Vol. 69, No. 3

Comparison of Ribotyping and Repetitive Extragenic Palindromic-PCR for Identification of Fecal *Escherichia coli* from Humans and Animals

C. Andrew Carson,[1]* Brian L. Shear,[1] Mark R. Ellersieck,[2] and Jennifer D. Schnell[1]

*Department of Veterinary Pathobiology[1] and Agricultural Experiment Station Statisticians,[2]
University of Missouri, Columbia, Missouri 65211*

This report compares the performances of two popular genotypic methods used for tracking the sources of fecal pollution in water, ribotyping and repetitive extragenic palindromic-PCR (rep-PCR). The rep-PCR was more accurate, reproducible, and efficient in associating DNA fingerprints of fecal *Escherichia coli* with human and animal hosts of origin.

Water is routinely monitored for compliance with government standards in the interest of public health (7, 17). Pollution from human and animal waste is traditionally indicated by the presence of commensal *Escherichia coli* (1). Though these organisms are essentially nonpathogenic, their presence warns of the possible concurrent existence of pathogenic microbes. Regulatory plans for remediation of impaired waterways, including establishment of daily allowance limits for pollutants (20), will now require accurate identification of host sources of fecal pollution (bacterial source tracking). Traditional methods, such as phage susceptibility (22) and the ratio of fecal coliforms to streptococci (6), have been routinely used as indicators of human or nonhuman pollution. Recently, DNA fingerprinting methods (4, 9, 15) and antibiotic resistance profiles (11, 21) have been reported as more accurate means to characterize fecal *E. coli* isolates with respect to host source. The latter approaches are based on the concept that human and nonhuman hosts harbor particular populations of *E. coli* which can be associated with host of origin. Examples of DNA-based procedures considered promising for bacterial source tracking include pulsed-field gel electrophoresis (14), ribotyping (3, 15, 16), ribosomal DNA heterogeneity (2), and repetitive extragenic palindromic-PCR (rep-PCR) (4). Numerous state water quality laboratories in the United States (including those of Delaware, Florida, Minnesota, Wisconsin, Washington, and Missouri) are currently applying either ribotyping or rep-PCR to bacterial source tracking for studies by environmental monitoring or regulatory agencies. However, we are unaware of reports of any controlled comparison of these two methods in regard to respective accuracy and efficiency of bacterial fingerprinting. Previous reports describing the performance of these two methods (3, 4) were based on evaluation of different collections of fecal *E. coli*, and different statistical methods of analysis were used. The present study was performed with a single collection of isolates and one program for pattern analysis. Comparison of the two procedures addressed accuracy of DNA pattern discrimination, reproducibility, ease of performance, resources required, and cost.

Fecal *E. coli* isolates. A collection of 482 fecal *E. coli* isolates from humans, cattle, swine, horses, dogs, chickens, turkeys, and migratory geese was used for this study. The collection contained 136 human isolates and 346 isolates from nonhuman hosts. Human samples were collected as anal swabs from volunteers and processed separately. Cattle, pig, chicken, turkey, and goose samples were collected from numerous production farms. Samples from the same species on a single farm were combined and mixed well prior to processing in the laboratory. Multiple horse and dog samples from each stable and boarding kennel, respectively, were similarly combined according to species. Fresh migratory goose feces were collected and combined at several locations while the birds were present. All fecal samples were cultured overnight in lactose broth (Becton Dickinson, Sparks, Md.) at 37°C. Fecal *E. coli* isolates were then selected by growth on mFc, mENDO (Les), MacConkey methylumbelliferyl-β-glucuronide, and Luria broth agar at prescribed temperatures (all products from Becton Dickinson). Final confirmation of isolates as fecal *E. coli* was accomplished with a BBL Crystal Identification Systems Enteric/Nonfermenter system (Becton Dickinson) with indole and oxidase tests. Table 1 indicates the numbers of fecal *E. coli* isolates from each host species, the number of individuals represented in each host class, and the geographic origin of samples. In view of the recent report confirming the existence of regional variation in strains of fecal *E. coli* in host species (10), it is important that all isolates were collected in Missouri.

