WATER RESEARCH 41 (2007) 3561–3574

ELSEVIER

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/watres

Identification of chicken-specific fecal microbial sequences using a metagenomic approach

Jingrang Lu, Jorge Santo Domingo, Orin C. Shanks*

US Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, 26 West Martin Luther King Drive, MS-387, Cincinnati, OH, USA

ARTICLE INFO

Article history:
Received 20 February 2007
Received in revised form
7 May 2007
Accepted 15 May 2007
Available online 24 May 2007

Keywords:
Chicken fecal DNA
Competitive genomic hybridization
Genetic markers
Fecal microbial community

ABSTRACT

In this study, we applied a genome fragment enrichment (GFE) method to select for genomic regions that differ among different fecal metagenomes. Competitive DNA hybridizations were performed between chicken fecal DNA and pig fecal DNA (CP) and between chicken fecal DNA and an avian DNA composite consisting of turkey, goose, and seagull fecal DNA extracts (CB) to enrich for chicken-specific DNA fragments. A total of 471 non-redundant chicken metagenomic sequences were retrieved and analyzed. All of the clone sequences were similar to prokaryotic genes, of which more than 60% could not be assigned to previously characterized functional roles. In general terms, sequences assigned characterized functional roles were associated with cellular processes (11.7%), metabolism (11.0%) and information storage and processing (13.4%). Approximately 53% of the non-redundant sequences are similar to genes present in intestinal bacteria belonging to *Clostridia* (20.9%), *Bacteroidetes* (15.0%), and *Bacilli* (17.3%). Twenty-five sequences from the CP and CB clone libraries were selected to develop chicken fecal-specific PCR assays. These assays were challenged against fecal DNA extracted from 21 different animal species, including mammals and birds. The results from the host-specificity studies showed that 12 of the assays had a high degree of specificity to chicken feces. In addition, three assays were specific to chicken and turkey while another four assays tested positive to more than two avian species, suggesting a broader distribution of some of the enriched gene fragments among different avian fecal microbial communities. Fecal pollution signals were detected using chicken-specific assays in contaminated water samples, although the PCR assays showed different detection limits. These results indicate the need for multiple assays to detect poultry fecal sources of pollution. The competitive DNA hybridization approach used in this study can rapidly select for numerous chicken fecal metagenomic regions that can be used as potential genetic markers for fecal source tracking.

Published by Elsevier Ltd.

1. Introduction

Poultry farming is a worldwide practice of meat production that has significantly increased in the last few decades. For example, the per capita consumption of chicken in the United States alone was estimated to be over 84 pounds in 2004, which represents a 200% increase in less than 20 years ⟨http://www.ers.usda.gov/Data/foodconsumption/FoodAvailIndex.htm⟩. As a result of this increase in production, fecal matter has become a significant by-product of the poultry industry, which in many cases has been used as fertilizer in the form of raw or composted manure. A potential risk arising from the

*Corresponding author. Tel.: +1 513 569 7085; fax: +1 513 569 7328.
E-mail address: santodomingo.jorge@epa.gov (J. Santo Domingo).
0043-1354/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.watres.2007.05.033

PLAINTIFF'S EXHIBIT
L

526

disposal of poultry fecal waste is the spread of enteric pathogens, such as *Escherichia coli* O157:H7, *Salmonella* spp., and *Campylobacter* spp. (Altekruse et al., 1997; Martin 1998) and viruses (Guan et al., 2000). These pathogens can reach watersheds after rainfall events, and thereby increase risks associated with recreational use of waterways. Furthermore, environmental concerns also include high nutrient loads (such as nitrate and phosphate) from runoff to streams, ponds, and ground water (Stuven et al., 2001). Methods that can specifically detect poultry fecal pollution are therefore needed to assist in the development and evaluation of adequate management practices targeting pollution control.

A limited number of studies have reported on the use of genotypic methods to identify the presence of poultry fecal contamination in surface waters. Ribotyping and rep-PCR DNA fingerprint techniques targeting *E. coli* isolates have been applied to discriminate among different animal fecal sources, including chicken and human fecal sources (Parveen et al., 1999; Carson et al., 2001; Hartel et al., 2002). However, the successful application of these genotypic methods depends on the development of large fingerprint databases of indicator bacterial isolates, primarily *E. coli* (Jenkins et al., 2003). Moreover, the use of *E. coli* for fecal source identification has been recently criticized in light of the abundance of secondary habitat populations that are capable of adapting to conditions outside of the animal gut and as a result contribute to the levels of fecal indicator bacteria in watersheds (Gordon et al., 2002). Recently, Field and coworkers used library-independent methods based on ribosomal 16S rRNA gene (i.e., 16S rDNA) sequences of *Bacteroides*-like bacteria to discriminate between human and ruminant feces (Bernhard and Field, 2000a; Field et al., 2003). These *Bacteroides* markers have been used in the identification of non-point sources of fecal pollution in coastal and inland waters (Bernhard and Field 2000b; Lamendella et al., 2007). Analyses of bacterial rDNA sequences from chicken fecal DNA extracts suggest that the chicken cecum and ileum are inhabited by a diverse bacterial community (Zhu et al., 2002; Lu et al., 2003b). However, different from cattle and human fecal microbial communities, thus far no studies have demonstrated the value of 16S rDNA sequences to design host-specific genetic markers. Moreover, to date, there are no non-16S rDNA library-independent assays that can determine the presence of chicken fecal pollution in watersheds.

