Applied and Environmental Microbiology, June 2006, p. 4012–4019 Vol. 72, No. 6
0099-2240/06/$08.00+0 doi:10.1128/AEM.02764-05

Development of Goose- and Duck-Specific DNA Markers To Determine Sources of *Escherichia coli* in Waterways

Matthew J. Hamilton,[1,2] Tao Yan,[2] and Michael J. Sadowsky[1,2,3]*

Department of Microbiology,[1] BioTechnology Institute,[2] and Department of Soil, Water, and Climate,[3] University of Minnesota, St. Paul, Minnesota 55108

Received 22 November 2005/Accepted 29 March 2006

The contamination of waterways with fecal material is a persistent threat to public health. Identification of the sources of fecal contamination is a vital component for abatement strategies and for determination of total maximum daily loads. While phenotypic and genotypic techniques have been used to determine potential sources of fecal bacteria in surface waters, most methods require construction of large known-source libraries, and they often fail to adequately differentiate among environmental isolates originating from different animal sources. In this study, we used pooled genomic tester and driver DNAs in suppression subtractive hybridizations to enrich for host source-specific DNA markers for *Escherichia coli* originating from locally isolated geese. Seven markers were identified. When used as probes in colony hybridization studies, the combined marker DNAs identified 76% of the goose isolates tested and cross-hybridized, on average, with 5% of the human *E. coli* strains and with less than 10% of the strains obtained from other animal hosts. In addition, the combined probes identified 73% of the duck isolates examined, suggesting that they may be useful for determining the contribution of waterfowl to fecal contamination. However, the hybridization probes reacted mainly with *E. coli* isolates obtained from geese in the upper midwestern United States, indicating that there is regional specificity of the markers identified. Coupled with high-throughput, automated macro- and microarray screening, these markers may provide a quantitative, cost-effective, and accurate library-independent method for determining the sources of genetically diverse *E. coli* strains for use in source-tracking studies. However, future efforts to generate DNA markers specific for *E. coli* must include isolates obtained from geographically diverse animal hosts.

The contamination of waterways by pathogenic microorganisms is a persistent threat to public health (40, 43). The waterborne pathogens can be transmitted though drinking water systems, by water-related recreational activities, and by consumption of shellfish (3, 8). Contamination of waterways with fecal material has generally been considered the major source of waterborne pathogens, and total maximum daily loads (TMDLs) are currently being used to abate this type of pollution and restore waterways to their designated uses. Identification of the sources of fecal contamination is a vital component of TMDL determinations, providing information about the type, magnitude, and location of pollutant inputs (46). The sources of fecal coliform bacteria in the environment include runoff from feedlots, manure-amended agricultural land, wildlife, malfunctioning septic systems, urban runoff, sewage discharge, and soilborne bacteria (21, 27).

Phenotypic and genotypic techniques have been used to determine potential sources of fecal bacteria found in surface waters (4, 5, 9, 11, 14, 19, 27, 31, 33, 37–39, 41), and *Escherichia coli* and *Enterococcus* sp. strains are the most widely examined bacteria in such studies. The majority of these methodologies require construction of known-source libraries to differentiate among environmental isolates originating from different animal sources (41). However, since the size of the host source libraries is often limited (many libraries consist of about 35 to about 2,500 isolates [27]), they do not permit adequate determination of potential sources of environmental *E. coli* and *Enterococcus* isolates. Moreover, the utility of known-source libraries is further limited by the lack of representation due to temporal and geographic variations in bacterial genotypes within and between animal species (13, 16, 24, 38), the presence of multiple strains in a single animal (31), host animal diet variation (17), the presence of soil- and alga-borne indicator organisms (7, 21), the presence of transient inhabitants of gastrointestinal tracts, and the great genetic diversity of microorganisms used for source-tracking analyses (27, 31).

Based on these shortcomings, investigators have evaluated the use of library-independent methods to define sources of fecal bacteria in the environment. These methods, which avoid issues of library size and isolate diversity, use both growth-dependent and growth-independent technologies. Enteric viruses have been investigated for use in growth- and library-independent analyses of fecal pollution sources. These studies have revealed that viruses from various animal sources exhibit some level of host specificity (26, 28, 34), and molecular assays have been developed to examine the usefulness of viruses in microbial source-tracking studies (12, 25). Bernhard and Field have been developing 16S rRNA gene-based genetic markers for growth- and library-independent analysis of *Bifidobacterium* and *Bacteroides-Prevotella* for source identification purposes (4, 5). Recently, Dick and coworkers reported effective use of a microplate subtractive hybridization method to define host-specific 16S rRNA-based genetic markers for *Bacteroides*

* Corresponding author. Mailing address: 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.

sp. strains (10). In a separate study, Dick and coworkers (9) analyzed *Bacteroidales* 16S rRNA gene sequences from the feces of eight animals and designed host-specific PCR primers to identify pig- and horse-derived fecal pollution in water. Similarly, Scott and coworkers (39) reported isolation of a host-specific marker gene of *Enterococcus faecium*, coding for a putative virulence factor (*esp*), that allows determination of sources of enterococci in waterways. While these methods show great promise as microbial source-tracking tools, they may be limited by the inability to obtain high-throughput data and by the expense and limitations associated with the use of PCR with environmental samples. In addition, neither system allows correlation with fecal coliform or *E. coli* counts that are commonly obtained by government agencies for freshwater systems.

In this paper, we describe the development and validation of host source-specific genetic markers for *E. coli* strains originating from Canada geese (*Branta canadensis*). These markers were shown to be useful for determining sources of fecal pollution in Lake Superior, and they are useful for high-throughput studies. Instead of randomly screening for host source-specific genes, we took a genomic comparison approach by using suppression subtractive hybridization (SSH) to define host source-specific markers. The SSH technique has been found to be useful for examining genetic diversity in *E. coli* (32), identifying genetic differences between closely related strains (2, 32), examining pathogenicity determinants in *E. coli* (22), and developing probes to examine natural bacterial communities (30). More importantly, the SSH approach has been found to be an effective tool for the development of strain- and host source-specific marker probes (1, 10, 15, 20, 23, 29).

