTS

We would like to thank LeeAnn Johnson, John Ferguson, Mary Brown, and Priscilla Dombek for help with the *E. coli* strains. We would also like to thank Dr. Lai-King Ng for the generous gift of positive control strains and cloned *tet* resistance genes.

This study was supported in part from grants from the University of Minnesota Agricultural Experiment Station, and from the Legislative Commission on Minnesota Resources,

through the Environment and Natural Resources Trust Fund and the MN Future Resource Fund (to MJS), and by funding from the Life Sciences Summer Undergraduate Research Program (to AB).

REFERENCES

1. **Aarestrup, F. M., A. M. Seyfarth, H. D. Emborg, K. Pedersen, R. S. Hendriksen, and F. Bager.** 2001. Effect of abolishment of the use of antimicrobial agents for growth promotion on occurrence of antimicrobial resistance in fecal enterococci from food animals in Denmark. Antimicrob. Agents Chemother. **45:**2054-2059.
2. **Akkina, J. E., and R. Johnson.** 1999. Antibiotic Use in U.S. Livestock Production, p. 20-. *In* Antimicrobial Resistance Issues in Animal Agriculture. USDA/APHIS/VS/ CEAHI/CEI, Washington, D.C.
3. **Altschul, S. F., T. L. Madden, A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research **25:**3389-3402.
4. **Arzese, A. R., L. Tomasetig, and G.A. Botta.** 2000. Detection of *tetQ* and *ermF* antibiotic resistance genes in *Prevotella* and *Porphyromonas* isolates from clinical specimens and resident microbiota of humans. J. Antimicrob. Chemother. **43:**1523–1524.
5. **Blake, D. P., R. W. Humphry, K. P. Scott, K. Hillman, D. R. Fenlon, and J. C. Low.** 2003. Influence of tetracycline exposure on tetracycline resistance and the carriage of tetracycline resistance genes within commensal *Escherichia coli* populations. J. Appl. Microbiol. **94:**1087-1097.

6. **Boyce, J. M.** 2001. Consequences of inaction: importance of infection control practices. Clin. Infect. Dis.. **33:**S133-137.

7. **Chopra, I., and M. Roberts.** 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol. Molec. Biol. Rev. **65:**232-260.

8. **DeFlaun, M. F., and S. B. Levy.** 1989. Genes and their varied hosts. *In:* S. B. Levy and R. V. Miller (eds.), Gene Transfer in the Environment, p. 1-32. McGraw-Hill, New York.

9. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66:**2572-2577.

10. **DuPont, H. L., and J. H. Steele.** 1987. The human health implication of the use of antimicrobial agents in animal feeds. Vet. Quart. **9:**309-320.

11. **Guillaume, G., D. Verbrugge, M. Chasseur-Libotte, W. Moens, and J. Collard.** 2003. PCR typing of tetracycline resistance determinants (Tet A-E) in *Salmonella enterica* serotype Hadar and in the microbial community of activated sludges from hospital and urban wastewater treatment facilities in Belgium. FEMS Microbiol. Ecol. **32:**77-85.

12. **Jones, M. E., J. A. Karlowsky, D. C. Draghi, C. Thornsberry, D. F. Sahm, and D. Nathwani.** 2003. Epidemiology and antibiotic susceptibility of bacteria causing skin and soft tissue infections in the USA and Europe: a guide to appropriate antimicrobial therapy. Int. J. Antimicrob. Agents. **22:**406-419.

13. **Lenart, J., A. A. Andersen, and D. D. Rockey. 2002**. Growth and development of tetracycline-resistant *Chlamydia suis*. Antimicrob. Agents Chemother. **45:**2198-2203.

14. **Levy, S. B.** 2001. Antibiotic resistance: consequences of inaction. Clin. Infect. Dis. **33:**S124-129.

15. **Levy, S. B., G. B. Fitzgerald, and A. B. Macone.** 1976. Changes in intestinal flora of farm personnel after introduction of tetracycline-supplemented feed on a farm. N. Eng. J. Med. **295:**583-588.

16. **Marshall, B., C. Tachibana, and S. B. Levy.** 1983. Frequency of tetracycline resistance determinant classes among lactose-fermenting coliforms. Antimicrob. Agents Chemother. **24:**835-840.

17. **Mathew, A. G., D. B. Arnett, P. Cullen, and P. D. Ebner.** 2003.Characterization of resistance patterns and detection of apramycin resistance genes in *Escherichia coli* isolated from swine exposed to various environmental conditions. Int. J. Food Microbiol. **89:**11-20.

18. **Ng, L. K., I. Martin, M. Alfa, and M. Mulvey.** 2001. Multiplex PCR for the detection of tetracycline resistant genes. Molec. Cell. Probes. **15:**209-215.

19. **Poyart, C., J. Celli, and P. Trieu-Cuot.** 1995. Conjugative transposition of Tn916-related elements from *Enterococcus faecalis* to *Escherichia coli* and *Pseudomonas fluorescens*. Antimicrob. Agents Chemother. **39:**500-506.

20. **Roberts, M. C.** 1996. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. FEMS Microbiol Rev **19:**1-24.

21. **Roberts, M. C.** 2003. Tetracycline therapy: update. Clin. Infect. Dis. **36:**462-467.

22. **Sengeløv, G., B. Halling-Sørensen, and F. M. Aarestrup.** 2003. Susceptibility of *Escherichia coli* and *Enterococcus faecium* isolated from pigs and broiler chickens to

tetracycline degradation products and distribution of tetracycline resistance determinants in *E. coli* from food animals. Vet. Microbiol. **95:**91-101.

23. **Smilack, J. D.** 1999. The tetracyclines. Mayo Clin. Proc. **74:**727-729.
24. **U.S. Congress, Office of Technology Assessment.** 1995. Impacts of Antibiotic-Resistant Bacteria, OTA-H-629, GPO 052-003-01446-7. Government Printing Office, Washington, D.C.
25. **Wegener, H. C., F. M. Aarestrup, P. Gerner-Smidt, and F. Bager.** 1999. Transfer of antibiotic resistant bacteria from animals to man. Acta Vet. Scand. Suppl. **92:**51-57.

Table1. *E. coli* isolates used in this study and their animal sources.

Animal source of *E. coli*	No. isolates used for MIC analysis	No. isolates used for multiplex PCR
Cat	46	9
Cow	158	24
Deer	74	1
Turkey	82	30
Duck	70	1
Human	176	30
Sheep	48	15
Goose	122	3
Dog	47	9
Pig	182	131
Horse	66	3
Chicken	151	66
Goat	41	3
Total	1263	325

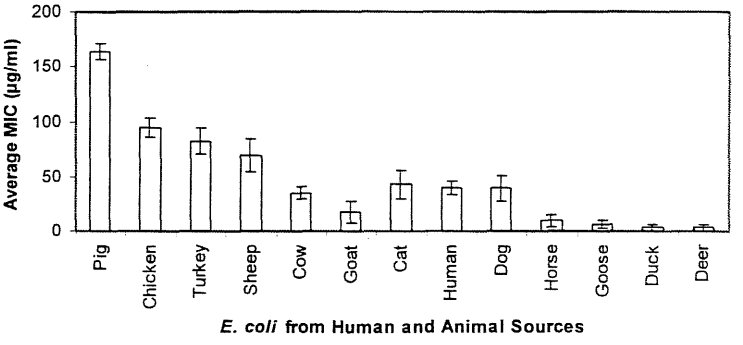
FIGURE LEGENDS

Figure 1. Average minimum inhibitory concentration (MIC) of tetracycline for *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, cats, humans, dogs, horses, geese, ducks, and deer, as determined by the plate dilution method.

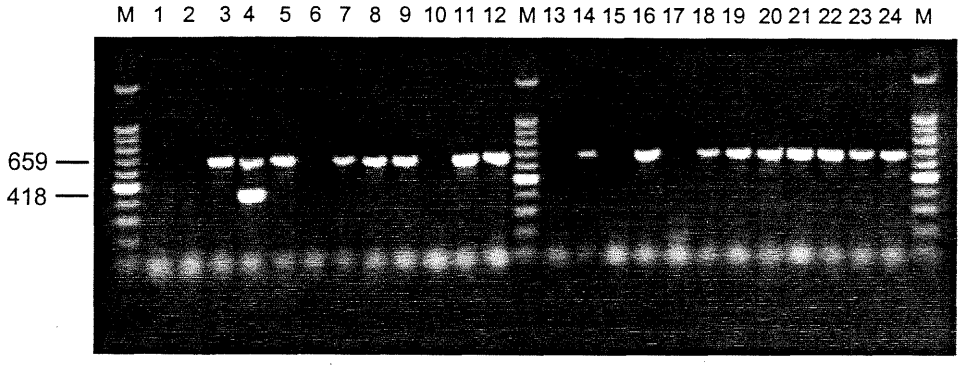
Figure 2. Representative agarose gel of PCR products from non-clinical *E. coli* using primer Group I, containing primers for *tetB*, *tetC*, and *tetD*. Lanes: 1, no template control; 2, *E. coli* H25; 3, *E. coli* H45; 4, *E. coli* H77; 5, *E. coli* P282; 6, *E. coli* P284; 7, *E. coli* P285; 8, *E. coli* P286; 9, *E. coli* P289; 10, *E. coli* P290; 11, *E. coli* P291; 12, *E. coli* P293; 13, *E. coli* P294; 14, *E. coli* P295; 15, *E. coli* P296; 16, *E. coli* P297; 17, *E. coli* P298; 18, *E. coli* P300; 19, *E. coli* P304; 20, *E. coli* P307; 21, *E. coli* P308; 22, *E. coli* P309; 23, *E. coli* P310; and 24, *E. coli* P312. *E. coli* isolate numbers beginning with P and H were isolated from pigs and horses, respectively. Molecular weight markers (100 bp ladder) are in lanes designated M. Sizes of amplicons in base pairs are indicated in the margins.

Figure 3. Frequency of *tetM, tetA, tetD, tetC,* and *tetB* in *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, cats, humans, dogs, horses, geese, ducks, and deer, as determined by colony multiplex PCR. The tetracycline genes *tetE, tetG, tetK, tetL, tetO, tetS, tetA(P), tetQ,* and *tetX* were not found among any of the 325 *E. coli* isolates tested.