Functional genes involved in host–microbial interactions may represent a good pool of targets for host-specific assays. Some of these functional genes are hypothesized to be microbial surface proteins, while others may be associated with cellular processes and metabolism (Hooper et al., 2002; Xu et al., 2003). However, a limited number of studies have used genes involved in host–microbial interactions as potential fecal community markers (Scott et al., 2005; Shanks et al., 2006b). This is probably due to the small number of microbial genes known to be involved in host–microbial interactions and the limited sequence information for these genes.

To maximize the recovery of DNA targets potentially involved in host–microbial interactions, a recently described competitive DNA hybridization approach named genome fragment enrichment (GFE) was used by Shanks et al. (2006b) to select for cattle-specific markers using metagenomic DNA. In this study, we used GFE to identify unique fragment sequences specific to chicken feces. Two different enrichment experiments were performed using metagenomic DNA from chicken versus pig (CP) or DNA extracts from non-chicken avian species (turkeys, seagulls and Canadian geese) (CB). Sequences obtained after competitive DNA hybridization were used to design chicken fecal-specific PCR assays.

2. Materials and methods

2.1. *Sample collection and DNA extraction*

Fecal samples were collected from diverse geographic locations in the United States (Florida, New Jersey, West Virginia, Delaware, Ohio, Texas, Nebraska, and Georgia) and from one location in the Republic of China (Shandong). Samples from the following animals were used to test the host specificity and host distribution of the potential host-specific markers: *Sus scrofa* (pig), *Capra aegagrus* (domestic goat), *Ovis aries* (sheep), *Equus caballus* (horse), *Bos taurus* (bovine), *Gallus gallus* (chicken), *Meleagris gallopavo* (turkey), *Anser* sp. (Canadian goose), *Larus californicus* (seagull), *Treron* sp. (pigeon), *Canis latrans* (coyote), *Sciurus carolinensis* (gray squirrel), *Odocoileus virginianus* (whitetail deer), *Didelphis virginiana* (possum), *Canis familiaris* (dog), *Felis catus* (cat), *Lynx rufus* (bobcat), *Procyon lotor* (raccoon), *Erinaceus* sp. (hedgehog), *Loragyps atratus* (black vulture), and *Homo sapiens* (human) (Table 1). Feces were collected aseptically, placed into sterile conical tubes with screw caps, and stored at −80 °C until required. Total DNA was

Table 1 – Fecal and water samples used to test host specificity and host distribution

Test type[a]	Animal type, sampling location	Number of samples[b]
Host specificity (composite samples)	Pig, DE	10 (2)
	Cow, WV	17 (3)
	Cow, DE	11 (1)
	Human, WV	16 (3)
	Goat, DE	10 (2)
	Sheep, DE	11 (3)
	Horse, WV	5 (1)
	House cat, WV	11 (1)
	Domestic dog, WV	13 (1)
	Deer, WV	6 (1)
	Coyote, TX	10 (1)
	Squirrel, TX	4 (1)
	Possum, TX	2 (1)
	Seagull, WV	8 (1)
	Canadian goose, WV	16 (1)
	Turkey, DE	11 (1)
	Turkey, OH	6 (1)
	Pigeon, WV	2 (1)
Host specificity (individual samples)	Turkey, WV	11
	Turkey, OH	7
	Turkey, China	4
	Duck, FL	1
	Duck, GA	25

Table 1 (*continued*)

Test type[a]	Animal type, sampling location	Number of samples[b]
	Duck, China	9
	Pigeon, WV	2
	Pigeon, China	5
	Goose, GA	27
	Goose, WV	21
	Goose, NJ	2
	Seagull, FL	8
	Seagull, DE	3
	Seagull, GA	13
	Cow, WV	9
	Human, WV	13
	Pig, DE	9
	Black vulture, TX	1
	Racoon, TX	1
	Hedgehog, WV	1
	Bob cat, TX	1
Host distribution	Chicken, DE	12
	Chicken, WV	15
	Chicken, OH	9
	Chicken, GA	4
	Chicken, China	30
Fecally impacted water samples	Possibly chicken water, GA	2
	Possibly chicken water, DE	6
	Possibly goose water, NJ	9
	Possibly human water, OH	5
	Possibly cow water, NE	20
	Possibly pig water, OH	2