MATERIALS AND METHODS

***E. coli* strains.** The *E. coli* strains used in SSH and subsequent specificity analyses were obtained from a previous library of unique isolates obtained from the feces of 12 known animal host sources (cats, chickens, cows, deer, dogs, ducks, Canada geese, goats, horses, pigs, sheep, and turkeys) and humans (11, 27). All *E. coli* isolates were obtained in Minnesota and Wisconsin from 1998 to 2005. Unique strains were defined as isolates from a single host animal that had DNA fingerprint similarity coefficients less than 90%. Horizontal fluorophore-enhanced repetitive extragenic palindromic PCR (HFERP) DNA fingerprints (27) for *E. coli* strains obtained from goose and human sources were analyzed for genetic relatedness using Pearson's product-moment correlation coefficient with 1% optimization, and dendrograms were generated using the unweighted pair group method with arithmetic means. Based on these analyses, five strains from geese (Go66, Go90, Go126, Go172, and Go206) and five strains from humans (Hu51, Hu130, Hu132, Hu188, and Hu252) that showed maximum differences in genetic relatedness were chosen for suppression subtractive hybridization studies and subsequent probe development. An additional 200 unique *E. coli* isolates were obtained on multiple days in 2004 from the water column 2 m offshore in Lake Superior Harbor in Duluth, MN, as previously described (21). Twenty-seven of these strains were presumptively identified as strains that originated from geese based on HFERP DNA fingerprint comparisons and bootstrap analyses done using known-source fingerprint libraries (21, 27).

To determine if marker DNAs were capable of hybridizing with goose isolates from other geographic areas, 172, 100, 73, and 14 *E. coli* isolates were also obtained from Canada geese in Delaware, West Virginia, Wisconsin, and Indiana, respectively.

Isolation of environmental *E. coli*. Offshore lake water samples were collected from Lake Phalen (St. Paul, MN), an urban lake frequented by Canada geese, using standard procedures (6). Water samples (2 liters) were filtered through 0.45-μm Nuclepore polycarbonate membranes (Whatman, Florham Park, NJ). Bacteria on the membranes were resuspended in phosphate-buffered saline (pH 7.0) using a sterile magnetic stir bar and vortexing to facilitate suspension of the bacterial cells. A total of 1,152 *E. coli* isolates were isolated from the concentrated samples as previously described (11) and stored at −80°C before use.

Suppression subtractive hybridization. SSH was done using the CLONTECH PCR-Select bacterial genome subtraction kit (BD Biosciences CLONTECH, Mountain View, CA) according to the manufacturer's instructions. Genomic DNAs from the five goose *E. coli* strains and five human *E. coli* strains were prepared using a cesium chloride density gradient centrifugation method as previously described (35). Two-microgram aliquots of genomic DNAs from the five goose *E. coli* strains and five human *E. coli* strains were separately pooled and used as tester and driver DNAs, respectively. Prior to the subtraction procedure, 2-μg aliquots of each pooled sample were digested to completion with RsaI. SSH was repeated using PCR-amplified secondary subtraction products as tester DNAs to further enrich for tester-specific fragments. To create a library of potential DNA inserts that were specific for geese, the final subtraction products were cloned into the pGEM-T vector using a T/A cloning procedure (Promega, Madison, WI). A total of 192 clones were randomly selected and stored frozen at −80°C in 50% glycerol until use.

Identification of DNA sequences specific for *E. coli* from geese. The library of cloned potential goose-specific DNA fragments was screened for hybridization specificity using a dot blot procedure as described by Schleicher & Schuell, Keene, NH (http://www.schleicher-schuell.com/bioscience). Cloned insert DNAs were amplified by PCR using nested primers 1 (5′-TCGAGCGGCCGCCCGG GCAGGT-3′) and 2R (5′-AGCGTGGTCGCGGCCGAGGT-3′) provided in the CLONTECH SSH kit. PCRs were carried out using the following conditions: 94°C for 2 min, followed by 25 cycles of 94°C for 30 s, 68°C for 30 s, and 72°C for 1 min and a final elongation step of 2 min at 72°C. PCR products (0.5 μg) were spotted onto duplicate Nytran SuPerCharge nylon membranes (Schleicher & Schuell, Keene, NH) using a dot blot vacuum manifold (Gibco-BRL, Gaithersburg, MD) and the Minifold spotting protocol (Schleicher & Schuell, Keene, N.H.). Membranes were baked at 80°C for 2 h and prehybridized overnight at 42°C in a solution containing 6× SSC, 10× Denhardt's solution, 1% sodium dodecyl sulfate, and 100 μg denatured herring sperm DNA per ml (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) (36). Aliquots (125 ng) of RsaI-digested pooled genomic DNAs from the five human *E. coli* strains or five goose *E. coli* strains were labeled with [α-^{32}P]CTP using a random primer labeling kit (Invitrogen, Carlsbad, Calif.) according to the manufacturer's protocol. Probes were hybridized for 18 h at 46°C to membranes and washed under high-stringency conditions as previously described (36). Images were captured using a STORM 840 densitometer (Molecular Dynamics, Piscataway, NJ). Presumptive goose-specific DNA inserts were identified on the basis of visual differences in hybridization intensity.

Plasmids were isolated from presumptive goose-specific clones using a QIAprep Spin miniprep kit (QIAGEN, Valencia, CA) according to the manufacturer's protocol. Insert DNA was amplified by PCR using nested primers 1 and 2R as described above and electrophoresed on 2% agarose gels. DNAs were transferred to Nytran SuPerCharge nylon membranes as described previously (36). The membranes were probed with the RsaI-digested, pooled, genomic DNAs as described above.