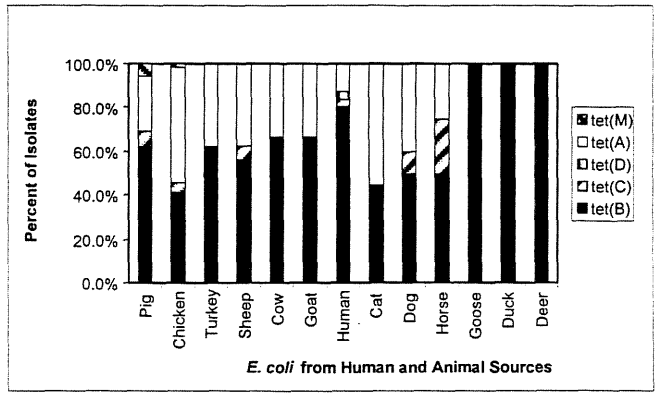
Figure 4. Percent of *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, cats, humans, dogs, horses, geese, ducks, and deer, containing multiple tetracycline resistance genes as determined by multiplex PCR using primers for *tetA, tetB, tetC, tetD, tetE, tetG, tetK, tetL, tetM, tetO, tetS, tetA(P), tetQ,* and *tetX.*



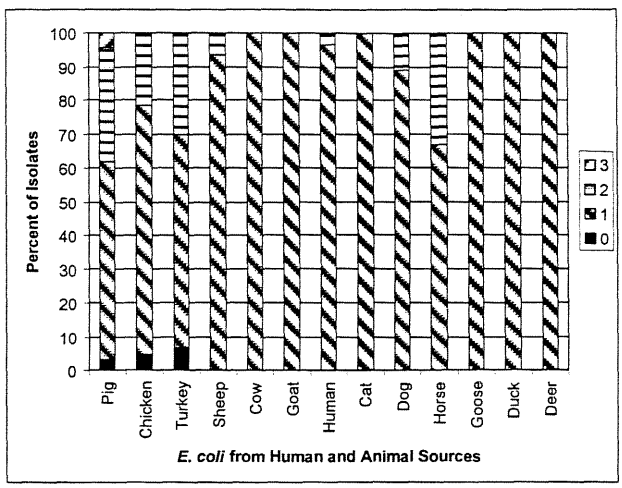
Bryan *et al.* Figure 1



Bryan *et al.* Figure2



Bryan *et al.* Figure 3



Bryan *et al.* Figure 4

Submitted Manuscript – DO NOT CITE

Duplication of Genotypes in DNA Fingerprint Libraries and a High Degree of Genetic Diversity Among Natural Populations of *Escherichia coli* from Different Animals Influences Accuracy of Determining Sources of Fecal Pollution

LeeAnn K. Johnson[†], Mary B. Brown, Ethan A. Carruthers, John A. Ferguson, Priscilla E. Dombek and Michael J. Sadowsky[*]

Department of Soil, Water, and Climate, University of Minnesota, St. Paul, MN 55108

Running Title: Diversity Among Natural *E. coli*

[*]Author to whom correspondence should be addressed:

Michael J. Sadowsky
University of Minnesota
Dept. of Soil, Water, and Climate
1991 Upper Buford Circle, 439 Borlaug Hall
St. Paul, MN 55108
Phone: 612-624-2706
FAX: 612-625-2208
Email: Sadowsky@soils.umn.edu

[†]Current Address: Minnesota Department of Agriculture, 90 West Plato Boulevard, Saint Paul, Minnesota 55107

ABSTRACT

A horizontal, fluorophore-enhanced, rep-PCR DNA fingerprinting technique (HFERP) was developed and evaluated as a means to differentiate human from animal sources of *Escherichia coli*. Box A1R primers and PCR was used to generate 2,466 rep-PCR and 1,531 HFERP DNA fingerprints from *E. coli* strains isolated from fecal material from known human and 12 animal sources: dogs, cats, horses, deer, geese, ducks, chickens, turkeys, cows, pigs, goats, and sheep. HFERP DNA fingerprinting reduced within gel grouping of DNA fingerprints and improved alignment of DNA fingerprints between gels, relative to that achieved using rep-PCR DNA fingerprinting. Jackknife analysis of the complete rep-PCR DNA fingerprint library, done using Pearson's product-moment correlation coefficient, indicated that 69.3-97.1% of animal and human isolates were assigned to the correct source groups, with a 82.2% average rate of correct classification. However, when only unique isolates were examined, isolates from a single animal having a unique DNA fingerprint, Jackknife analysis showed that 44.3–73.8% of the isolates were assigned to the correct source groups, with a 60.5% average rate of correct classification. The percentage of correctly classified isolates were about 15 and 17% greater for rep-PCR and HFERP, respectively, when analyses were done using the curve-based Pearson's product-moment correlation coefficient, rather than the band-based Jaccard algorithm. Rarefaction analysis indicated that despite the relatively large size of the known source database, genetic diversity in *E. coli* was very great, and is most likely accounting for our inability to correctly classify many environmental *E. coli* isolates. Taken together, our data indicates that duplication of genotypes within the DNA fingerprint library, database size, method of statistical analysis, and alignment of band data within and between gels impacts the accuracy of microbial source tracking methods.

INTRODUCTION

Protection of humans from pathogen contamination is dependent on the purity of waters designated for recreation, drinking, and shellfish harvesting. Bacterial pathogens have been listed as the second leading cause of impairment of rivers and streams, and the leading pollutant in estuaries (53). Restoration of impaired waters is currently being accomplished through the development of Total Maximum Daily Loads (TMDLs). Source assessment is an important component of TMDL development in which pollutants are identified and characterized by type, magnitude, and location (54). The implementation of TMDLs has provided one of the driving forces for the development of methods to distinguish between human and animal sources of fecal pollution. Sources of fecal coliform bacteria may include runoff from feedlots and manure-amended agricultural land, wildlife, inadequate septic systems, urban runoff, and sewage discharges. The ability to distinguish between human and animal sources of fecal contamination will be an important assessment tool for the evaluation of possible health risks and for the development of effective control strategies.

Both phenotypic and genotypic methods have been explored as means to study the ecology of fecal bacteria related to host specificity, and determining potential sources of fecal bacteria found in surface water (9,45,48). The mostly widely investigated bacteria for these studies have been *Escherichia coli* and *Enterococcus* sp. strains. Phenotypic approaches that have been explored to date include: fecal coliform/fecal streptococci ratios (10), antibiotic resistance profiles (15,16,21,37,61,62), coliphage typing (23,35), *Bacteroides* phage typing (39,51), and sorbitol-fermenting *Bifidobacterium* (31). In contrast, genotypic approaches including ribotyping (6,7,25,38,44), pulsed field gel electrophoresis (33,36,47), rep-PCR DNA fingerprinting (7,9,33), multilocus enzyme electrophoresis and virulence factors (12), 16S rRNA

analysis (15) and amplified fragment length polymorphism analysis (2,15), and PCR analysis of host specific 16S rDNA fragments from members of the genus *Bifidobacterium* and the *Bacteroides-Prevotella* group (3), have also been investigated as means to determine sources of fecal bacteria.. The use of these methods is based on the hypothesis that specific strains, or a strain's phenotypic or genetic attributes, are related to specific host animals (e.g. bacteria from the intestinal tracts of humans have a greater propensity to associate with humans than other animal species). This hypothesis, however, has only been tested in a limited manner with a minimum number of host animals and a minimal number of bacterial strains.

The majority of phenotypic and genotypic methodologies require the construction of known-source libraries (a host origin database) to differentiate among isolates, which is subsequently used to determine the host origin of unknown environmental isolates (48). However, in most cases, the size of the host origin databases are rather limited, consisting of 35 to about 500 isolates (1,5-7,9,15,18,19,33,34,36,37,44,46,59,60), making broader comparisons to larger populations of *E. coli* and *Enterococcus* in the environment difficult. In addition, temporal and geographic variation in bacterial genotypes within and between animal species (11,18,25,44), multiple strains within a single animal (33), and diet variation within a host animal (19) have been shown to influence the comprehensiveness of known source libraries. Moreover, while microbial source tracking studies done using phenotypic approaches and antibiotic resistance patterns have frequently used large known-source libraries, consisting of about 1000 – 6,000 isolates (5,14,16,21,61-63), many of the strains examined were isolated from the same source material or sample, and thus libraries may be biased due to the presence of multiple replications (clones) of the same bacterial genotype.

The rep-PCR DNA fingerprinting technique uses the polymerase chain reaction and primers based on highly conserved and repetitive nucleotide sequences to amplify specific portions of the microbial genome (24,32,42,50,55,57,58). When the PCR products are separated by agarose gel electrophoresis and visualized following staining with ethidium bromide, the resulting banding patterns produce a "fingerprint" unique to each strain. Bacteria having identical fingerprints are regarded as being the same strain, and those having nearly identical or similar banding patterns are regarded as being genetically related. While rep-PCR has proven to be a valuable tool to identify and track medically and environmentally important microorganisms (8,26,43,55), it has also been recently evaluated for its use as a source-tracking tool (1,7,9,29,33). The rep-PCR DNA fingerprinting technique is relatively quick, easy, and inexpensive to perform, and lends itself to high throughput applications, making it an ideal method for microbial source-tracking studies.

Initial studies done in our laboratory indicated that rep-PCR done with Box A1R primers and *E. coli* yielded a more consistent and complex DNA fingerprints than did studies done using REP primers (9). However, rep-PCR reactions done with Box, ERIC, and REP primers have all been evaluated in microbial source-tracking studies (1,7,9,33). Dombek *et al.* (9) used a minimal data set consisting of about 200 non-unique *E. coli* isolates and reported that 100% of chicken and cow isolates, and between 78-90% of human, goose, duck, pig and sheep isolates were correctly assigned to host source groups using rep-PCR DNA fingerprinting and Box AIR primers. Similarly, Carson *et al.* (7) reported that rep-PCR DNA fingerprinting done using Box A1R produced a 96.6% average rate of correct classification (ARCC) for human and non-human *E. coli* isolates, and McLellan *et al.* (33) reported a 79.3% ARCC for *E. coli* analyzed using rep-PCR done and REP primers.

While all these initial analyses indicated that the rep-PCR technique may be useful for determining animal sources of *E. coli*, these studies were done with relatively small datasets. Moreover, since rep-PCR, and most other genotypic methods, require the construction of libraries of known-source fingerprints, which is labor-intensive and time-consuming, it is very important that the fingerprint database is unbiased, and representative of the diversity of *E. coli* potentially present in animal hosts and in environmental samples. Furthermore, since the database itself can be influenced by many factors (52), including the reproducibility and alignment of DNA fingerprint patterns between and within gels, these variables need to be minimized by using highly standardized protocols and by avoiding known problem conditions.