[a] Different tests were performed to validate the host-specific assays. DNA extracts from fecal samples were used to test host specificity using composite as well as individual fecal samples. To determine host distribution of genetic markers DNA extracts from individual chicken fecal samples were used. DNA extracts from water samples were used to determine the presence of host specific markers in fecally impacted water bodies.
[b] Numbers in parentheses represent the number of composites used in each particular case. For example, DNA extracts from two sets of five different fecal samples were combined to create the two composites tested in host-specificity studies.

extracted from the fecal samples using the Mo Bio Fecal kit (Mo Bio Laboratories Inc., CA) or the FastDNA Kit (Q-Biogene; Carlsbad, CA) following the protocols provided by the manufacturers. Total DNA was eluted in 50 or 100 μl of molecular grade water, and DNA concentrations were measured using a NanoDrop® ND-1000 UV–Vis Spectrophotometer (NanoDrop Technologies, Inc., Berlin, Germany) using 2 μl of each DNA extract.

Water samples that were possibly contaminated by chickens (Georgia), Canadian geese (New Jersey), cow (Nebraska), pig (Ohio), and human (Ohio) (Table 1) were collected in sterile bottles and transported to the laboratory in ice coolers. Samples (100 ml) were filtered onto 47 mm polycarbonate membranes (0.2 μm pore size; Millipore Corporate, MA). Membranes were then transferred into sterile conical tubes and kept at −80 °C until further processing. DNA from water samples was extracted using the FastDNA Kit (Q-Biogene; Carlsbad, CA).

2.2. *Genome fragment enrichment of chicken fecal DNA versus other fecal DNA*

A modification of the GFE method developed by Shanks et al. (2006b) was used to enrich for chicken-specific metagenomic regions (Fig. 1). Briefly, genomic DNA extracts from individual chicken fecal samples were mixed to create a fecal microbial community DNA composite. The chicken DNA composite was called "tester". A similar approach was performed for the DNA extracts of other animals that were used as "blocker". The term tester is used as the pool of metagenomic DNA from which the markers are selected, while the term blocker is used as the pool of DNA that is used to block the sequences that are in common with the tester DNA. In the original GFE protocol, DNA extracts from only one individual were used as tester and blocker. A composite DNA pool ($n = 14$ for chicken; $n = 9$ for pig) was used to better represent the diversity of host metagenomes of both tester and blocker and thus decrease the potential assay cross-reactivity. Additionally, a smaller amount of tester and blocker was used in this study and the tester–blocker ratio was 1:15. DNA purification steps were performed using isopropanol precipitation instead of ethanol precipitation to improve the recovery yields. Washing steps after the competitive DNA hybridization steps were also modified to increase stringency (details provided below).

Two independent GFE experiments were performed in this study. In the first experiment, a composite of chicken metagenomic DNA (tester) and a composite of pig metagenomic DNA (blocker) were used to enrich for chicken-specific fragments. This experiment was labeled CP. In order to selectively retrieve fragments from the chicken fecal metagenome, a subsample of the tester DNA was labeled with biotin and used as capturing surface. To prepare the DNA capturing surface, approximately 10 μg of composite fecal DNA from 14 different chickens from West Virginia (714 ng per individual fecal sample) were mechanically sheared into approximately 100–900 base pair (bp) fragments, and labeled with biotin (Sigma) as described by Shanks et al. (2006a). Ten micrograms of a DNA extract composite from nine pigs was used to prepare the blocker solution. To prepare DNA used to enrich for host-specific fragments, sequence-specific oligonucleotide primers having both a common 5′ sequence and nine random residues (herein called K9 primers; gacactctcgagacatcaccggtacc-nnnnnnnnn) were linked to 1 μg of sheared chicken fecal DNA using Klenow polymerase extension (Shanks et al., 2006a). Metagenomic fragments modified with the K9 primers are herein described as K9-tagged DNA. The second GFE experiment, labeled CB, was conducted again using chicken metagenomic DNA as the tester, while a composite of metagenomic DNA from different birds ($n = 35$: 11 turkeys, 16 Canadian geese, and 8 seagulls) was used as the blocker. The blocker fecal DNA was prepared by mixing equal amounts of DNA for each animal type to reach a total of 10 μg of composite DNA.