DNA sequencing and analysis. Confirmed goose-specific DNA inserts were sequenced in both directions using pUC/M13 universal forward (5′-CGCCAG GGTTTTCCCAGTC ACGAC-3′) and reverse (5′-TCACACAGGAAACAGC TATGAC-3′) sequencing primers. Sequencing reactions were performed using BigDye (Applied Biosystems, Foster City, CA) sequencing chemistry at the Advanced Genetic Analysis Center, University of Minnesota, St. Paul. Translated sequences were analyzed using the BLASTX algorithm at NCBI (http://www.ncbi.nlm.nih.gov/BLAST) and the GenBank and *E. coli* databases.

Colony hybridization for probe evaluation and environmental application. The specificity of subtracted DNA inserts was evaluated by colony hybridization to 48 cat, 96 chicken, 96 cow, 96 deer, 96 dog, 81 duck, 135 goose, 42 goat, 78 horse, 210 human, 96 pig, 60 sheep, and 96 turkey *E. coli* isolates (27). An additional 27 *E. coli* strains isolated from Lake Superior Harbor in Duluth, MN, 1,152 isolates from Lake Phalen (St. Paul, MN), and 359 isolates from Canada geese obtained in Delaware, West Virginia, Wisconsin, and Indiana were also evaluated by colony hybridization. Probe specificity was evaluated using blind samples consisting of 96 randomly selected isolates obtained from geese, horses, pigs, sheep, and humans. *E. coli* strains were inoculated from frozen stocks onto Nytran SuPerCharge membranes (20 cm^2; Schleicher & Schuell, Keene, NH) using a 48-pin multiple inoculator. The membranes were placed onto the surfaces of LB (36) agar plates (22 by 22 cm; Qtray Genetix, United Kingdom) and incubated at 37°C for approximately 5 h. Colonies were lysed, and DNA on the membranes was processed as described previously (36). Membranes were prehybridized at 68°C overnight in a solution containing 6× SSC, 10× Denhardt's solution, and 100 μg denatured herring sperm DNA per ml. Probes from insert

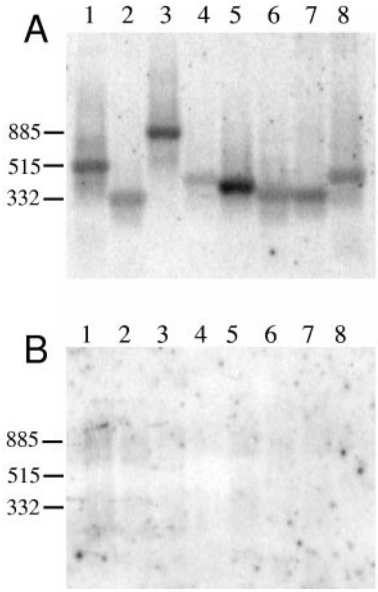

FIG. 1. Southern hybridization of eight SSH-derived, PCR-amplified insert DNAs with ^{32}P-labeled RsaI-digested pooled genomic DNAs from *E. coli* isolates obtained from geese (A) and humans (B). Panels A and B show duplicate membranes probed with the genomic DNAs.

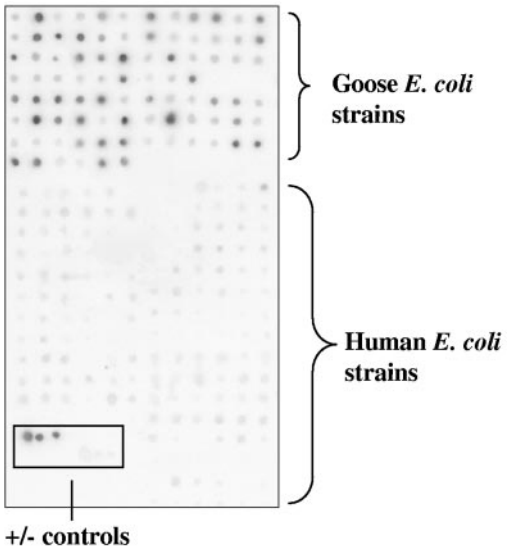

FIG. 2. Colony hybridization of ^{32}P-labeled GE11 insert DNA with unique *E. coli* isolates obtained from geese and humans. The positive and negative control strains are enclosed in a box.

DNAs (50 ng) were labeled using the Random Primer DNA labeling system (Invitrogen, Carlsbad, Calif.) according to the manufacturer's protocol. Membranes were hybridized overnight at 68°C in a solution containing 6× SSC, 10× Denhardt's solution, and 100 μg denatured herring sperm DNA per ml. Blots were finally washed in 0.1× SSC–0.1% sodium dodecyl sulfate at 65°C, and images were obtained as described below.

Quantitative image analysis. Quantitative image analysis was used to determine positive and negative signals on colony hybridization membranes. Images were captured using a STORM 840 densitometer (Molecular Dynamics, Piscataway, NJ) and were analyzed using the ScanAlyze version 2.50 software (http://rana.lbl.gov/EisenSoftware.htm). The normalized intensity of each spot was calculated by subtracting the median intensity of the background from the mean intensity of each spot. Normalized spot intensities were plotted using the Sigma Plot version 8.0 software (Systat Software, Point Richmond, CA), and a cutoff value was assigned based on normalized mean intensities of negative control spots plus three times the standard deviation.

Nucleotide sequence accession numbers. The nucleotide sequences obtained in this study have been deposited in the GenBank database under accession numbers DQ300500 to DQ300502 and DQ300504 to DQ300507.