Binary similarity coefficients are used to analyze presence/absence data (28), and simple banding data obtained from DNA fingerprints can be analyzed using binary coefficients such as Dice or Jaccard band matching algorithms. However, more quantitative algorithms, such as Pearson's product-moment correlation coefficient, can also be applied to complex DNA banding patterns, such as those found using rep-PCR. In this case, fingerprints are analyzed as densitometric curves, taking into account both peak position and height (intensity) (17).

In this study we created a large-scale, known-source rep-PCR DNA fingerprint database from 2,466 *E. coli* isolates obtained from 13 animal sources: cows, pigs, sheep, goats, turkeys, chickens, ducks, geese, deer, horses, dogs, cats, and humans. The database was assembled using a new fingerprinting method, horizontal, fluorophore-enhanced, rep-PCR (HFERP), and the usefulness of this method to differentiate human from animal sources of fecal *E. coli* was evaluated.

MATERIALS AND METHODS

Isolation of *E. coli* from known animal sources. Fecal samples, representing 13 animal source groups, were collected from wild and domesticated animals throughout Minnesota and western Wisconsin. Fresh fecal material was collected from individual animals as previously described (9) by swabbing the rectal or cloacal region using a Culturette7 swab transport system (BD Diagnostic System, Sparks, MD), or by collecting freshly voided feces with a sterile tongue depressor. Fecal samples were placed into sterile Whirl-Pak® bags (Nasco, Fort Atkinson, WI) and kept at 4°C until processed, usually within 6 hr. Fecal material was streaked onto mFC agar plates (Difco, BD Diagnostic Systems, Sparks, MD), and incubated at 44.5°C for 24 hours. Characteristic blue colonies (usually six) from mFC plates were picked and evaluated using selective and differential media as previously described (9). Isolates were used for subsequent studies if growth and color responses on all media were typical for *E. coli*. Isolates giving atypical responses with any test were further screened using API 20E test kits (bioMerieux, Inc., St. Louis, MO). Isolates yielding a "good" to "excellent" *E. coli* identification by the API 20E kit were used for DNA fingerprinting. Three *E. coli* colonies from each individual fecal sample were used for DNA fingerprinting and were stored at –80°C in 50% glycerol.

***E. coli* preparation and rep-PCR conditions**. *E. coli* isolates were streaked onto Plate Count Agar (Difco, BD Diagnostic Systems, Sparks, MD) and grown overnight at 37°C. Colonies were picked with a 1 µl sterile inoculating loop (Fisher Scientific, Pittsburgh, PA), suspended in 100 µl of distilled H_2O in 96-well microtiter plates, and 2 µl of the resulting suspension was used as template for PCR. The rep-PCR fingerprints were obtained using the Box A1R primer: 5'-

CTACGGCAAGG CGACGCTGACG-3', and PCR reactions were done as described previously (9,40,41). PCR was performed using an MJ Research PTC 100 (MJ Research, Waltham, MA) using the protocol specific for this thermocyclers and the Box A1R primer. PCR was initiated with an incubation at 95°C for 2 minutes, followed by 30 cycles, consisting of 94°C for 3 seconds, 92°C for 30 seconds, 50°C for 1 minute, and 65°C for 8 minutes (40). PCR reactions were terminated after an extension at 65°C for 8 min, and stored at 4°C. Reactions that were not used immediately for gel electrophoresis analysis were stored at -20°C.

Electrophoresis was done at 4°C for 17-18 hours at 70V with constant buffer recirculation (9,40). Gels were stained for 20 min in 0.5 μg/ml ethidium bromide prepared in 0.5× TAE buffer. Gel images were captured as TIF files using a FOTO/ Analyst Archiver electronic documentation system (Fotodyne Inc., Hartland, WI).

HFERP studies. Horizontal, fluorophore-enhanced, rep-PCR (HFERP) analyses were performed as follows: *E. coli* colonies were picked with a 1 μl sterile inoculating loop (Fisher Scientific, Pittsburgh, PA), suspended in 100 μl of 0.05 M NaOH in 96-well, low profile, PCR plates (MJ Research, Waltham, MA), heated to 95 °C for 15 min, and centrifuged at 640 RPM for 10 min in a Hermle/Labnet Z383K centrifuge. A 2 μl aliquot of the supernatant in each well was used as template for PCR using the protocol described above for rep-PCR. The primer consisted of a mixture of 0.09 μg of unlabeled Box A1R primer per μl and 0.03 μg of 6-FAM fluorescently labeled Box A1R primer per μl (Integrated DNA Technologies, Coralville, IA). The primer mixture was used at a final concentration of 0.12 μg per 25 μl PCR reaction. A 6.6 μl aliquot of a mixture of 50 μl Genescan-2500 ROX internal lane standard (Applied Biosystems, Foster City, CA) and 200 μl non-migrating loading dye (150 mg Ficoll 400 per ml, and 25 mg blue dextran per ml) was added to each 25 μl PCR reaction prior to loading the PCR reaction

into agarose gels, 12 µl of the resulting mixture was loaded per gel lane. DNA fragments were separated as described for rep-PCR, and HFERP images were captured using a Typhoon 8600 Variable Mode Imager (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) operating in the fluorescence acquisition mode using the following settings: green (532 nm) excitation laser; 610 BP 30 and 526 SP emission filters in the autolink mode with 580 nm beam splitter; normal sensitivity; 200 micron/pixel scan resolution; + 3 mm focal plane; and 800 V power.

Computer-assisted rep-PCR fingerprint analysis. Separated gel images (ROX-stained standards and HFERP banding patterns) were processed using ImageQuant image analysis software (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) and converted to 256 gray scale TIF images. Gel images were normalized and analyzed using BioNumerics v.2.5 software (Applied-Maths, Sint-Martens-Latem, Belgium). Rep-PCR gel lanes were normalized using the 1 kb ladder from 298 bp to 5090 bp, as external reference standards, while HFERP gel lanes were normalized using the Genescan-2500 ROX internal lane standard from 287 bp to 14,057 bp. Band matching for rep-PCR DNA fingerprints was accomplished by using the following BioNumerics settings: minimum profiling 5%, gray zone 5%, minimum area 0%, and shoulder sensitivity of 5; while band matching for HFERP DNA fingerprints was done by using 3% minimum profiling, 0% gray zone, 0% minimum area, and 0 shoulder sensitivity. DNA fingerprint similarities were calculated by using either the curve-based cosine or Pearson's product-moment correlation coefficient, with 1% optimization, or the band-based Jaccard coefficient. Dendrograms were generated using the unweighted pair-group method using arithmetic means (UPGMA). The percentages of known-source isolates assigned to their correct source group were calculated by using Jackknife analysis, with maximum similarities.

RESULTS

Evaluation of isolates. Most genotype-based bacterial source tracking methods rely on the construction and use of libraries of known-source fingerprints. Library construction is time consuming and expensive. It is often assumed, however, that isolates from sewage or fecal materials that grow on selective and differential media are bona fide *E. coli* or *Enterococcus* sp. strains. Of the 2,672 *E. coli* strains obtained from known human and animal sources using an array of selective and differential plating media, 219 isolates gave at least one atypical result when examined by routine biochemical screening tests. The biochemical characteristics of these isolates were examined further by using the API 20E system. Results of this analysis indicated the majority of these isolates, 167, were bona fide *E. coli,* while the remainder, 52, could not be confirmed as this bacterium. The latter group was not used in rep-PCR analysis or included in the DNA fingerprint database. This result indicates that it is important to confirm the identity of bacteria used in source tracking libraries, rather than relying solely on growth or reactions on selective/differential plate media.

Influence of duplicate *E. coli* strains on classification of known source library. While we previously described the use of rep-PCR DNA fingerprinting to determine sources of fecal bacteria (9), our initial studies, and many others, used libraries consisting of a relatively small number of samples, some of which were obtained from the same individual animal. Since results from several studies suggest that *E. coli* is genetically diverse and clonal in origin, and that this may influence the usefulness of this bacterium for source tracking studies (11), we evaluated this technology using a large library of *E. coli* obtained from 13 human and animal sources collected throughout Minnesota and Western Wisconsin (Table 1).

2,466 high-quality rep-PCR DNA fingerprints were generated using the Box A1R primer and template DNA from *E. coli* strains obtained from the 13 human and animal sources (Table 1). Jackknife analysis performed on the 2,466 DNA fingerprints from the entire known-source rep-PCR DNA fingerprint database, using Pearson's product-moment correlation coefficient, indicated that 69-97% of animal and human *E. coli* isolates were assigned into correct source groups (Table 2). This corresponds to an 82.2% average rate of correct classification for the 2,466 rep-PCR DNA fingerprints.

However, since identical DNA fingerprints from *E. coli* strains obtained from the same individual most likely represent isolates of clonal origin, and can artificially bias subsequent analyses of strain groupings (e.g. increase the average rate of correct classification) and the fidelity of the database, we eliminated duplicate DNA fingerprints originating from *E. coli* strains obtained from the same individual animal or human. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%.

Of the 2,466 DNA fingerprints analyzed, 1,535 (62%) remained in the "unique" DNA fingerprint library (Table 1). The influence of duplicate DNA fingerprints on the correct classification of library strains is shown in Table 2. When the 1,535 DNA fingerprints from the unique *E. coli* isolates were examined, Jackknife analyses indicated that only 44-74% of the isolates were assigned to the correct source group (Table 2). The average rate of correct classification for these 1,535 unique rep-PCR DNA fingerprints was 60.5%. Taken together, these results indicate that inclusion of duplicate DNA fingerprints in the library can artificially influence strain groupings and increase percentages of strains correctly assigned to source groups.

Influence of library size on usefulness of DNA fingerprint libraries. We also determined whether *E. coli* isolates obtained in this study were sufficient to capture the genetic diversity present within the *E. coli* populations sampled. *E. coli* isolates with rep-PCR DNA fingerprint similarities of 90% or greater (based on cosine coefficient, 1% optimization and UPGMA) were assigned to the same genotype. By this definition, 657 genotypes were identified from the 1,535 unique *E. coli* isolates in the known-source database. The isolates were randomized, and a rarefaction curve was constructed by summing the number of genotypes that accumulated with the successive addition of isolates. Results in Figure 1 show that despite a library size of 1,535 DNA fingerprints, genetic diversity has not been saturated. This is evidenced by the apparent first order relationship between isolate numbers (sampling effort) and accumulation of new genotypes. Moreover, 58.75% of the genotypes from isolated strains, across all animal groups, occurred only once in the database and a limited number occurred multiple times (Figure 2). Consequently, such a library is most likely not optimal for determining sources of unknown fecal bacteria from water, and if used would result in a large proportion of environmental strains not being classified to correct source groups.