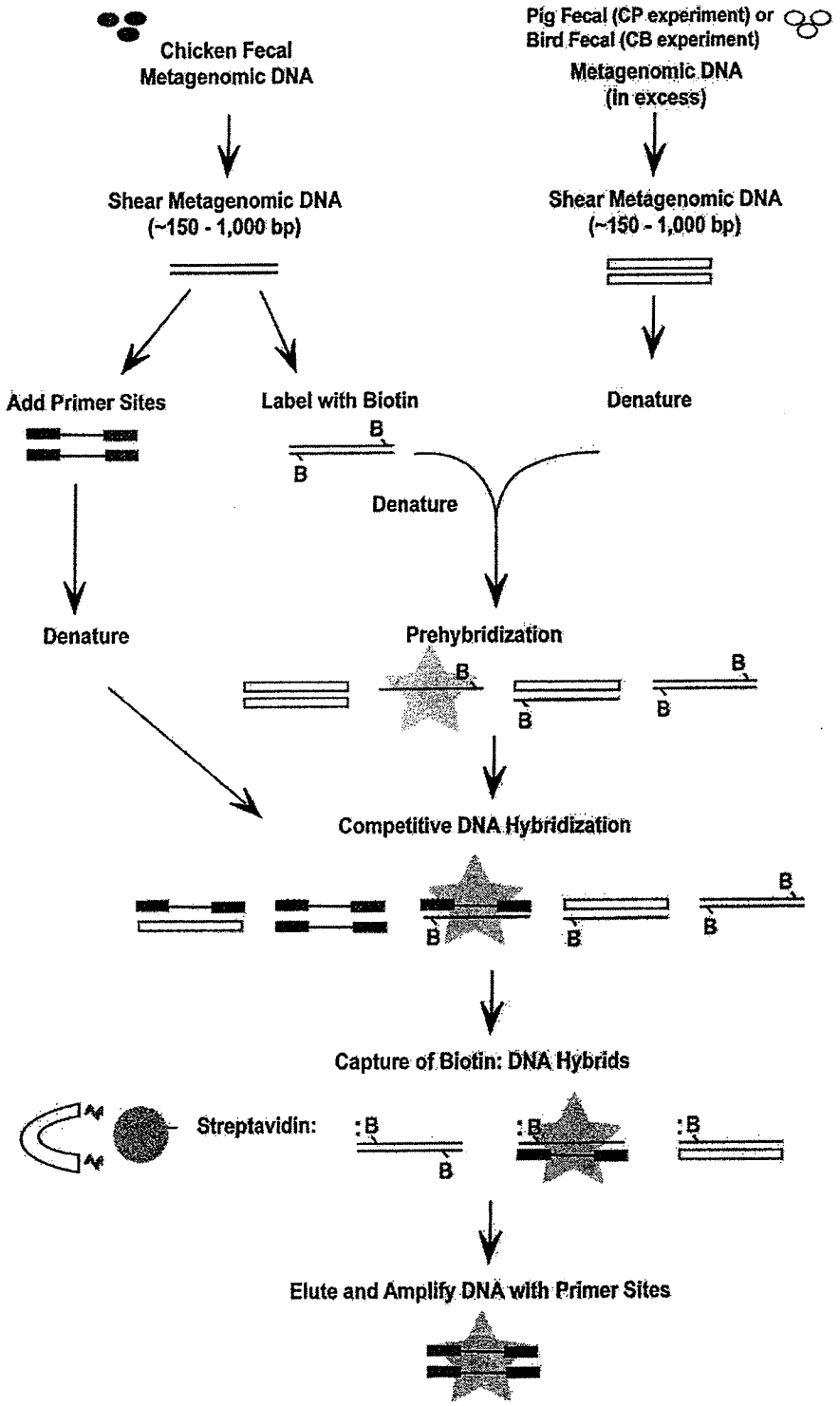

Fig. 1 – Schematic representation of the DNA enrichment method used to select for chicken and avian fecal community DNA sequences. Two experiments were performed using biotin-labeled, sheared chicken fecal metagenomic DNA (tester DNA). In one experiment, tester DNA was challenged against metagenomic DNA fragments from porcine fecal DNA extracts (CP). In a separate experiment, tester DNA was challenged against metagenomic DNA fragments from a bird composite DNA fecal extract (CB). DNA hybrids were isolated by streptavidin binding. Clone libraries were developed and randomly selected clones were sequenced to determine their potential protein function.

In both GFE experiments, the pre-hybridization solution was prepared by mixing 100 ng of the genomic DNA capturing surface and 1.5 μg of the blocker DNA solution overlaid with mineral oil. This solution was heated at 98 °C for 2 min before 4 μl of 5 M NaCl was added, and then the mixture was allowed to pre-hybridize for 20 min at 55 °C. The tagged K9-tagged

chicken genomic DNA (100 ng) was incubated separately at 98 °C for 2 min and mixed with 4 μl of 5 M NaCl before it was transferred to ice. The entire K9-tagged fecal DNA solution was added to the pre-hybridization solution, and incubated overnight at 55 °C. The conditions of the pre-hybridizations and hybridizations in the two experiments for CP and CB were the same. DNA hybrids were isolated by streptavidin binding and the captured K9-tagged genomic fragments were amplified by lone-linker PCR (Grothus, 1993). All PCR reactions were performed using a MJ Research DNA Engine Tetrad 2 thermal cycler (Bio-Rad, Hercules, CA). In each case, PCR products from the previous round were used for the next enrichment round. Each experiment was conducted in triplicate using the same preparations of capturing surface, blocking, and tester DNAs. PCR products from five reactions, for each enrichment round were pooled and cloned into pCR4.1 TOPO following the manufacturer's instructions (Invitrogen). Cloning libraries of enriched fragments were developed for rounds 1 and 2 of the CB experiment and for round 2 of the CP experiment.