RESULTS

Isolation of goose-specific DNA fragments. Following SSH, 192 putative goose-specific DNA clones were randomly selected to create a DNA subtraction library. The cloned insert DNAs were initially screened using a dot blot protocol to determine hybridization specificity. Twenty clones exhibited increased hybridization intensity when they were probed with labeled RsaI-digested genomic DNAs from the five pooled *E. coli* strains from geese compared to the intensity seen with the pooled *E. coli* genomic DNAs from humans. The hybridization specificity of DNA inserts from these clones was further evaluated by Southern hybridization using the probes that were used in the dot blot hybridizations. Southern hybridization analyses indicated that 17 cloned insert DNAs were goose specific. The results of Southern hybridization analyses of a representative group of eight goose-specific insert DNAs are shown in Fig. 1.

Analysis of insert specificity. While Southern hybridization analyses confirmed that several of the cloned DNA fragments hybridized specifically to goose genomic DNAs, this specificity analysis was limited to probes derived from the goose and human *E. coli* strains used in the initial SSH procedure. To examine the hybridization specificity of the clones in more detail, colony hybridization experiments were done to identify cloned insert DNAs that hybridized with many *E. coli* strains from geese and with only a few strains from humans. A library consisting of 135 and 210 unique *E. coli* isolates from geese and humans, respectively, was cultured on nylon membranes and individually probed with 14 of the ^{32}P-labeled PCR-amplified insert DNAs from the confirmed goose-specific clones. The remaining three cloned insert DNAs were not evaluated further since they were duplicates of existing clones. A representative image of a colony hybridization membrane is shown in Fig. 2. DNAs from the five goose *E. coli* strains and five human *E. coli* strains that were used in SSH were used as references for determining positive and negative hybridization signals, respectively, and quantitative image analysis was performed to determine the pixel intensities of the individual colony spots (Fig. 3). The cutoff value was determined to be the mean intensity of the five human strains plus three times the standard deviation. Based on these analyses, 7 of 14 (50%) goose-specific DNA inserts (GA9, GB2, GD5, GE3, GE11, GF5, and GG11) exhibited specific hybridization with goose *E. coli* strains compared to the hybridization seen with strains isolated from humans (Table 1). The insert DNAs hybridized with 20.7 to 48.1% of the 135 unique goose strains tested. In contrast, the insert DNAs tested cross-hybridized with 1 to 10% of the 210 *E. coli* strains from humans. Insert DNAs GB2 and GE11 hybridized to the greatest number of goose isolates (48.1% of the isolates). Together, the seven probes hybridized with about 76% of the *E. coli* strains from geese and cross-hybridized, on

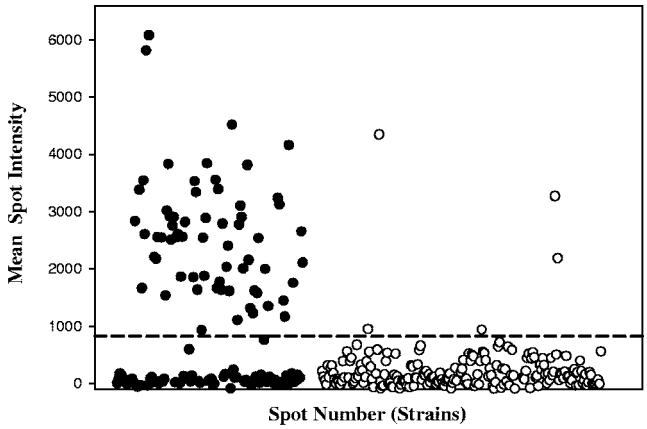

FIG. 3. Pixel intensities for a colony hybridization membrane containing 134 and 209 unique *E. coli* strains isolated from geese (●) and humans (○), respectively. The membrane was probed with ^{32}P-labeled DNA of marker insert GE11. The dashed line indicates a cutoff value for determining positive and negative signals.

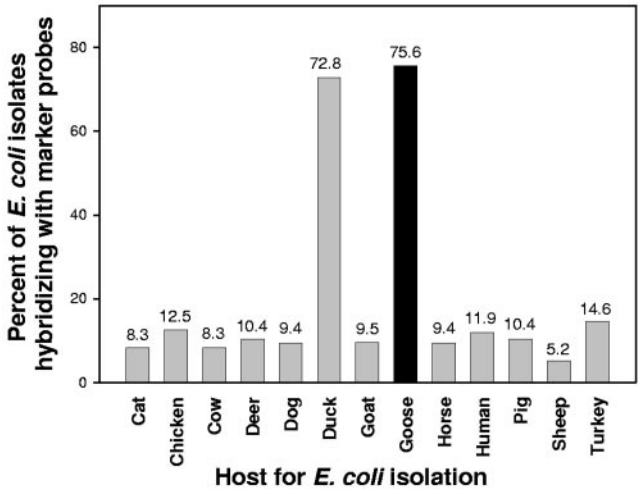

FIG. 4. Percentages of *E. coli* strains hybridizing with ^{32}P-labeled, pooled insert GB2 and GE11 marker DNAs obtained by colony hybridization and pixel intensity analysis. The values above the bars are hybridization percentages.

average, with 5% of the human *E. coli* strains. These hybridization experiments were repeated twice in triplicate to verify the results.

Host specificity determination. Since the goose-specific marker DNAs identified will ultimately be used to examine *E. coli* in natural habitats, it is important to determine whether the probes cross-hybridize with *E. coli* from other host animal species. To examine this, we hybridized each ^{32}P-labeled insert DNA probe to 891 unique *E. coli* strains isolated from cats, chickens, cows, deer, ducks, goats, horses, humans, pigs, sheep, and turkeys. The results, summarized in Fig. 4, showed that the probes hybridized to 76% of the goose isolates examined. Similarly, the probes cross-hybridized to 73% of the duck isolates. In contrast, the probes cross-hybridized with a limited number of *E. coli* isolates from other host species, and the greatest cross-hybridization occurred with *E. coli* isolates from turkeys (14.6%) and chickens (12.5%). These results indicated that the greatest cross-hybridization occurred with *E. coli* isolates from avian hosts. The mean frequency of false-positive cross-hybridization of the probes to *E. coli* from other host sources was about 9%.