HFERP DNA fingerprinting. In our studies we noted that cluster analysis of rep-PCR DNA fingerprint data often produced groupings that were more related to the gels from which they originated, than the host animal from which they were isolated. We hypothesized that within-gel clustering of DNA fingerprints was in part due to intrinsic gel-to-gel variation, differential DNA migration in repeated runs of the same and different PCR samples, and the inability to correct for heat and buffer-induced gel distortion across and between single and multiple gels. Since DNA fingerprint libraries are assembled from many different gels, this could have a major impact on the fidelity of DNA fingerprint libraries and their subsequent use for

tracking sources of unknown fecal bacteria. To overcome these major limitations, we developed and evaluated the use of a horizontal, fluorophore-enhanced, rep-PCR (HFERP) technique as a means to differentiate human from animal sources of fecal bacteria. In this method, alignment, correction, and normalization of fluorescently-labeled, rep-PCR DNA fingerprint bands within and between gels is facilitated by the use of internal ROX-labeled molecular weight markers that are present in each lane. The technique is similar to that previously described for use with a DNA sequencer (56), but instead uses a standard horizontal agarose gel and a dual-wavelength scanner. An example of an unseparated HFERP gel displaying the ROX-labeled internal lane standard and 6-FAM-labeled Box A1R DNA fingerprints is shown in Figure 3A, and the separated gel images are shown in Figures 3B and 3C. Typically, and with our *E. coli* strains, 12 to 20 DNA bands per strain were revealed using the HFERP technique.

To test whether HFERP reduced within-gel groupings of DNA fingerprints, we analyzed DNA fingerprints from 40 *E. coli* strains obtained from dogs on 2 different gels using Pearson's product-moment coefficient. Results of these studies indicated that rep-PCR DNA fingerprints from strains run on the same gel were, on average, 50% (range 29 – 57%) more likely to be grouped together as the same strains analyzed by using the HFERP technique (data not shown). This indicates that HFERP method considerably reduces within gel grouping of DNA fingerprints. In addition, the HFERP method reduced alignment difficulties due to within- and between-gel variation in band migration found with rep-PCR gels (Figure 4).

The repeatability of the rep-PCR and HFERP DNA fingerprinting methods was examined by fingerprinting a single, reference, control *E. coli* strain (pig isolate number 294) that was included on each gel. DNA fingerprints from 29 and 41 repetitions of *E. coli* control pig strain 294, each from a separate gel, were generated by using the rep-PCR and HFERP methods,

respectively. When analyzed using the curve-based Pearson's correlation coefficient, the rep-PCR DNA fingerprints had an average similarity of 88%, whereas the HFERP-derived DNA fingerprints had an average similarity of 92%. Taken together, our results indicate that the HFERP technique has the ability to improve alignment of bands and the repeatability of banding patterns across different gels that are used to create DNA fingerprint libraries. This technology will have application to other DNA fingerprinting methods that rely on the use of PCR primers.

Analysis of HFERP-generated DNA fingerprint library. Of the 1,535 previously selected unique *E. coli* isolates from animals and humans (Table 1), 1,531 were subjected to HFERP DNA fingerprinting using a combination of fluorescently labeled and unlabeled Box A1R PCR primers. Jackknife analyses of HFERP gels done using the curve-based Pearson's correlation coefficient indicated that 38-73% of the isolates were assigned to the correct source group using this technique (Table 3). For the curve-based analysis, the HFERP technique had the lowest percent of correctly classified strain in cases where the numbers of analyzed fingerprints were relatively small (for sheep, horses, and goats). The average rate of correct classification for the unique HFERP-generated DNA fingerprints was 59.9%.

In contrast, Jacknife analyses of HFERP-generated DNA fingerprints done using the band-based Jaccard analysis showed that only 8-56% of the *E. coli* isolates were assigned to the correct source group, with a 43.0% average rate of correct classification. This indicates that for this type of data, the Pearson's product-moment correlation coefficient was superior to Jaccard's band matching algorithm for assigning known isolates to the correct source groups. Interestingly, results in Table 3 also show that despite problems associated with within- and between-gel variation, within-gel grouping of isolates, and repeatability issues, Jacknife analysis of rep-PCR DNA fingerprints, analyzed using Pearson's correlation coefficient, indicated that 48-74% of the

isolates were assigned to the correct source group, a 60.9% average rate of correct classification. Analysis of rep-PCR DNA fingerprint data using the Jaccard band-based method was not as useful in separating *E. coli* isolates into their correct source group as was the curve-based method.

Groupings of fingerprint data. In some instances, it may be sufficient to identify unknown watershed *E. coli* isolates to larger groupings, rather than to individual animal types. To determine if the HFERP-generated DNA fingerprint data from our library of unique *E. coli* isolates grouped well into larger categories, we assembled DNA fingerprints from pets (dogs and cats), domesticated animals (chickens, cows, goats, horses, pigs, sheep, and turkeys), wild-life (deer, ducks, and geese), and humans, and used Jacknife analysis to assess the percent of correctly classified strains. Results in Table 4 show that the HFERP DNA fingerprints, analyzed using Pearson's product-moment correlation coefficient, correctly classified 83.2, 53.8, 71.4, and 59.1% of the isolates into the domesticated, human, wildlife, and pet categories, respectively. The average rate of correct classification for these groups was 74.3%. However, when DNA fingerprints were analyzed using Jaccard's coefficient, the average rate of correct classification was 66.2%. As before, the least precision was found in categories having the smallest number of fingerprints, pets and humans, suggesting that there is an apparent relationship between the number of fingerprints analyzed and the percentage of correctly classified isolates.

In microbial source tracking studies it may often be useful to determine if unknown isolates belong to either animal or human source groups, rather than to more specific categories. Results in Table 5 show that about 94% and 54% of *E. coli* from animals and humans, respectively, were assigned to the correct source groups using HFERP-generated DNA fingerprints and Pearson's correlation coefficient. The average rate of correct classification was

88.2 and 86.1% for analyses done using Pearson's and Jaccard's algorithms, respectively. The lower percentage of correctly classified human isolates may, in part, be due to the smaller size of fingerprints analyzed for this category.

DISCUSSION

The purpose of this study was to develop a large-scale known-source *E. coli* rep-PCR DNA fingerprint library that can be used to identify the source of *E. coli* bacteria isolated from impacted watersheds in Minnesota. The known-source DNA fingerprint library included 2,466 *E. coli* isolates obtained from nearly 1,000 individuals belonging to thirteen source groups: cows, pigs, sheep, goats, turkeys, chickens, ducks, geese, deer, horses, dogs, cats, and humans. Earlier work in our laboratory, examining a much smaller subset of *E. coli* isolates, indicated that the rep-PCR technique had the necessary sensitivity and resolving power to differentiate between strains of fecal coliform bacteria originating from different human and animal sources (9). However, in our earlier studies, and those done by several researchers, the size of the host origin databases were limited, consisting of 35 to about 500 isolates (1,5-7,9,15,18,19,33,34,36,37,46, 59,60). The relatively small size of these libraries do not take into account the tremendous genetic diversity present in *E. coli* (11,20,33) and enterococci, and makes broader comparisons to larger populations of these organisms in the environment difficult.

In our studies reported here we show that increasing the size of the known source library to 2,466 isolates did not necessarily lead to an increase in the ability to correctly assign strains to the correct source group. In fact, the average rate of correct classification decreased 4.2% using the larger library reported here, relative to what was seen using a smaller library in our previous

studies (9). This may in part be due to the uncovering of increased genetic diversity among isolates, increased accumulation of errors due to gel-to-gel variation, or the presence of duplicate genotypes (DNA fingerprints) from the same individual within our original library. Reduction in the percentage of known-source *E. coli* isolates that were correctly classified was especially apparent when our unique library of 1535 *E. coli* isolates was examined. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%. Since DNA fingerprints from *E. coli* strains obtained from the same individual represent isolates of clonal origin, these duplicate strains (or fingerprints) can artificially bias the average rate of correct classification and the fidelity of the database. Results in Table 2 show that there was a 21.7% reduction in the average rate of correct classification by using the unique DNA fingerprint library, relative to that seen with the complete library. Moreover, the 60.5% average rate of correct classification found with the unique library was less than we previously reported using a smaller library of *E. coli* strains containing duplicate DNA fingerprints from the same individual animal (9), and less than reported by other authors using libraries containing duplicate entries (7,33). More importantly, our results show that failure to remove identical fingerprints from analyses resulted in an overestimation of the ability of the database to assign isolates to their correct source group, perhaps in part due to the clonal composition of *E. coli* populations (11,20,33). Taken together, our results indicate that inclusion of duplicate DNA fingerprints in the library can artificially influence strain groupings and incorrectly increases percentages of strains correctly assigned to source groups.