2.3. *Sequencing and data analysis*

Individual clones were grown in Luria Broth plus ampicillin as the selective agent and cells were then added directly to M13-PCR assays to screen for inserts. PCR assays (25 μl) contained 1× ExTaq PCR buffer (Panvera), 2.5 mM (each) dATP, dCTP, dGTP, and dTTP, 0.2 μM of M13F and M13R primers, 0.064% bovine serum albumin (Sigma-Aldrich), 0.625 U ExTaq, and 1 μl of cells. Amplification conditions included an initial incubation at 94 °C for 3 min followed by 20 cycles of 94 °C (30 s), 52 °C (20 s), and 72 °C (40 s). Inserts were confirmed using agarose gel electrophoresis and PCR products were purified using Qiaquick 96 Plate (Qiagen). Sequencing was carried out using Big Dye terminator chemistry and capillary gel electrophoresis (Applied Biosystem PRISM 3730XL DNA Analyzer) at the Cincinnati Children's Hospital Medical Center Genomics Core Facility (Cincinnati, OH). Sequences were generated for each clone using M13 forward and reverse primers. Sequence editing and alignment were performed using Sequencher software (Gene Codes Corporation, Ann Arbor, MI). The putative protein transcript of each sequence was annotated based on the biochemical function of similar gene sequences using BLASTX with the non-redundant (NR) GenBank database (Altschul et al., 1997). BLASTX sequence matches with E values of $\leqslant 10^{-3}$ and sequence identities of $\geqslant$30% were considered to be similar protein sequences (Breitbart et al., 2003; Pearl et al., 2000). To organize sequences into functional gene categories, the DNA sequences were grouped according to the database of Clusters of Orthologous Groups (COG) of proteins ⟨http://www.ncbi.nlm.nih.gov/COG/; Tatusov et al., 1997⟩. Enriched sequences were assigned bacterial class annotations based on the top BLASTX hit (lowest E-value score) GenBank NR database.

2.4. *Primer design and PCR tests*

Primers were designed using Primer Designer software (version 2.01; Cary, NC) under the following conditions: no hairpin, no primer dimer formation, and annealing temperature of 64 or 65 °C. Assays were optimized through temperature gradients using various concentrations of fecal DNA templates. Primers were tested for host specificity against fecal DNA composites for each animal type listed above. The primers that showed host specificity to chicken composites were further challenged against fecal DNA extracts from individual cow, human, pig, turkey, Canadian goose, seagull, duck, and pigeon specimens (Table 1). Host-specific assays were used to measure host distribution of each genetic marker with individual chicken fecal samples. In addition, selected chicken fecal-specific PCR assays were challenged against DNA extracted from water samples presumed to be impacted with chicken fecal contamination. PCR assays specific to *Bacteroidetes* spp. and *Clostridium coccoides* were used to determine the presence of potential PCR inhibitors in all DNA extracts and for the potential presence of fecal pollution in water samples (Bernhard and Field, 2000a; Matsuki et al., 2002). A positive signal from the general *Bacteroides* spp. and *C. coccoides* assays in samples that tested negative for the host-specific assays was used as evidence for the absence of PCR inhibition. All tests were performed using two DNA concentrations including 1 and 10 ng μl^{-1}. For host-specificity studies, DNA for each fecal sample was first extracted and then equal amounts of each DNA extract were mixed to create fecal DNA composites. To test host distribution, individual DNA extracts from each targeted animal were tested. The presence of PCR products was visualized using 2% agarose gel electrophoresis and GelStar as the nucleic acid stain (FMC Bioproducts; Rockland, ME).

3. Results and discussion

3.1. *Analysis of GFE Sequences*

A total of 471 clones were characterized in this study, 196 from the CP experiment (chicken fecal DNA versus pig fecal DNA) and 275 from the CB experiment (chicken fecal DNA versus a composite containing turkey, seagull, and geese fecal DNA). Eighty sequences did not have significant similarity to sequences in the NR protein database. Based on top BLASTX hits, the analyzed sequences were similar to genes encoded in 19 bacterial groups (classes) and Archaea. *Clostridia*-like sequences were the most abundant group (20.9%), with many sequences showing the greatest similarity (top BLASTX hit based on E- values) to *C. perfringens*, *C. tetani*, *C. thermocellum*, and *Moorella thermoacetica* proteins (Fig. 2). *Bacilli*-like sequences were the second most abundant group (17.4%), with sequences similar to members of the genera *Enterococcus*, *Lactobacillus*, and *Streptococcus*. *Bacteroidetes*-like sequences represented 15.1% of the total clones analyzed with sequences showing similarity to *Bacteroides* spp., *Cytophaga hutchinsonii*, and *Porphyromonas gingivalis*. Other sequences showed similarity to proteins from *Actinobacteria* (8.0%) and *Cyanobacteria* (4.9%) and to α- (1.8%), β- (3.1%), γ- (9.5%), and δ-Proteobacteria (5.8%). A few sequences (1.7%) showed similarity to archaeal sequences. Interestingly, some of the sequences partially matched pathogenic bacterial genes, such as *E. coli* (93–100% identity), *Listeria monocytogenes* (84% identity), *Salmonella enterica* (96–100% identity), and *Shigella flexneri* (87–100% identity). Although the sequences in the NR