Ribotyping. Ribotyping was performed according to a previously reported procedure (3). Briefly, the fecal *E. coli* isolates were cultured, DNA was extracted and digested with restriction enzyme, fragments were separated by electrophoresis, and a labeled rRNA probe was used to generate ribotype patterns.

rep-PCR. The rep-PCR was essentially performed by a slight modification of a previously reported procedure (4). Fecal *E. coli* bacteria were isolated from specimens as described under "Fecal *E. coli* isolates." Whole-cell suspensions of *E. coli* cultures were lysed with Lyse-n-Go PCR reagent (Pierce Chemical Co., Rockford, Ill.). PCR products were produced with the BOX A1R primer (4). Electrophoresis was performed in a 1.5% SeaKem agarose gel (BioWhittaker, Rockland, Maine) at 100 V for 4 h at room temperature.

---

* Corresponding author. Mailing address: 201 Connaway Hall, University of Missouri, Columbia, MO 65211. Phone: (573) 884-7640. Fax: (573) 884-0521. E-mail: carsonc@missouri.edu.

TABLE 1. Fecal *E. coli* isolates used in study

Host source	No. of isolates	No. of individuals represented	Location[a]
Human	136	28	C Missouri
Cattle	62	35[b]	N, C, S Missouri
Pig	47	42[b]	C Missouri
Horse	51	16	C, S Missouri
Dog	41	23	C Missouri
Chicken	52	39[b]	C, S Missouri
Turkey	51	14[b]	C, S Missouri
Goose	42	20[b]	C Missouri
Total	482	215	

[a] N, north; C, central; S, southern.
[b] Approximate.

Statistical analysis of DNA patterns. Gel images of DNA fingerprints were captured with a Kodak EDAS 290 system (Kodak Co., Rochester, N.Y.). Fingerprint patterns were analyzed with Bionumerics software, version 3.0 (Applied Maths, Kortrijk, Belgium), with ribotype bands between 500 bp and 22.0 kb and rep-PCR bands between 300 bp and 10.0 kb.

Maximum similarity coefficients were derived by the curve-based Pearson correlation method (13). Discriminant analysis of fecal *E. coli* patterns in the database of known-host samples was accomplished by cross-validation (Jackknife method). Isolates initially entered into the database in association with a known host of origin were individually removed and re-presented as test subjects (12) for association with one of the eight host classes. This exercise, repeated for each of the 482 fecal *E. coli* isolates, determined the accuracy with which DNA patterns of isolates were assigned to each source class. The number of isolates correctly assigned to the proper host class by discriminant analysis is the rate of correct classification (RCC). The RCC was established for isolates in each of the eight host classes, and an average RCC (ARCC) was calculated for ribotyping and rep-PCR. A two-class comparison of performance was also done for human and nonhuman (pooled) patterns. As a further test of class discrimination, the holdout method of cross-validation (12) was performed by randomly selecting 25% of the isolates in each host class for removal from the database. The removed isolates were then presented as "unknowns" for assignment to host classes. This method is considered to be a more rigorous test of the predictive power of the databases (11), and in this instance, 120 of the 482 fecal *E. coli* isolates were held out for cross-validation.

TABLE 3. Comparison of RCCs of human and nonhuman classes by ribotyping and rep-PCR

Host class	No. of isolates	RCC (%) by:		Statistical difference[a] (P)
		Ribotyping	rep-PCR	
Human	136	87.50	97.06	Yes (0.006)
Nonhuman	346	86.42	96.24	Yes (0.001)
Total	482			
ARCC		86.96	96.65	Yes (0.001)

[a] Row-by-column chi-square test.

Comparison of discriminant analysis by ribotyping and rep-PCR was done by use of the row-by-column chi-square test.

Ribotyping pattern assignment. Patterns of the fecal *E. coli* isolates were composed of between 6 and 12 bands. Approximately 85% of the patterns were highly resolved initially and suitable for analysis. Isolates representing the 15% which were not clearly legible were reprocessed to achieve the desired resolution. Assignment of ribotyping patterns to host class by Jackknife analysis is shown in Table 2. Boldface values, on a diagonal across the table, indicate percentages of isolates correctly assigned to host classes. RCCs ranged between 50.98% for turkey and 95.24% for goose. The ARCC was 72.78%. Table 3 shows the RCCs for human and nonhuman (pooled) ribotyping patterns as 87.50% for human and 86.42% for nonhuman patterns. The ARCC was 86.96%.