Results presented here also show that despite our use of an increased number of individuals in our library for DNA fingerprinting, we still failed to capture the genetic diversity

present in *E. coli*. Populations of *E. coli* have been shown to be very diverse (49) and this is evidenced by rarefaction analysis results shown in Figure 1. Despite having a known source library or over 1500 unique isolates, the number of genotypes uncovered by DNA fingerprinting continued to increase at a constant rate. Moreover, across all animal hosts, the majority of these fingerprints occurred only once. For a library to be truly representative it needs to be large enough to capture all the unknowns present in an environmental sample, otherwise strain assignment will most likely be incorrect, or a large number of isolates will be characterized as being unknowns or cosmopolitan. Since the rarefaction curve in Figure 1 has not become asymptotic, our data cannot be used to predict the ultimate size that this library needs to be. However, data presented in Figure 2 indicates that with our current library size, each new isolate added to the library only has a greater than 50% chance of being new. It has been suggested that a library size of 20,000 to 40,000 isolates may be needed to capture all the genetic diversity present in *E. coli* (Mansour Samadpour, personal communication). One suggested strategy to avoid this under-representation problem in large regional or national libraries, is to develop moderate sized libraries for a highly confined geographical region, wherein isolates are only obtained from the animals in the study area. In this way only animals pertinent to the study site, and those likely to have an impact on the targeted watershed, need to be examined in detail

We also report here the development and evaluation of HFERP as an alternative to the standard rep-PCR method. HFERP was shown to reduce gel-to-gel variability and illegitimate clustering of fingerprints within gels. HFERP utilizes a fluorescent-labeled rep-primer (6-FAM-labeled Box A1R) in the PCR reaction, and a size standard set labeled with a second fluorophore (ROX) in each gel lane. Previously, Versalovic, *et al.* (56) and Rademaker, *et al.* (40) reported on the use of fluorophore-enhanced rep-PCR (FERP), whereby polyacrylamide gel

electrophoresis and automated DNA sequencers were used to separate and detect bands generated by the FERP protocol. While the more automated method presented by these authors has some advantages, the increased cost of analyses and the limited dynamic range of fragment size separation on sequencing gels did not make this technique useful in our applications. Thus, in our HFERP studies we separated PCR products using horizontal agarose gel electrophoresis in the same manner as the standard rep-PCR protocol. This allows for the separation of a large range of DNA band sizes using more standard laboratory equipment. Moreover, the presence of a size standard in each lane of the HFERP gel allows for the very accurate normalization of bands within and between gels, which corrects for band-migration variation that occurs during electrophoresis. The result of the normalization process is that fingerprint patterns from different gels can be accurately compared. It should be noted, however, that the intensity of HFERP bands are more variable than those generated by rep-PCR, and that some of the gains achieved by more precise alignment of bands may be offset by more variation in band intensity. We found that this variation in intensity can be overcome by the careful mixing of all reagents in the PCR master mix and greater pipetting precision when loading gels (data not presented). Further improvements to increasing the intensity of HFERP-generated DNA fingerprints may also be obtained by varying the ratio of labeled to unlabeled primer and the final concentration of the primer mixture in PCR reactions. Nevertheless, our results clearly show that HFERP-derived DNA fingerprint bands are more precisely aligned than the rep-PCR bands. In addition, we show that HFERP DNA fingerprints generated by our method reduce within gel groupings of fingerprints, which can have profound ramifications for the assembly of libraries and the analysis of unknown environmental isolates.

A variety of similarity measures exist. Binary similarity coefficients are mostly used to analyze presence/absence data (28) and band- matching data obtained from DNA fingerprints can be analyzed using binary coefficients. However, quantitative similarity coefficients require a measure of relative abundance (27). Quantitative coefficients can be applied to DNA fingerprints when the fingerprints are analyzed as densitometric curves that take into account both peak position and intensity (peak height). Häne, *et al.* (17) demonstrated that for complex DNA fingerprints, such as those produced with the techniques we used here, a curve-based method such as Pearson's product-moment correlation coefficient more reliably identified similar or identical DNA fingerprints than band matching formulas, such as simple matching, Dice, or Jaccard. Results presented here confirm that the curve-based Pearson's product-moment correlation coefficient was superior to the band-based Jaccard algorithm is correctly assigning isolates to the correct source group. Similarly, Louws and co-workers (30) reported that curve-based statistical methods worked best for analysis of complex banding profiles generated by rep-PCR, since comparison of curve data is less dependent on DNA concentration in loaded samples and is relatively insensitive to background differences in gels. More recently, Albert *et al.* (1) performed a statistical evaluation of rep-PCR DNA fingerprint data and reported that *k*-nearest neighbor's classification was similar to Person's product-moment coefficient in its ability to correctly classify fingerprints of 584 *E. coli* isolates.

Further refinements to the Jackknife analysis, including the pooling of source groups into domesticated, human, and wild-life categories, were found to improve the ability to correctly classify isolate to their respective source groups. Over 83, 53, and 71% of domesticated animals, humans, and wild-life animals, respectively, were correctly classified using this approach with the unique DNA fingerprint library analyzed by HFERP. When all animal sources were pooled

into one group, the overall correct classification rate for humans and animals by HFERP was improved to about 94 and 54%, respectively, when analyzed using the curve-based Pearson's correlation coefficient. Accordingly, these results indicated that (1) broader classifications of source groups should be used when appropriate, or (2) a targeted subset of the DNA fingerprint database should be used to more precisely determine sources of fecal pollutants in watersheds where specific source groups are known to be present. The pooling of source groups into a more limited number of categories has previously been shown to increase the average rate of correct classification following discriminant analysis of antibiotic resistance (16,21,62), ribotype (6,7), and rep-PCR DNA fingerprint analyses (7).

In summary, our results suggest that HFERP-generated Box A1R DNA fingerprints of *E. coli* are useful to differentiate between different *E. coli* subtypes of human and animal origin and that this method reduces within gel groupings of DNA fingerprints, and ensures more proper alignment and normalization of fingerprint data. However, our results further indicate that other important issues must also be resolved to more fully understand the potential applications and limitations of this and other library-based microbial source tracking methodologies. Among these are questions concerning the inclusion of identical DNA fingerprints from the same animal in the library, the number of fingerprints that must be included in an *E. coli* known source library to adequately capture the diversity of *E. coli* genotypes that exist among potential host animals, and ultimately, whether *E. coli* exhibits a sufficient level of host specificity to allow unambiguous assignment of unknown environmental *E. coli* to specific host animals.

ACKNOWLEDGEMENTS

This work was supported, in part, by funding from the Legislative Commission on Minnesota Resources through the Environment and Natural Resources Trust Fund and the MN Future Resource Fund, the Metropolitan Council Environmental Services, and The University of Minnesota Agricultural Experiment Station (to M.J.S.).

We would like to thank Priscilla Dombek for DNA fingerprinting assistance, Linda Kinkel and Anita Davelos for help with rarefaction analysis, and Vivek Kapur, Todd Markowski and Bruce Witthun for help with laser image analysis.

REFERENCES

1. **Albert, J. M., J. Munkata-Marr, L. Tenorio, and R. L. Siegrist.** 2003. Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli.* Environ. Sci. technol. **37:**4554-4560.
2. **Arnold, C., L. Metherell, G. Willshaw, A. Maggs, and J. Stanley.** 1999. Predictive fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: high resolution typing method with phylogenetic significance. J. Clin. Microbiol. **37:**1274-1279.
3. **Bernhard, A. E., and K. G. Field.** 2000. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl. Environ. Microbiol. **66:**1587-1594.

4. **Boualam, M., L. Mathieu, S. Fass, J. Cavard, and D. Gatel.** 2002. Relationship between coliform culturability and organic matter in low nutritive waters. Water Res **36:** 2618-2626.

5. **Burnes, B. S.** 2003. Antibiotic resistance analysis of fecal coliforms to determine fecal pollution sources in a mixed-use watershed. Environ. Monit. Assess. **85:**87-98.

6. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and A. Asfaw.** 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. Appl. Environ. Microbiol. **67:**1503-1507.

7. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and J. D. Schnell.** 2003. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. Appl Environ Microbiol **69:**1836-1839.

8. **de Bruijn, F. J.** 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. Appl. Environ. Microbiol. **58:**2180-2187.

9. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66:**2572-2577.

10. **Geldreich, E. E., and B. A. Kenner.** 1969. Concepts of fecal streptococci in stream pollution. J. Wat. Poll. Cont. Fed. Suppl. **41:**R336-R332.

11. **Gordon, D. M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiol. **147:**1079-1085.

12. **Gordon, D. M., S. Bauer, and J. R Johnson.** 2002. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. Microbiol. **148:**1513-1522.

13. **Grabow, W. O. K., T. E. Neubrech, C. S. Holtzhausen, and J. Jofre.** 1995. *Bacteroides fragilis* and *Escherichia coli* bacteriophages: excretion by humans and animals. Wat. Sci. Tech. **31:**223-230.

14. **Graves, A. K., A. T. Hagedorn, M. Mahal, A. M. Booth, and R. B. Reneau Jr.** 2002. Antibiotic resistance profiles to determine sources of fecal contamination in a rural Virginia watershed. J. Environ. Qual. **31:**1300-1308.

15. **Guan, S., R. Xu, S. Chen, J. Odumeru, and C. Gyles.** 2002. Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources. Appl. Environ. Microbiol. **68:**2690-2698.

16. **Hagedorn, C., S. L. Robinson, J. R. Filtz, S. M. Grubbs, T. A. Angier, and R. B. Reneau, 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.

17. **Häne, B.G., K. Jäger, and H. Drexler.** 1993. The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprinting profiles than band matching algorithms. Electrophor. **14:**967-972.

18. **Hartel, P. G., J. D. Summer, J. L. Hill, J. Collins, J. A. Entry, W. and I. Segars.** 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. of Environ. Qual. **31:**1273-1278.

19. **Hartel, P. G., J. D. Summer, and W. I. Segars.** 2003. Deer diet affects ribotypo diversity of *Escherichia coli* for bacterial source tracking. Water Res. **37:**3263-3268.

20. **Hartl, D. L., and D. E. Dykhuizen.** 1984. The population genetics of *Escherichia coli.* Annu. Rev. Genet. **18:**31-68.

21. **Harwood, V. J., J. Whitlock, and V. H. Withington.** 2000. Classification of the antibiotic resistance patterns of indicator bacteria by discriminant analysis: Use in predicting the source of fecal contamination in subtropical Florida waters. Appl. Environ. Microbiol. **66:**3698-3704.

22. **Havelaar, A. H., W. M. Pot-Hogeboom, K. Furuse, R. Pot, and M. P. Horman.** 1990. F-specific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin. J. Appl. Bacteriol. **69:**30-37.

23. **Hsu, F-C., Y.-S. Carol Shieh, J. van Duin, M. J. Beekwilder, and M. D. Sobsey.** 1995. Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. Appl. and Environ. Microbiol. **61:**3960-3966.

24. **Hulton, C. S. J., C. F. Higgins, and P. M. Sharp.** 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli, Salmonella typhimurium* and other enterobacteria. Mol. Microbiol. **5:**825-834.

25. **Jenkins, M. B., P. G. Hartel, T. J. Olexa, and J. A. Stuedemann.** 2003. Putative temporal variability of *Escherichia coli* ribotypes from yearling steers. J Environ Qual. **32:**305-309.

26. **Judd, A. K., M. Schneider, M. J. Sadowsky, and F. J. de Bruijn.** 1993. Use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. Appl. Environ. Microbiol. **59:**1702-1708.