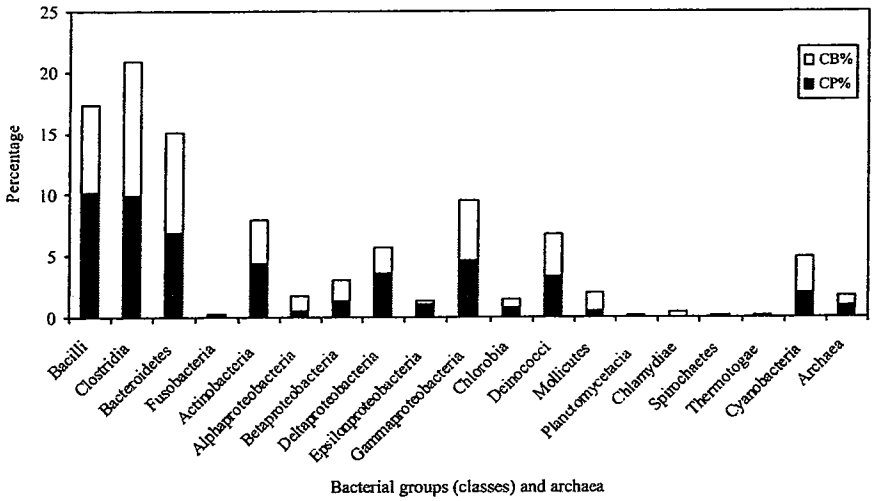

Fig. 2 – Bacterial groups associated with the identified chicken fecal metagenomic fragments (using top BLAST hit and lowest E- value; $n = 471$). CP clones are in solid bars while CB clones are represented with open bars.

database is biased toward culturable microorganisms, bacterial class designations were made to have an idea of the diversity of bacterial populations associated with the enriched DNA libraries.

The abundance of *Clostridia*- and *Bacilli* (mainly *Lactobacilladales*)-like proteins in the metagenomic libraries is not surprising in light of findings from 16S rDNA-based studies indicating that *Clostridia* and *Lactobacilladales* represent approximately 65% and 23%, respectively, of the intestinal (cecum) bacterial community of broiler chickens (Lu et al., 2003a; Zhu et al., 2002). In contrast, *Bacteroidetes* are not as numerically dominant as *Clostridia* and *Bacilli* in the chicken intestine, normally representing only 1–5% of 16S rDNA chicken clone libraries. While *Bacteroidetes* are not as abundant in chickens as in other gut systems (e.g., humans), our results further confirm that as a group *Bacteroidetes* possess a high number of bacterial host-specific genes, some of which might be involved in host–microbial interactions (Shanks et al., 2006b, Gordon et al., 2002). In addition, these results further confirm the high selection process of the GFE technique. Other gut bacteria represented in the GFE libraries and commonly identified in 16S rDNA gut libraries are *Bifidobacterium* spp., *E. coli*, *S. enterica*, and *Campylobacter* spp. (Zhu et al., 2002; Lu et al., 2003b). Some clones in the GFE library are similar to genes found in environmental bacteria like *Arthrobacter* spp., *Corynebacterium* spp., *Pseudomonas* spp., and *Geobacter* spp.. These organisms might be transitory in the gut and therefore similar sequences would not be good candidate genes for chicken-specific PCR assays.

Most fragments in which the top BLASTX sequence match showed significant similarity to a *Bacteriodetes* protein frequently showed similarity to the same protein in other *Bacteroidetes* species, suggesting that some of these genes could be playing important roles in this bacterial group. In contrast, in several cases when the sequence matched *Clostridia* or *Bacilli* proteins, other potential matches suggested a link to proteins from other organisms including *Paracoccus denitrificans*, *Corynebacterium efficiens*, *Trichodesmium erythraeum*, and *Chlamydophila pneumoniae*. Altogether, these data suggest that *Clostridia* and *Lactobacillus* genes similar to non-fecal bacterial genes might not be good candidates for the development of host-specific PCR assays.

In both CP and CB libraries, more than 60% of the bacterial clones were similar to poorly characterized genes (e.g., 63.9% were predicted as genes with unknown function). Sequences similar to *Clostridia* had the largest proportion of enriched fragments associated with uncharacterized functions (56.6%), while *Bacteroidetes* and *Bacilli* had a smaller proportion (i.e., 30.6% and 45.6%, respectively). Of the fragments associated with characterized function genes (36.1% of total analyzed sequences), 55 (11.7%) sequences are associated with cellular process, 52 (11.0%) sequences are associated with metabolic process, and 63 (13.4%) sequences are associated with information storage and processing (e.g., DNA repair and DNA replication) (Fig. 3). Sequences with high similarity to characterized genes from *Bacteroidetes*, *Clostridia*, and *Bacilli* were used for PCR assay development (Table 2).