Hybridization specificity was also evaluated by using a blind sample consisting of 96 isolates, including 19 goose, 20 horse, 20 pig, 20 sheep, and 17 human *E. coli* strains. The seven probes evaluated (GA9, GB2, GD5, GE3, GE11, GF5, and GG11) hybridized with 14 of 19 goose strains (73.7%) and only 6 of 77 (7.8%) of the strains from other animals (data not shown).

TABLE 1. Goose-specific marker DNAs isolated using suppression subtractive hybridization

Marker DNA	% of *E. coli* isolates hybridizing with marker DNA probes	
	Goose isolates (n = 135)[a]	Human isolates (n = 210)
GA9	27.4	1.9
GB2	48.1	3.3
GD5	30.4	9.1
GE3	23.6	7.2
GE11	48.1	4.8
GF5	20.7	10.0
GG11	31.1	1.0

[a] n is the total number of strains examined by colony hybridization.

Environmental *E. coli* and geographic analyses. To examine the correlation between the results obtained using the new markers described in this paper and the results obtained using other methods, we isolated about 200 *E. coli* strains from Duluth harbor water and analyzed them first by using the HFERP DNA fingerprinting technique and then by hybridization using combined ^{32}P-labeled insert DNAs GB2 and GE11. Of the 200 *E. coli* isolates examined, 27 (13.5%) were identified as isolates that likely originated from geese using the HFERP DNA fingerprinting technique, a comprehensive known-source DNA fingerprint library, and ID bootstrap analysis with a P value of ≥0.9 (27). When isolates were screened by colony hybridization to a pooled GB2/GE11 insert DNA probe, 22 of 27 strains hybridized with the probe. This corresponded to 81.5% agreement between HFERP classification and marker probe analysis using the GB2/GE11 screening system described here. The applicability of DNA marker technology was also demonstrated by screening randomly selected environmental *E. coli* isolates from Lake Phalen, a local urban lake frequently affected by Canada geese. Of the 1,152 isolates examined, 301 (26.1%) tested positive with the GB2 and GE11 probes.

To determine if the DNA markers used could identify *E. coli* from geese obtained from other geographic regions of the United States, we hybridized probes GB2 and GE11 with an additional 359 goose isolates obtained from Delaware, Indiana, Wisconsin, and West Virginia. The results of this experiment demonstrated that only 24.0% of the isolates hybridized to the marker DNAs (data not shown). Probes GB2 and GE11 hybridized to 20, 28, 38, and 20% of the goose *E. coli* strains from Delaware, Indiana, Wisconsin, and West Virginia, respectively.

TABLE 2. Insert marker DNAs showing hybridization specificity with *E. coli* isolates from geese

Insert DNA	Length (bp)	Protein homolog in database	GenBank accession no.	No. of identical amino acids/ no. examined[a]	E value
GA9	515	Type III secretion apparatus protein (*E. coli* O157:H7 EDL933)	AAG57987	61/161 (37)	1.00E-26
GB2	332	AIDA-I adhesin-like protein (*E. coli* O157:H7 RIMD 0509952)	BAB33785	81/123 (65)	2.20E-40
GD5	885	TraT (*E. coli* plasmid R1)	AAT85681	112/132 (85)	2.00E-57
GE3	380	NikB (*E. coli* O157:H7 RIMD 0509952)	NP_052661	30/88 (34)	2.00E-05
GE11	336	AIDA-I adhesin-like protein (*E. coli* O157:H7 RIMD 0509952)	BAB33785	81/123 (66)	1.00E-40
GF5	346	ORF5 (no significant homologous proteins in database) (*E. coli* B171)	AAB36834	57/58 (98)	2.00E-27
GG11	427	Type III secretion protein EprH (*E. coli* O157:H7 RIMD 0509952)	BAB37142	31/101 (32)	1.00E-11

[a] The values in parentheses are the percentages of identity with database entries.

Sequencing and BLAST searches. The seven confirmed goose- and duck-specific DNA inserts were sequenced in both directions, and translated sequences were subjected to BLASTX analyses using *E. coli* protein databases at NCBI. The sequenced inserts were between 332 and 885 bp long. The results of BLASTX homology searches are summarized in Table 2. The GB2 and GE11 inserts, each of which hybridized to about 48% of the *E. coli* strains from geese, were 93% identical to each another at the nucleotide level. When the sequences were translated, there was significant amino acid homology (65% and 66% amino acid identity, respectively) to the C-terminal fragment of the AIDA-I adhesin-like protein of *E. coli* O157:H7 (GenBank accession no. BAB33785). The GD5 insert product exhibited 89% amino acid identity to a fragment of the TraT complement resistance protein of *E. coli* (accession no. AAT85681), and the GF5 insert was 98% identical to ORF5 in *E. coli*, with no significant matches to any entries in the database. Other matches with less than 50% amino acid identity to proteins in the database included two type III secretion machinery proteins from *E. coli* O157:H7 (accession no. AAG57987 and BAB37142) and a NikB nickase (accession no. NP_052661).

DISCUSSION

In this study, SSH was successfully used to identify seven DNA markers with high levels of hybridization specificity for *E. coli* strains originating from geese. Combined, the marker DNAs were capable of identifying about 76% of the goose *E. coli* strains examined and 73% of the duck *E. coli* strains examined. In contrast, on average, the probes cross-reacted with about 10% of the *E. coli* isolates from other host species. As our goal was to identify sequences specific for goose strains, we adapted the standard SSH protocol by using pooled genomic DNAs from multiple goose and human strains as the tester and driver DNAs, respectively. By using pooled genomic DNAs rather than DNA from a single strain, we expected that more genetic diversity among the goose *E. coli* isolates could be uncovered and that the subtraction products obtained would more likely be present in other goose isolates than in *E. coli* strains from humans. Thus, the method employed was expected to enrich for sequences found in all of the pooled tester genomes rather than fragments present in only a single genome. This hypothesis was shown to be true by the presence of very similar, but not identical, DNA sequences in inserts GB2 and GE11. An additional clone with 100% identity to GE11 was also identified using this approach, but it was not used in further analyses (data not shown).