rep-PCR pattern assignment. Fingerprints generated by rep-PCR were composed of between 18 and 30 bands. Over 95% of the initial patterns were of high quality and did not require reprocessing. Table 4 shows assignment of rep-PCR patterns to host classes. RCC ranged between 66.67% for horse and 97.87% for pig. The ARCC was 88.14%. Table 3 shows the RCCs for human and nonhuman (pooled) rep-PCR patterns as 97.06% for human and 96.24% for nonhuman patterns. The ARCC was 96.65%

Comparison of ribotyping and rep-PCR. Table 5 shows the statistical analysis of the performance of ribotyping and rep-PCR with all eight host classes considered. RCC and ARCC values were compared by a row-by-column chi-square test. rep-PCR was significantly superior to ribotyping with respect to RCCs of human, pig, dog, and turkey patterns. Although not statistically significant, there was an indication of superior performance in favor of rep-PCR for cattle and chicken patterns. With respect to goose patterns, ribotyping showed an advantage, although not statistically significant. rep-PCR was signif-

TABLE 2. Assignment of ribotyping patterns to host classes by Jackknife analysis

Host class	*E. coli* patterns (%) assigned to class[a]							
	Human	Cattle	Pig	Horse	Dog	Chicken	Turkey	Goose
Human	**87.50**	2.94	2.94	2.21	2.21	0.74	1.47	0.00
Cattle	12.90	**70.97**	0.00	4.84	1.61	3.23	1.61	4.84
Pig	12.77	2.13	**68.09**	4.26	0.00	10.64	2.13	0.00
Horse	19.61	0.00	5.88	**60.78**	5.88	5.88	1.96	0.00
Dog	9.76	0.00	0.00	12.20	**75.61**	0.00	0.00	2.44
Chicken	9.62	1.92	7.69	3.85	0.00	**73.08**	1.92	1.92
Turkey	27.45	3.92	1.96	7.84	1.96	5.88	**50.98**	0.00
Goose	0.00	2.38	0.00	2.38	0.00	0.00	0.00	**95.24**

[a] RCCs of patterns to host class are in boldface. The ARCC was 72.78%.

TABLE 4. Assignment of rep-PCR patterns to host classes by cross-validation

Host class	*E. coli* patterns (%) assigned to class[a]							
	Human	Cattle	Pig	Horse	Dog	Chicken	Turkey	Goose
Human	**97.06**	0.004	0.00	0.74	0.00	0.00	1.47	0.74
Cattle	1.61	**83.87**	1.61	1.61	0.00	1.61	1.61	8.06
Pig	0.00	0.00	**97.87**	0.00	2.13	0.00	0.00	0.00
Horse	9.80	3.92	3.92	**66.67**	1.96	1.96	5.88	7.84
Dog	2.44	0.00	0.00	0.00	**97.56**	0.00	0.00	0.00
Chicken	3.85	0.00	5.77	1.92	0.00	**84.62**	3.85	0.00
Turkey	1.96	0.00	1.96	0.00	1.96	0.00	**94.12**	0.00
Goose	7.14	2.38	0.00	7.14	0.00	0.00	0.00	**83.33**

[a] RCCs of patterns to host class are in boldface. The ARCC was 88.14%.

TABLE 5. Comparison of RCCs of eight host classes
by ribotyping and rep-PCR

Host class	RCC (%) by:		Statistical difference[a] (P)
	Ribotyping	rep-PCR	
Human	87.50	97.06	Yes (0.006)
Cattle	70.97	83.87	No (0.192)
Pig	68.09	97.87	Yes (0.001)
Horse	60.78	66.67	No (0.680)
Dog	75.61	97.56	Yes (0.010)
Chicken	73.08	84.62	No (0.230)
Turkey	50.98	94.12	Yes (0.004)
Goose	95.24	83.33	No (0.158)
ARCC	72.78	88.14	Yes (0.001)

[a] Row-by-column chi-square test.

icantly superior to ribotyping in ARCC. Table 3 indicates the comparative performances of the two methods in classification of patterns of human and nonhuman (pooled) isolates. In assignment of both human and nonhuman patterns the rep-PCR was significantly superior. Similarly the rep-PCR showed the better overall performance in regard to ARCC. Results of the holdout method, whereby 25% of isolates in each host class were removed from the database and used as test isolates, were as follows. RCCs for ribotyping were as follows: human, 82.35%; cattle, 81.25%; pig, 66.67%; horse, 69.23%; dog, 70.00%; chicken, 61.53%; turkey, 64.28%; and goose, 80.00%. The ARCC was 71.91%. RCCs for rep-PCR were as follows: human, 100.00%; cattle, 93.75%; pig, 100.00%; horse, 76.92%; dog, 100.00%; chicken, 92.30%; turkey, 100.00%, and goose, 90.00%. The ARCC was 94.12%.