Generally, there were no major differences in COG categories between CP and CB libraries, although cellular processes associated with cell motility and metabolic functions associated with lipid transport were present only in the CP library, while metabolic functions associated with nucleotide transport were present only in the CB library (Fig. 3). These results suggest that regardless of the type of blocking DNA used, it is possible to obtain similar COG subcategories of genes specific to the chicken microbial community by using the GFE approach. As stated above, the same gene was represented several times within a library (e.g., clone CB-R2-27) and shared between the different libraries (e.g., clone CP2-10) (Table 2). Considering the complexity of the metagenomes and the limited number of clones analyzed in this study, the probability of randomly enriching for the same gene in independent libraries is significantly low. Consequently, these results provide further evidence of the effectiveness of GFE as an approach to select for genes that are unique to the microbial community under study.

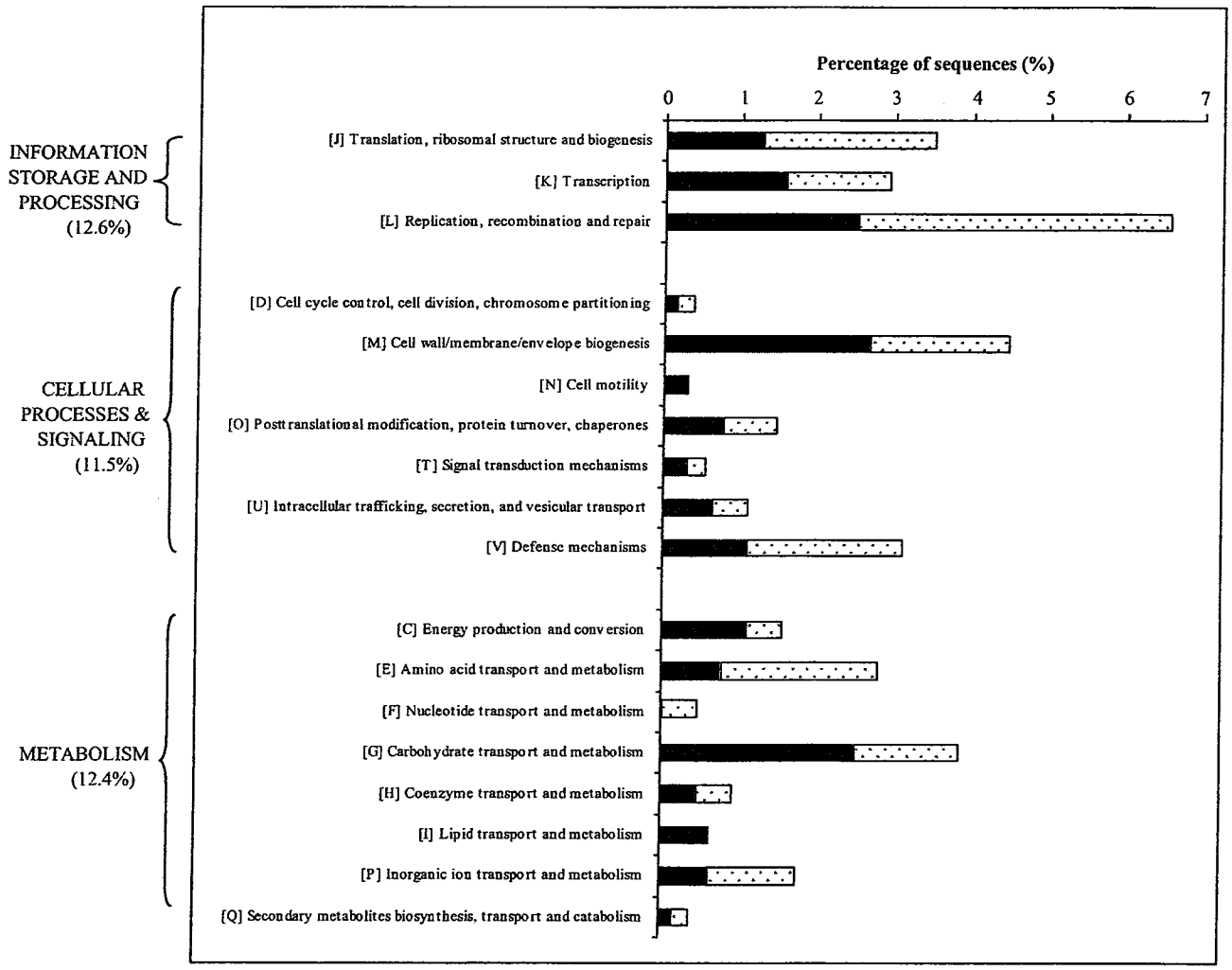

Fig. 3 – Function annotation of enriched chicken fecal DNA sequences (COG classification of top BLASTX hit; E- value $\leqslant 10^{-3}$; n = 471). CP clones are represented with open bars while CB clones are represented with solid bars, total 18%: chicken fecal DNA versus the fecal DNA of turkey, seagull, and Canadian goose. Not included in this figure are the clones associated with poorly characterized categories (31% and 32% in CP and CB clone libraries, respectively).