One downside of using multiple tester DNAs is reduced subtraction efficiency due to the increased complexity introduced into the reaction. Generally, genome subtraction yields greater than 25% tester-specific sequences after screening (CLONTECH, Mountain View, CA), compared to the approximately 9% efficiency that was observed in this study. However, reduced efficiency was not found to be an issue with the screening procedures that we employed, and for our purposes increased hybridization specificity and the ability to identify more isolates are the most important parameters. Seven goose-specific insert DNAs exhibited increased hybridization with strains isolated from geese compared to the hybridization with isolates obtained from humans. While these insert DNAs each hybridized with less than one-half of the goose isolates tested, revealing genetic diversity in goose *E. coli* strains, together the inserts identified 76 and 72% of the *E. coli* isolates from goose and ducks, respectively. Consequently, subsequent field studies should be done using pooled insert DNAs as hybridization probes.

When the sequences were translated, the products of the nearly identical insert DNAs GB2 and GE11 exhibited 65% amino acid identity to the C-terminal portion of the AIDA-I adhesin-like protein of *E. coli* strain O157:H7. This result suggests that inserts GB2 and GE11 are fragments of an unidentified adhesin-like gene. As adhesins mediate the attachment of bacteria to host tissues (45), it seems plausible and logical that this putative gene may mediate the attachment of specific *E. coli* isolates to the goose intestinal tract. Attachment to the host intestinal epithelium is the necessary first step in gut colonization (45), and, therefore, the putative gene may be responsible for preferential colonization of the goose host. If this hypothesis is validated by experimental in vivo colonization data, other adhesin genes that participate in host-specific colonization may also represent ecologically meaningful markers that can be targeted for microbial source-tracking purposes.

Since together the seven DNA inserts hybridized with 76% of goose isolates, we examined whether the probes cross-hybridized with isolates from cats, chickens, cows, deer, ducks, goats, horses, humans, pigs, sheep, and turkeys. Interestingly, the seven probes cross-hybridized with 73% of the *E. coli* isolates from ducks and with 14.6 and 12.5% of the isolates from turkeys and chickens, respectively, but with only about

10% of the *E. coli* strains from other hosts. However, the results of preliminary studies indicated that the GB2 and GE11 probes cross-hybridized with 11 and 9% of gull and tern *E. coli* isolates, respectively. Presumably, these results are due to the close genetic relationship between chickens, ducks, geese, and turkeys and may indicate that the intestinal tracts of some avian species can be colonized by the same *E. coli* strain. Alternately, they may reflect the cosmopolitan nature of some *E. coli* strains (47), a transient intestinal population structure (18), a lack of host specificity in this subgroup of *E. coli*, or the presence of multiple adhesins that mediate colonization (44).

Recently, Soule et al. (42) used a microarray approach to identify several DNA markers from *Enterococcus* sp. that were subsequently used to develop host-specific PCR primers. While many of the markers identified were specific for *Enterococcus* isolates from targeted host species, they often failed to detect a high percentage of the isolates from these hosts. However, other markers detected from 27 to 45% of the enterococci from targeted host species, but they also detected 1.1 to 7.1% of the nontargeted isolates. This result is similar to cross-reactions that we found in the current study using the DNA probes (Fig. 4). In contrast, Bernhard and Field (4) and Dick et al. (10) reported that PCR primers for *Bacteroidales* did not detect nontargeted hosts, suggesting that the markers which they used were more specific than those found in our study. However, these authors analyzed diluted fecal samples and DNAs rather than individual colonies, making direct comparisons to our method difficult.

Results obtained from screening water isolates from Lake Superior with the combined GB2/GE11 probe compared favorably with results obtained using the HFERP DNA fingerprinting method for assigning isolates to host source groups. Of the 27 isolates assigned to goose sources by HFERP, 22 (81.5%) had a positive hybridization signal with the GB2/GE11 probe. While the library-dependent HFERP method was previously shown to correctly identify about 70% of the waterfowl isolates in a known-source library (27) and far fewer environmental isolates, the method described here is a vast improvement for accurately and quantitatively determining the host origins of environmental isolates. Moreover, with the library-independent hybridization-based marker method there are fewer false-positive and false-negative reactions than there are with the HFERP and other techniques that have been evaluated recently, except for host-specific PCR analysis (14). The applicability of this DNA marker technology was also evaluated by screening *E. coli* isolates from Lake Phalen, a local urban lake frequently affected by Canada geese. The results of this analysis indicated that 26% of the 1,152 isolates examined hybridized with the GB2 and GE11 probes. These data further illustrate that the DNA markers identified can be used for environmental isolates. Considerably greater numbers of environmental isolates will most likely be found if hybridizations are done using the seven combined markers. Large-scale field studies using the combined seven probes will be done in the summer of 2006 to assess the impact of geese on Lake Superior beaches.

To assess whether the DNA markers allowed detection of goose *E. coli* strains from different geographic regions, we obtained isolates from eastern and midwestern United States. The results of our studies indicated that the combined GB2 and GE11 probes identified only 24% of the isolates examined. While the level of identification most likely would increase if all seven marker probes were used, our results suggest that *E. coli* strains are geographically distributed. Since the library that we used was constructed with goose *E. coli* strains isolated from two locations in Minnesota, it is not surprising that the highest percentage of strains identified were isolated in Wisconsin, a bordering state. Consequently, future efforts in which SSH is used to generate DNA markers specific for animal hosts should be done with tester strains originating from several regions of the United States.