Previous studies of ribotyping (3, 15) and rep-PCR (4) have been reported with respect to the capacity of these methods to identify host sources of known-host isolates of fecal *E. coli*. Direct comparison of performances of the two methods, however, was difficult due to variation in test performance and means used for statistical analysis. In the present study, the two methods were compared by using a single collection of fecal *E. coli* isolates, one program for pattern analysis, and a constant means for validation of discriminant analysis and critical evaluation of results. *E. coli* isolates from human and seven nonhuman sources were included. Discriminant analysis of assignment of the isolates of each of the eight classes was one major criterion considered for each test. Accuracy of assignment of isolates to one of two classes, human and nonhuman (pooled), was another measured criterion. In the present study the rep-PCR performed better in most RCC and ARCC functions (Tables 3 and 5), and in most instances the differences were statistically significant. We speculate that the reason that rep-PCR excelled over ribotyping in accuracy of pattern classification may relate to the larger number of features which the former method records. These variables, translated into numbers of bands, typically range between 18 and 30 for rep-PCR while ribotyping patterns contain between 6 and 12 bands. Greater availability of information or richness of features achieves better pattern discrimination (5).

RCCs in the ribotyping portion of the present study varied in comparison to a previously reported study (3), and there was some improvement in ARCC. Reasons for the improvement are not certain since the collections of *E. coli* samples and

analytical programs differed from those previously used. Discriminant analysis of rep-PCR patterns in the present study, however, was generally quite comparable to RCC and ARCC percentages previously reported (4).

Though Jackknife analysis of ribotyping and rep-PCR methods yielded data indicative of power to discriminate between fecal *E. coli* isolates from various host sources, the holdout method was used to confirm the accuracy of cross-validation. RCCs for the holdout procedure, with ribotyping patterns, ranged from approximately 5 to 10 percentage points lower or higher than those generated by the Jackknife procedure. The ARCCs for holdout and Jackknife, however, were nearly the same. With respect to the rep-PCR, the holdout method yielded RCCs which ranged from about 2 to 10 percentage points lower or higher than those for the Jackknife procedure. The ARCC was nearly 6 percentage points higher for the holdout method. Row-by-column chi-square analyses of the Jackknife and holdout results for ribotyping indicated no significant difference between the two means of discriminant analysis. Chi-square analysis of the rep-PCR results indicated a significant difference only between the two ARCCs. Therefore, we concluded that, in general, the holdout method did validate the accuracy of the Jackknife procedure.

Practical considerations in application of the two subject procedures for bacterial source identification include ease of performance, cost, and potential for universal application. Reproducibility, technical skill and equipment required, commitment of personnel time, associated efficiency, throughput volume, cost, and robustness must all be taken into account. With respect to both procedures, approximately 5 days are required to culture selected and proven isolates of fecal *E. coli*. The rep-PCR is highly reproducible and generates high-quality patterns approximately 95% of the time. Ribotyping, by contrast, results in initially well-resolved patterns approximately 85% of the time. Manual ribotyping requires a total of 10 to 12 days for total processing while rep-PCR requires only 7 to 8 days. Ribotyping is more rigorous and requires more skilled technician time, and there are more individual steps in the procedure. Ribotyping is performed with purified DNA, and gel patterns must be transferred to Southern blots (18) for hybridization. The efficiency of ribotyping is lower, the cost is higher, and prospects for universal application are less than those for rep-PCR. In summary, rep-PCR is considered superior to the manual ribotyping method.

Implementation of any strategy for bacterial source tracking based on DNA fingerprinting methods (including ribotyping and rep-PCR) will also require recognition of potential limitations. Observations have been made which indicate that there are regional differences in enteric flora of humans and animals (10). It may be expected that *E. coli* strains which populate the intestinal tracts of cattle in one geographic location will differ from those which are typical for cattle in another location. Therefore, it may be necessary to establish a database of fecal *E. coli* strains isolated from human and nonhuman hosts for each watershed in which bacterial source tracking is done. The necessary size of such a database remains to be resolved. Efforts to develop new source tracking methods, based on bacterial markers peculiar to enteric bacteria of the various host species (human and nonhuman), have begun, partly to avoid the database requirement. A promising exam-

ple of this approach is based on *Bacteroides-Prevotella* ribosomal DNA PCR markers which distinguish isolates from human and cattle feces (2).