27. **Krebs, Charles J**. 1999. Ecological methodology. Benjamin/Cummings, Menlo Park, CA.

28. **Legendre, P., and L. Legendre.** 1998. Numerical Ecology, 2nd English edition. Elsevier Science, Amsterdam, The Netherlands.

29. **Lipman, J. A., A. de Nijs, T. J. G. M. Lam, and W. Gaastra.** 1995. Identification of *Escherichia coli* strain from cows with clinical mastitis by serotyping and DNA polymorphism patterns with REP and ERIC primers. Vet. Microbiol. **43:**13-19.

30. **Louws, F. J, J. L. W. Rademaker, and F. J. de Bruijn.** 1999. The three Ds of PCR-based genomic analysis of phytobacteria: diversity, detection, and disease diagnosis. Ann. Rev. Phytopath. **37:**81-125.

31. **Mara, D. D., and J. I. Oragui.** 1983. Sorbitol-fermenting bifidobacteria as specific indicators of human fecal pollution. J. Appl. Bacteriol. **55:**349-357.

32. **Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D. A. Morrison, G. J. Boulnois, and J.-P. Claverys.** 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucl. Acids Res. **20:**3479-3483.

33. **McLellan, S. L., A. D. Daniels and A. K. Salmore.** 2003. Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution using DNA fingerprinting. Appli. Environ. Microbiol. **69:**2587-2594.

34. **Nebra, Y., X. Bonjoch, A. R. Blanch.** 2003. Use of *Bifidobacterium dentium* as an indicator of the origin of fecal water pollution. Appl. Environ. Microbiol. **69:**2651-2003.

35. **Osawa, S., K. Furuse, and I. Wantanabe.** 1981. Distribution of ribonucleic acid coliphages in animals. Appl. Environ. Microbiol. **41:**164-168.

36. **Parveen, S., N. C. Hodge, R. E. Stall, S. R. Farrah, and M. L. Tamplin.** 2001. Phenotypic and genotypic characterization of human and nonhuman *Escherichia coli.* Wat. Res. **35:**379-386.

37. **Parveen, S., R. L. Murphree, L. Edmiston, C. W. Kasper, K. M. Portier, and M. L. Tamplin.** 1997. Association of multiple-antibiotic-resistance profiles with point and nonpoint sources of *Escherichia coli* in Apalachicola bay. Appl. Environ. Microbiol. **63:**2607-2612.

38. **Parveen, S., K. M. Portier, K. Robinson, L. Edmiston, and M. L. Tamplin.** 1999. Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Appl. Environ. Microbiol. **65:**3142-3147.

39. **Puig, A., J. Jofre, and R. Araujo.** 1997. Bacteriophages infecting various *Bacteroides fragilis* strains differ in their capacity to distinguish human from animal faecal pollution. Wat. Sci Tech. **35:**359-362.

40. **Rademaker, J. L. W., F. J. Louws, and F. J. deBruijn.** 1998. Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting, suppl. 3, chapter 3.4.3, p.1-26. *In* A. D. L. Akkermans, J. D. van Elsas and F. J. de Bruijn (ed.), Molecular microbial ecology manual. Luwer Academic Publishers, Dordrecht, The Netherlands.

41. **Rademaker, J. L. W., F. J. Louws, U. Rossbach, P. Vinuesa, and F. J. deBruijn.** 1999. Computer-assisted pattern analysis of molecular fingerprints and database construction, suppl. 4, chapter 7.1.3, p.1-33. *In* A. D. L. Akkermans, J. D. van Elsas and F. J. de Bruijn (ed.), Molecular microbial ecology manual. Kluwer Academic Publishers, Dordrecht, The Netherlands.

42. **Sadowsky, M. J., and H.-G. Hur**. 1998. Use of endogenous repeated sequences to fingerprint bacterial genomic DNA, 399-413. *In* J.R. Lupski, G. Weinstock, and F. J. de Bruijn (ed.), Bacterial genomes: structure and analysis. Chapman & Hall, New York, NY.

43. **Sadowsky, M. J., L. L. Kinkel, J. H. Bowers, and J. L. Schottel.** 1996. Use of repetitive intergenic DNA sequences to classify pathogenic and disease-suppressive *Streptomyces* strains. Appl. Environ. Microbiol. **62:**3489-3493.

44. **Scott, T. M., S. Parveen, K. M. Portier, J. B. Rose, M. L. Tamplin, S. R. Farrah, A. Koo, and J. Lukasik.** 2003. Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. **69:**1089-1092.

45. **Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik.** 2002. Microbial source tracking: current methodology and future directions. Appl. Environ. Microbiol. **68:**5796-5803.

46. **Seurinck, S., W. Verstraete, and S. D. Siciliano.** 2003. Use of 16S-23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify nonpoint fecal sources. Appl. Environ. Microbiol. **69:**4942-4950.

47. **Simmons, G. M., Jr., S. A. Herbein, and C. M. James.** 1995. Managing nonpoint fecal coliform sources to tidal inlets. Water Res. Update. **100:**64-74.

48. **Simpson, J. M., J. W. Santo Domingo and D. J. Reasoner.** 2003. Microbial source tracking: state of the science. Environ. Sci. Tech. **36:**5280-5288.

49. **Souza, V., M. Rocha, A. Valera, and L. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. **65:**3373-3385.

50. **Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37:**1015-1026.

51. **Tartera, C. and J. Jofre.** 1987. Bacteriophages active against *Bacteroides fragilis* in sewage-polluted waters. Appl. Environ. Microbiol. **53:**1632-1637.

52. **Tyler, K. D., G. Wang, S. D. Tyler, and W. M. Johnson.** 1997. Factors affecting reliability and reproducibility of amplification-based DNA fingerprinting of representative bacterial pathogens. J. Clin. Microbiol. **35:**339-346.

53. **U.S. Environmental Protection Agency.** 2000. National Water Quality Inventory: 1998 Report to Congress. EPA-841-R-00-001. Office of Water, Washington, D.C.

54. **U.S. Environmental Protection Agency.** 2001. Protocol for Developing Pathogen TMDLs. EPA 841-R-00-002. Office of Water, Washington, DC.

55. **Versalovic, J., F. J. de Bruijn, and J. R. Lupski.** 1998. Repetitive sequence-based PCR (rep-PCR) DNA fingerprinting of bacterial genomes, p. 437-454. *In* F. J. de Bruijn, J. R. Lupski, and G. M. Weinstock (eds.), Bacterial genomes: physical structure and analysis. Chapman & Hall, New York, NY.

56. **Versalovic, J., V. Kapur, T. Koeuth, G. H. Mazurek, T. S. Whittam, J. M. Musser, and J. R. Lupski.** 1995. DNA fingerprinting of pathogenic bacteria by fluorophore-enhanced repetitive sequence-based polymerase chain reaction. Arch. Pathol. Lab. Med. **119:**23-29.

57. **Versalovic, J., T. Koeuth, and J. R. Lupski.** 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucl. Acids Res. **19:**6823-6831.

58. **Versalovic, J., M. Schneider, F. J. de Bruijn, and J. R. Lupski.** 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol. Cell. Biol. **5:**25-40.

59. **Wheeler, A. L., P. G. Hartel, D. G. Godfrey, J. L. Hill, and W. I. Segars.** 2002. Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking. J. Environ. Qual. **31:**1286-1293.

60. **Whitlock, J. E., D. T. Jones, and V. J. Harwood.** 2002. Identification of the sources of fecal coliforms in and urban watershed using antibiotic resistance analysis. Water Res. **36:**4273-4282.

61. **Wiggins, B. A.** 1996. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. Appl. Environ. Microbiol. **62:**3997-4002.

62. **Wiggins, B. A., R. W. Andrews, R. A. Conway, C. L. Corr, E. J. Dobratz, D. P. Dougherty, J. R. Eppard, S. R. Knupp, M. C. Limjoco, J. M. Mettenburg, J. M. Rinehardt, J. Sonsino, R. L. Torrijos, and M. E. Zimmerman.** 1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Appl. Environ. Microbiol. **65:**3483-3486.

63. **Wiggins, B.A., P. W. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M. Smith, E. N. Stevens, and A. K. Varner.** 2003. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. Appl. Environ. Microbiol. **69:**3399-3405.

Table 1. Animal source groups and rep-PCR DNA fingerprints generated from *E. coli* isolates.

Animal Source Group	Individuals Sampled	Total Fingerprints	Unique Fingerprints[a]
Cat	37	108	48
Chicken	86	231	144
Cow	115	299	191
Deer	64	179	96
Dog	71	196	106
Duck	42	122	81
Goat	36	104	42
Goose	73	200	135
Horse	44	114	79
Human	197	307	211
Pig	111	303	215
Sheep	37	101	61
Turkey	69	202	126
Total	**982**	**2,466**	**1,535**

[a]Identical *E. coli* genotypes from each individual animal were removed.

Table 2. Percentage of known-source rep-PCR DNA fingerprints assigned to the correct source group by Jackknife analysis[a].

Animal Source	All Fingerprints (n=2,466)	Unique Fingerprints (n=1,535)
	Percent Correctly Classified Isolates	
Pets[b]	91.8 (279)[d]	61.7 (95)
Chicken	81.4 (188)	59.7 (86)
Cow	79.6 (238)	55.0 (105)
Deer	85.5 (145)	55.2 (53)
Waterfowl[c]	81.4 (262)	66.2 (143)
Goat	97.1 (101)	66.7 (28)
Horse	69.3 (79)	44.3 (35)
Human	78.3 (240)	59.2 (125)
Pig	77.9 (236)	63.7 (137)
Sheep	79.0 (80)	47.5 (29)
Turkey	88.6 (179)	73.8 (93)
Overall	82.2 (2,027)	60.5 (929)

[a]Done using Pearson's product moment correlation coefficient with 1% optimization and maximum similarities.
[b]Pet group consists of cats and dogs
[c]Waterfowl group consists of ducks and geese.
[d]Values in parentheses are number of isolates correctly classified.