In the past, the development of microbial source-tracking techniques has focused on library-dependent methods (37, 41). However, these methods suffer from the need to develop and maintain large reference libraries for comparisons with environmental isolates. Additionally, geographic and temporal variability in isolates, transportability issues, the inability to assign many environmental isolates to source groups, the large library sizes needed to adequately capture genetic diversity, and the high levels of false-positive and false-negative assignments make these methods difficult to implement at a large and economically feasible scale (14, 27). In contrast, library-independent methods that screen for host-specific and ecologically meaningful genes alleviate many of these issues. These genes most likely would not be influenced by geographic and temporal variability, as they would be stable in bacterial isolates obtained recently from a specific host source. While library-independent marker gene approaches have recently been investigated as source-tracking tools with members of the genus *Bifidobacterium* and the *Bacteroides-Prevotella* group (4, 5), these organisms are rarely quantified in routine analyses of fecal bacteria in waterways. Conversely, *E. coli* is becoming one of the most frequently monitored indicators of fecal contamination of freshwater systems, and thus, source-tracking information obtained using the markers reported here can be easily coupled with existing and new fecal count data for TMDL analyses and abatement strategies. Recently, a library-independent marker gene method has also been developed for *Enterococcus* species (39), allowing similar analyses for saltwater environments.

Since waterways are most often contaminated by fecal bacteria originating from several different sources rather than a single animal host species, it is frequently necessary to screen large numbers of isolates for accurate determination of host sources (31). The development of host-specific DNA fragments for screening by colony hybridization provides a cost-effective quantitative method for simultaneous analysis of many bacterial isolates. Moreover, this method can be easily adapted for automated, rapid, and high-throughput macro- and microarray screening strategies, reducing the time and expense of analyzing the thousands of isolates needed for large-scale and accurate source-tracking studies.

In summary, our results provide evidence that SSH is an effective tool for identification of ecologically meaningful marker DNAs that can be used to identify a large number of genetically diverse *E. coli* isolates originating from geese. While our initial studies indicated that these markers can be effectively used as hybridization probes to determine the source of environmental *E coli* isolates, more extensive field testing is needed before large-scale microbial source-tracking

studies can be initiated. Nevertheless, we believe that the SSH approach will allow us to identify additional markers for *E. coli* strains from humans and other animals and to obtain more comprehensive information about sources of fecal contamination in waterways. Coupled with high-throughput, automated macro- and microarray screening, these markers may provide a cost-effective, quantitative, and accurate method for determining sources of genetically diverse *E. coli* strains for use in water quality analyses and TMDL determinations.

ACKNOWLEDGMENTS

This work was supported in part by grants from the University of Minnesota Agricultural Experiment Station and the BioTechnology Institute (to M.J.S.) and by training grant 2T32-GM008347 from the National Institutes of Health (to M.J.H.).

We thank Satoshi Ishii, Sam Myoda, Cindy Nakatsu, Don Stoeckel, and Greg Kleinheinz for providing *E. coli* isolates and John Ferguson for help with the blind studies, cluster analyses, and library maintenance.