The hypothesis for application of the subject technology is that various hosts harbor particular and identifiable enteric bacteria. Traceback of bacterial fingerprints to host source would, therefore, be possible. Several reports (15, 19, 21) appear to substantiate this hypothesis. Conflicting studies (8) indicate that enteric bacterial subpopulations change in transition from their intestinal (primary) to their environmental (secondary) habitat and that traceback of environmental isolates of fecal bacteria to host species of origin may be impossible. Further studies must be done to clarify this important question.

REFERENCES

1. **American Public Health Association.** 1995. Standard methods for the examination of water and wastewater, 19th ed. American Public Health Association, Washington, D.C.
2. **Bernhard, A. E., and K. G. Field.** 2000. A PCR assay to discriminate human and animal feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. Appl. Environ. Microbiol. **66:**4571–4574.
3. **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.
4. **Dombek, P. E., L. K. Johnson, S. J. 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.
5. **Duda, R. O., P. E. Hart, and D. G. Stork.** 2001. Pattern classification, 2nd ed. John Wiley and Sons, Inc., New York, N.Y.
6. **Feacham, R.** 1975. An improved role for faecal coliform to faecal streptococci ratios in the differentiation between human and nonhuman pollution sources. Water Res. **9:**689–690.
7. **Fleisher, J. M., D. Kay, R. L. Salmen, F. Jones, M. D. Wyer, and A. F. Godfree.** 1996. Marine waters, contaminated with domestic sewage: nonenteric illnesses associated with bather exposure in the United Kingdom. Am. J. Public Health **86:**1228–1234.
8. **Gordon, D. M., and F. FitzGibbon.** 1999. The distribution of enteric bacteria from Australian mammals: host and geographical effects. Microbiology **145:**2663–2671.
9. **Guan, S., R. Xu, S. Chen, J. Odumeru, and C. Gyles.** 2002. Development of a procedure for discriminating among *Escherichia coli* isolates from animals and human sources. Appl. Environ. Microbiol. **68:**2690–2698.
10. **Hartel, P. G., J. D. Summer, J. L. Hill, J. V. Collins, J. A. Entry, and W. I. Segers.** 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. Environ. Qual. **31:**1273–1278.
11. **Harwood, V. J., J. Whitlock, and V. Withington.** 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in tropical waters. Appl. Environ. Microbiol. **66:**3698–3704.
12. **Huberty, C. J.** 1994. Applied discriminant analysis. John Wiley and Sons, Inc., New York, N.Y.
13. **Jobson, J. D.** 1991. Applied multivariate data analysis, vol. 2. Springer-Verlag, New York, N.Y.
14. **Kariuki, S., C. Gilks, J. Kimari, A. Olanda, F. Muyodi, P. Waiyaki, and C. A. Hart.** 1999. Genotype analysis of *Escherichia coli* strains isolated from children and chickens living in close contact. Appl. Environ. Microbiol. **65:**472–476.
15. **Parveen, S., K. M. Portier, K. Robinson, L. Edminston, 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.
16. **Regnault, B., F. Grimont, and P. A. D. Grimont.** 1997. Universal ribotyping method using chemically labeled oligonucleotide probe mixture. Res. Microbiol. **148:**649–659.
17. **Sauer, T. J., T. C. Daniel, D. J. Nichols, C. P. West, P. A. Moore, Jr., and G. L. Wheeler.** 2000. Runoff water quality from poultry litter-treated pasture and forest sites. J. Environ. Qual. **29:**515–521.
18. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
19. **Souza, V., M. Rocha, A. Vlaera, and L. E. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. Appl. Environ. Microbiol. **65:**3373–3385.
20. **U.S. Environmental Protection Agency.** 1998. 1998 TMDL tracking system data version 1.0. Total maximum daily load program. Office of Water, U.S. Environmental Protection Agency, Washington, D.C. [Online.]
21. **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.
22. **Zierdt, C. H., E. A. Robertson, R. L. Williams, and J. D. MacLowry.** 1980. Computer analysis of *Staphylococcus aureus* phage typing data from 1957 to 1975, citing epidemiological trends and natural evolution within phage typing system. Appl. Environ. Microbiol. **39:**623–629.