Table 3. Unique *E. coli* isolates correctly classified into source groups by rep-PCR and HFERP DNA fingerprinting methods.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]			
		rep-PCR		HFERP	
		Pearson	Jaccard	Pearson	Jaccard
Pets[b]	154	61.7 (95)[d]	45.5 (70)	59.1 (91)	44.8 (69)
Chicken	144	59.7 (86)	38.9 (56)	63.2 (91)	31.9 (46)
Cow	189	55.0 (104)	47.6 (90)	62.0 (117)	48.2 (91)
Deer	96	55.2 (53)	36.5 (35)	62.2 (60)	42.6 (41)
Waterfowl[c]	216	66.2 (150)	52.8 (114)	70.4 (152)	56.5 (122)
Goat	42	66.7 (27)	59.5 (25)	47.6 (20)	42.9 (18)
Horse	78	44.3 (35)	34.2(27)	52.6 (41)	32.1 (25)
Human	210	59.2 (124)	47.4(100)	53.8 (113)	45.2 (95)
Pig	215	63.7 (137)	43.7 (94)	54.4 (117)	36.3 (78)
Sheep	61	7.5 (29)	39.3 (24)	37.7 (23)	8.2 (5)
Turkey	126	73.8 (93)	52.4 (66)	73.0 (92)	54.8 (69)
Overall	1,531	60.9 (933)	45.8 (701)	59.9 (917)	43.0 (659)

[a]Based on Jackknife analysis with 1% optimization and maximum similarities using curve-based (Pearson's product moment correlation coefficient) or band-based (Jaccard's coefficient) similarity calculations.
[b]Pet group consists of cats and dogs.
[c]Waterfowl group consists of ducks and geese.
[d]Values in parentheses are number of isolates correctly classified.

Table 4. Percentage of *E. coli* isolates correctly classified into domestic, human and wildlife source groups by using the HFERP DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]	
		Pearson	Jaccard
Domesticated[b]	855	83.2 (711)[e]	77.5 (663)
Human	210	53.8 (113)	45.2 (95)
Wildlife[c]	312	71.4 (223)	59.6 (186)
Pets[d]	154	59.1 (91)	44.8 (69)
Overall	1,531	74.3 (1,138)	66.2 (1,013)

[a]Done using Jackknife analysis with 1% optimization and maximum similarities using curve-based Pearson's product-moment correlation coefficient and band-based Jaccard similarity calculations.
[b]Domesticated group includes, chickens, cows, goats, horses, pigs, sheep and turkeys.
[c]Wildlife group includes deer, ducks and geese.
[d]Pet group includes dos and cats.
[e]Values in parentheses are number of isolates correctly classified.

Table 5. Percentage of *E. coli* isolates correctly classified into human and animal source groups by using the HFERP DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]	
		Pearson	Jaccard
Animal	1321	93.7 (1,237)[b]	92.6 (1,223)
Human	210	53.8 (113)	45.2 (95)
Overall	1,531	88.2 (1,350)	86.1 (1,318)

[a]Done using Jackknife analysis with 1% optimization and maximum similarities using curve-based Pearson's product moment correlation coefficient and band-based Jaccard's similarity calculations.
[b]Values in parentheses are number of isolates correctly classified.

Figure Legends

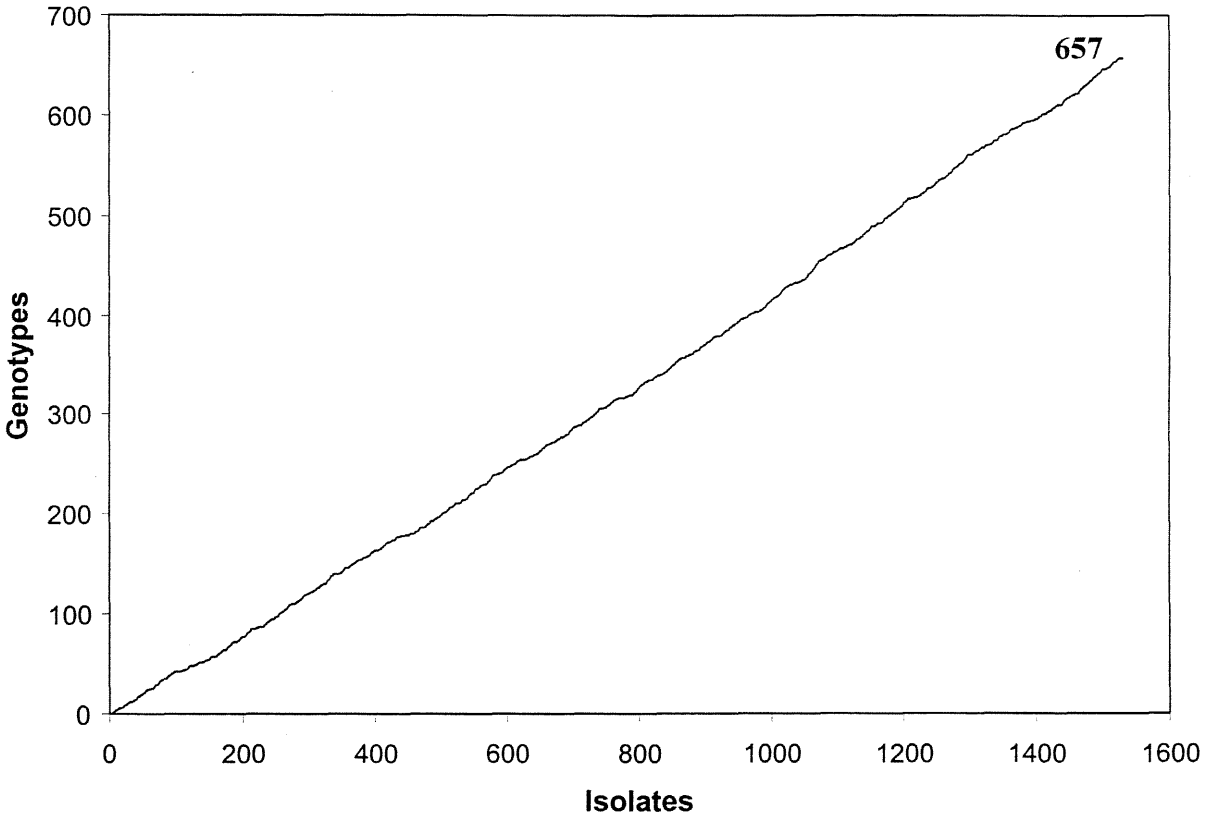
Figure 1. Accumulation curve of genotypes from *E. coli* isolates. Of 1,535 unique *E. coli* isolates in the known-source database with rep-PCR DNA fingerprint similarities of 90% or greater (based on the cosine coefficient), 657 genotypes were identified. The isolates were randomized, and a rarefaction curve was constructed by summing the number of genotypes that accumulated with the successive addition of isolates.

Figure 2. Frequency of occurrence of genotypes among rep-PCR DNA fingerprints from unique *E. coli* isolates. Analysis was limited to the 657 genotypes identified among the 1,535 unique *E. coli* isolates with rep-PCR DNA fingerprint similarities of 90% or greater.

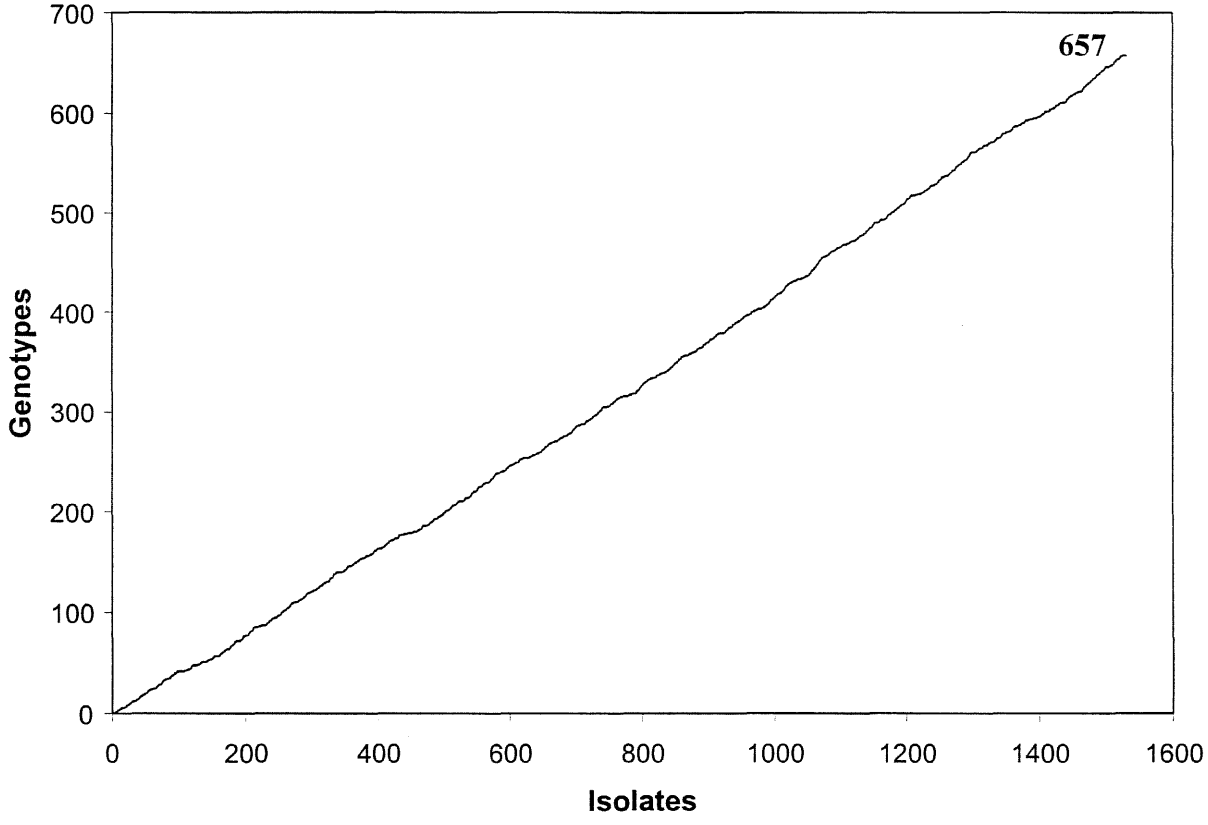
Figure 3. Representative examples of HFERP DNA fingerprint images. Genomic DNAs from 24 *E. coli* strains were subjected to HFERP DNA fingerprint analysis using a mixture of unlabeled Box A1R and 6-FAM fluorescently labeled Box A1R primers. Each lane contained Genescan-2500 ROX internal lane standards and HFERP DNA fingerprints. The combined, dual colored, HFERP image (A) was captured using a Typhoon Imager and two emission filters. Values in margin are in base pairs. Individual images of the HFERP DNA fingerprints (B) and Genescan-2500 ROX internal lane standard (C) were acquired using one filter at a time.