REFERENCES

1. **Agron, P. G., R. L. Walker, H. Kinde, S. J. Sawyer, D. C. Hayes, J. Wollard, and G. L. Andersen.** 2001. Identification by subtractive hybridization of sequences specific for *Salmonella enterica* serovar Enteritidis. Appl. Environ. Microbiol. **67:**4984–4991.
2. **Akopyants, N. S., A. Fradkov, L. Diatchenko, J. E. Hill, P. D. Siebert, S. A. Lukyanov, E. D. Sverdlov, and D. E. Berg.** 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **95:**13108–13113.
3. **Alexander, L. M., A. Heaven, A. Tennant, and R. Morris.** 1992. Symptomatology of children in contact with sea water contaminated with sewage. J. Epidemiol. Community Health **46:**340–344.
4. **Bernhard, A. E., and K. G. Field.** 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. Appl. Environ. Microbiol. **66:**4571–4574.
5. **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.
6. **Bordner, R., and J. A. Winter.** 1978. Microbiological methods for monitoring the environment, water, and wastes. EPA 600/8-78-017. U.S. Environmental Protection Agency, Washington, D.C.
7. **Byappanahalli, M. N., D. A. Shively, M. B. Nevers, M. J. Sadowsky, and R. L. Whitman.** 2003. Growth and survival of *E. coli* and enterococci populations in the macro-alga *Cladophora* (*Chlorophyta*). FEMS Microbiol. Ecol. **46:**203–211.
8. **Centers for Disease Control and Prevention.** 1998. Outbreak of *Vibrio parahaemolyticus* infections associated with eating raw oysters, Pacific Northwest, 1997. Morb. Mortal. Wkly. Rep. **147:**45762.
9. **Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters, and K. G. Field.** 2005. Host distributions of uncultivated fecal *Bacteroidales* reveal genetic markers for fecal source identification. Appl. Environ. Microbiol. **71:**3184–3191.
10. **Dick, L. K., M. T. Simonich, and K. G. Field.** 2005. Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. Appl. Environ. Microbiol. **71:**3179–3183.
11. **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.
12. **Fong, T. T., D. W. Griffin, and E. K. Lipp.** 2005. Molecular assays for targeting human and bovine enteric viruses in coastal waters and their application for library-independent source tracking. Appl. Environ. Microbiol. **71:**2070–2078.
13. **Gordon, D. M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiology **147:**1079–1085.
14. **Griffith, J. F., S. B. Weisberg, and C. D. McGee.** 2003. Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. J. Water Health **1:**141–151.
15. **Harakava, R., and D. W. Gabriel.** 2003. Genetic differences between two strains of *Xylella fastidiosa* revealed by suppression subtractive hybridization. Appl. Environ. Microbiol. **69:**1315–1319.
16. **Hartel, P. G., J. D. Summer, J. L. Hill, J. Collins, J. A. Entry, and W. I. Segars.** 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. Environ. Qual. **31:**1273–1278.
17. **Hartel, P. G., J. D. Summer, and W. I. Segars.** 2003. Deer diet affects ribotype diversity of *Escherichia coli* for bacterial source tracking. Water Res. **37:**3263–3268.
18. **Hartl, D. L., and D. E. Dykhuizen.** 1984. The population genetics of *Escherichia coli*. Annu. Rev. Genet. **18:**31–68.
19. **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.
20. **Hsieh, W. J., and M. J. Pan.** 2004. Identification *Leptospira santarosai* serovar shermani specific sequences by suppression subtractive hybridization. FEMS Microbiol. Lett. **235:**117–124.
21. **Ishii, S., W. B. Ksoll, R. E. Hicks, and M. J. Sadowsky.** 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. Appl. Environ. Microbiol. **72:**612–621.
22. **Janke, B., U. Dobrindt, J. Hacker, and G. Blum-Oehler.** 2001. A subtractive hybridisation analysis of genomic differences between the uropathogenic *E. coli* strain 536 and the *E. coli* K-12 strain MG1655. FEMS Microbiol. Lett. **199:**61–66.
23. **Janssen, P. J., B. Audit, and C. A. Ouzounis.** 2001. Strain-specific genes of *Helicobacter pylori*: distribution, function and dynamics. Nucleic Acids Res. **29:**4395–4404.
24. **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.
25. **Jiang, S., R. Noble, and W. Chu.** 2001. Human adenoviruses and coliphages in urban runoff-impacted coastal waters of Southern California. Appl. Environ. Microbiol. **67:**179–184.
26. **Jiménez-Clavero, M. A., C. Fernández, J. A. Ortiz, J. Pro, G. Carbonell, J. V. Tarazona, N. Roblas, and V. Ley.** 2003. Teschoviruses as indicators of porcine fecal contamination of surface water. Appl. Environ. Microbiol. **69:**6311–6315.
27. **Johnson, L. K., M. B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky.** 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. Appl. Environ. Microbiol. **70:**4478–4485.
28. **Ley, V., J. Higgins, and R. Fayer.** 2002. Bovine enteroviruses as indicators of fecal contamination. Appl. Environ. Microbiol. **68:**3455–3461.
29. **Liu, L., T. Spilker, T. Coenye, and J. J. LiPuma.** 2003. Identification by subtractive hybridization of a novel insertion element specific for two widespread *Burkholderia cepacia* genomovar III strains. J. Clin. Microbiol. **41:**2471–2476.
30. **Mau, M., and K. N. Timmis.** 1998. Use of subtractive hybridization to design habitat-based oligonucleotide probes for investigation of natural bacterial communities. Appl. Environ. Microbiol. **64:**185–191.
31. **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. Appl. Environ. Microbiol. **69:**2587–2594.
32. **Mokady, D., U. Gophna, and E. Z. Ron.** 2005. Extensive gene diversity in septicemic *Escherichia coli* strains. J. Clin. Microbiol. **43:**66–73.
33. **Parveen, S., N. C. Hodge, R. E. Stall, S. R. Farrah, and M. L. Tamplin.** 2001. Genotypic and phenotypic characterization of human and nonhuman *Escherichia coli*. Water Res. **35:**379–386.
34. **Pina, S., M. Puig, F. Lucena, J. Jofre, and R. Girones.** 1998. Viral pollution in the environment and in shellfish: human adenovirus detection by PCR as an index of human viruses. Appl. Environ. Microbiol. **64:**3376–3382.
35. **Sadowsky, M. J., R. E. Tully, P. B. Cregan, and H. H. Keyser.** 1987. Genetic diversity in *Bradyrhizobium japonicum* serogroup 123 and its relation to genotype-specific nodulation of soybeans. Appl. Environ. Microbiol. **53:**2624–2630.
36. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. **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.
38. **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.
39. **Scott, T. M., T. M. Jenkins, J. Lukasik, and J. B. Rose.** 2005. Potential use of a host associated molecular marker in *Enterococcus faecium* as an index of human fecal pollution. Environ. Sci. Technol. **39:**283–287.
40. **Sharma, S., P. Sachdeva, and J. S. Virdi.** 2003. Emerging water-borne pathogens. Appl. Microbiol. Biotechnol. **61:**424–428.
41. **Simpson, J. M., J. W. Santo Domingo, and D. J. Reasoner.** 2003. Microbial source tracking: state of the science. Environ. Sci. Technol. **36:**5280–5288.

42. **Soule, M., E. Kuhn, F. Loge, J. Gay, and D. R. Call.** 2006. Using DNA microarrays to identify library-independent markers for bacterial source tracking. Appl. Environ. Microbiol. **72:**1843–1851.
43. **Szewzyk, U., R. Szewzyk, W. Manz, and K. H. Schleifer.** 2000. Microbiological safety of drinking water. Annu. Rev. Microbiol. **54:**81–127.
44. **Torres, A. G., and J. B. Kaper.** 2003. Multiple elements controlling adherence of enterohemorrhagic *Escherichia coli* O157:H7 to HeLa cells. Infect. Immun. **71:**4985–4995.
45. **Torres, A. G., X. Zhou, and J. B. Kaper.** 2005. Adherence of diarrheagenic *Escherichia coli* strains to epithelial cells. Infect. Immun. **73:**18–29.
46. **U.S. Environmental Protection Agency.** 2001. Protocol for developing pathogen TMDLs. EPA 841-R-00-002. Office of Water, U.S. Environmental Protection Agency, Washington, D.C.
47. **Whitlock, J. E., D. T. Jones, and V. J. Harwood.** 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. Water Res. **36:**4273–4282.