Figure 4. Comparison of DNA fingerprint patterns of a reference *E. coli* strain generated using rep-PCR and HFERP. (A) rep-PCR DNA fingerprint patterns were assembled from 29 individual

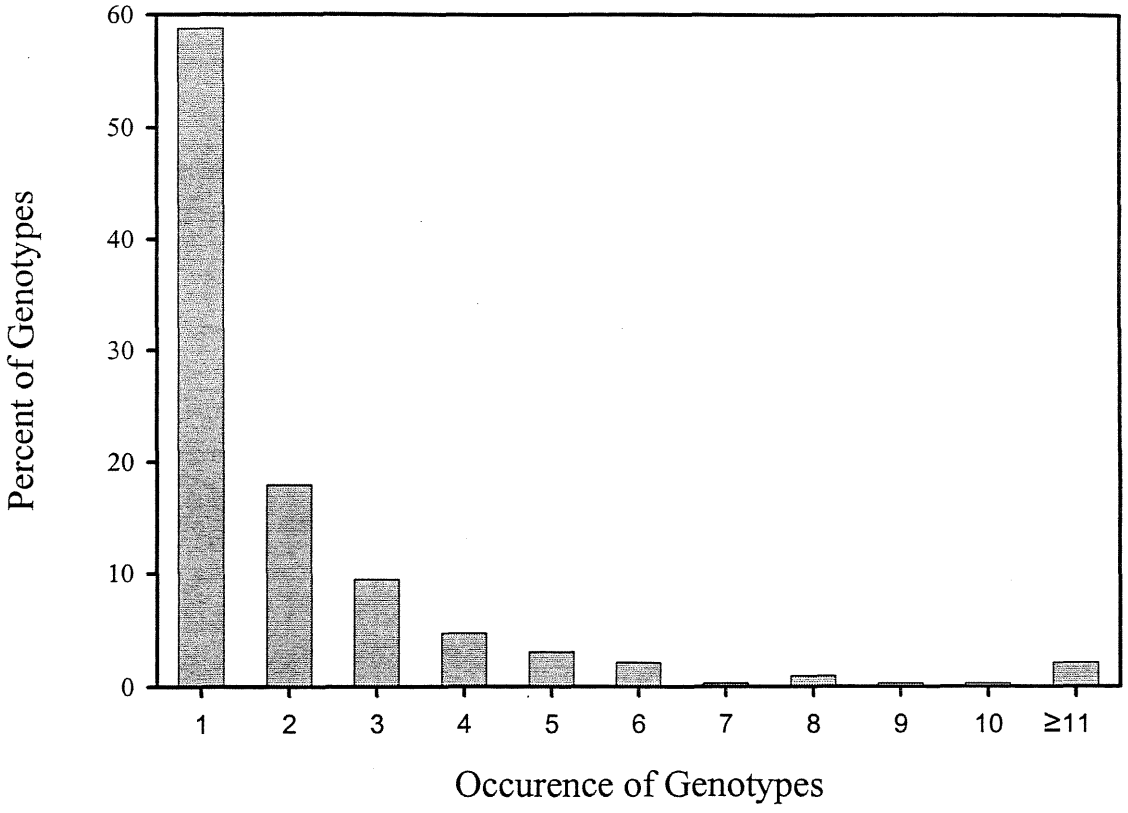
PCR reactions, each of which was run on a separate agarose gel. Fingerprints were generated using *E. coli* isolate P294 as template DNA and the Box A1R primer. (B) HFERP DNA fingerprint patterns were assembled from 29 individual PCR reactions each, of which was run on a separate agarose gel. Fingerprints were generated using *E. coli* isolate P294 as template DNA and a mixture of unlabeled Box A1R and 6-FAM fluorescently labeled Box A1R primers. Bands were aligned using Genescan-2500 ROX internal standards, which were present in each lane. Similarities were determined using the cosine algorithm of Bionumerics and dendrograms were generated using the unweighted pair-group method using arithmetic means (UPGMA).



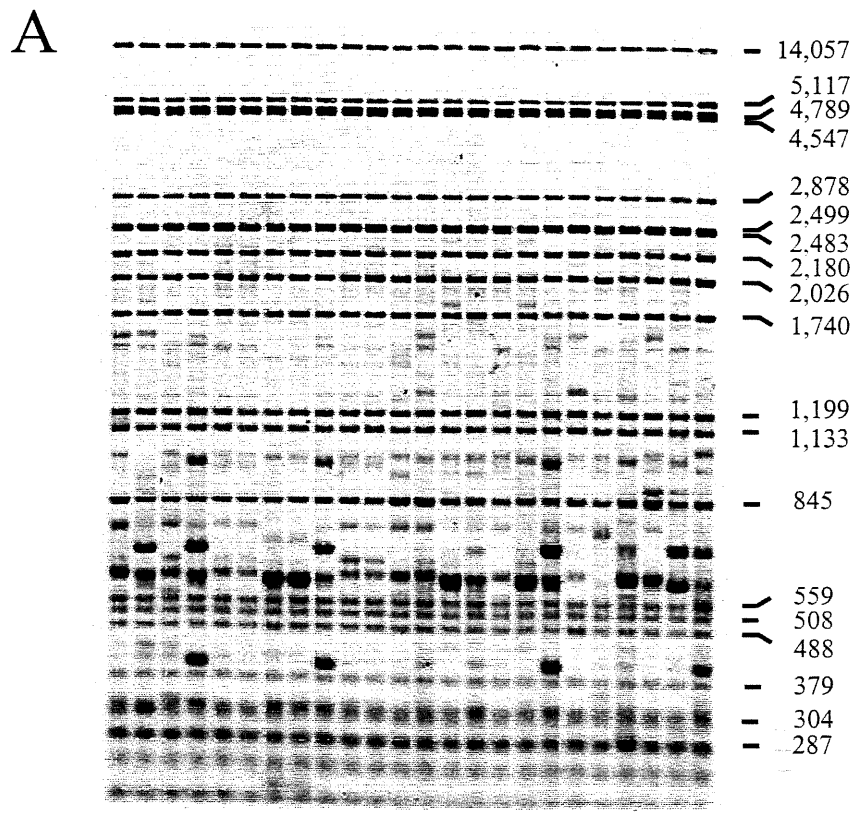
Johnson *et al.* – Figure 1



Johnson *et al.* – Figure 1

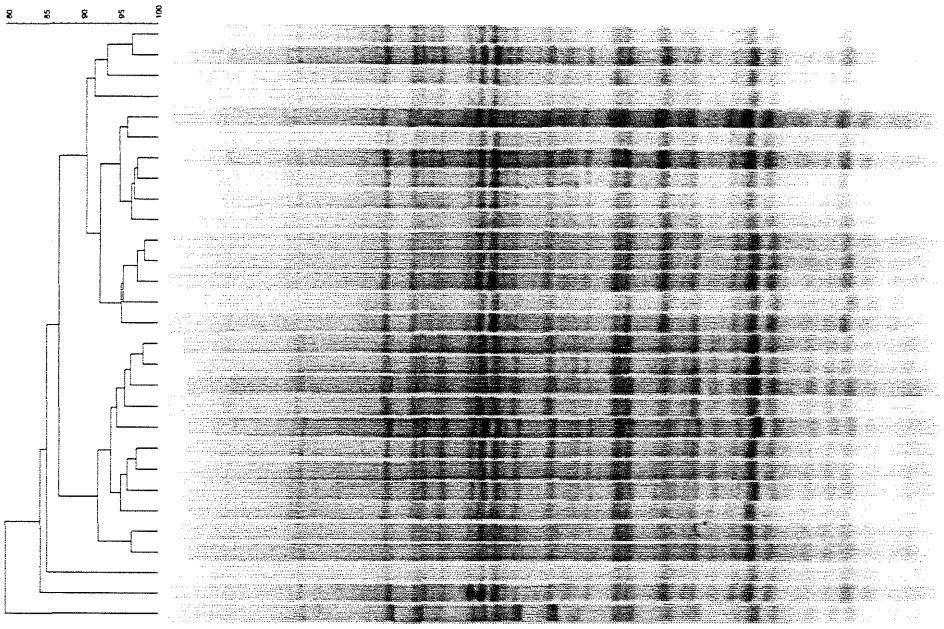


Johnson *et al.* – Figure 2

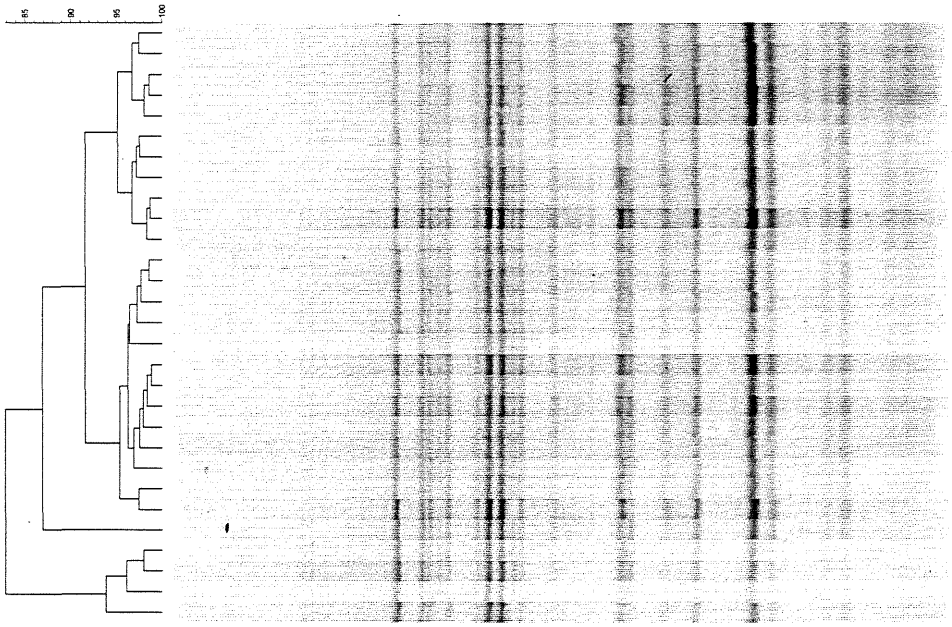


Johnson *et al.* – Figure 3

A



B



Johnson *et al.* – Figure 4