3.2. *Development of host-specific primers*

Twenty-five sequences were selected from the clone libraries to develop host-specific PCR assays based on the following criteria: (1) showed similarity to *Clostridia*-, *Bacteroidetes*- and *Lactobacillales*-like proteins; (2) showed similarity to characterized proteins involved in information storage, cellular processes, and metabolism; and (3) showed similarity to membrane-associated proteins.

To test host specificity, the PCR assays were challenged against composite fecal DNA extracts from non-target animals and from the chicken composite sample used in the GFE experiments. All primers tested discriminated between testers and blockers (pig in CP and turkey/goose/seagull in CB, respectively). Of the CP-based assays, ten were shown to be chicken specific, three produced amplification products for both chicken and turkey fecal DNA, four assays were positive with fecal DNA from all birds tested (chicken, turkey, seagull, and geese), and two cross-reacted with non-avian hosts (Table 2). Two assays from the CB sequences were specific to chicken fecal DNA, while the rest also cross-reacted with sheep and/or goat fecal DNA. Of the ten PCR assays targeting cellular related processes, eight of them were shown to be chicken or bird specific. These results are compatible with previous studies showing that functional gene sequences associated with cellular processes are good targets for host-specific PCR assays (Shanks et al., 2006b).

Of the seven DNA sequences possibly related to metabolism, five of them were specific to chicken fecal bacteria (Table 2). Those five DNA sequences were similar to cell membrane proteins involved in the transport of carbohydrates (e.g., CP1-26), inorganic ions (e.g., CP1-24 and CP2-9), amino acids (e.g., CP1-55 and CP3-49), and nucleotides (CP2-10). Sequences related to proteins involved in information storage/processing and poorly characterized functions (CP1-25, CP3-48, CB-R2-42, CP1-40, and CP1-74) were specific to a narrow range of hosts (i.e., chicken and/or turkey). Therefore, these results suggest that host-specific sequences can be found in several COG categories.

Table 2 – Description of primers tested for host specificity

Clone #	Fragment size/PCR product size (DNA bp)	Forward and reverse primer sequences (5′→3′)	COG category	Top BLASTX hit organism (lowest E value)	Expect values	Amino acid sequence length (% identity)	Primer specificity
CB-R2-10	326/306	CCATCCACAGCACGTCGTA AGATCTTCATCCAGTACGGCA	Cellular processes (chaperones)	*Bacteriaroides fragilis*	4E-27	108 (50)	Chicken and goat
CB-R2-27	614/607	CGAAGCGGAGAAGAACAAGA GTTCCGCAAGGTAGAGGAAA	Metabolism (inorganic ion)	*B. thetaiotaomicron*	2E-44	205 (45)	Chicken, goat, and sheep
CB-R2-28	344/327	GGCAAGCCTCAATCGCAT GTTCTGGTCGTTGGGCTGA	Cellular processes (signal transduction)	*B. fragilis*	3E-35	115 (61)	Chicken and sheep
CB-R2-34	418/261	GTCCAGGATTTCGTGGGA AAGGAGCAGCTGACGGCA	Information storage and processing	*Clostridium thermocellum*	5E-26	115 (52)	Chicken, pigeon, and sheep
CB-R2-42	627/265	GAGGAGATCTATATTTGCCTGA CGGAGGATATCCTACGATCA	General function prediction only	*Desulfitobacterium hafniense*	1E-03	93 (33)	Chicken
CB-R2-80	589/287	CGTGAATTTCCGCTACGA CCTCTTCGTTGCGTCCCA	Cellular processes (wall/membrane)	*B. fragilis*	1E-25	125 (45)	Chicken
CP1-1	623/281	GGCAGGCATCAAGTCAACA TGGCAAAAGCAACTGTCATGGCA	Cellular processes (cell division)	*C. tetani*	3E-16	99 (41)	Chicken and other birds
CP-1-10	383/350	AGGAGCATTTGTCGCCCTA GGTAAAGCTGCCCGGTAATA	Cellular processes (defense)	*B. fragilis*	9E-31	96 (88)	Chicken
CP1-24	549/379	TACCCGCAACGGGGAGAA GCGATGATACGCTTTCCGAA	Metabolism (inorganic ion)	*B. fragilis*	3E-13	138 (33)	Chicken
CP1-25	575/445	CTGGAGATCATCGTTGAGAGA TAGGCTCAAGCAGTACCGGA	Information storage and processing	*C. perfringens str.*	4E-58	165 (65)	Chicken and turkey
CP1-26	544/442	CCTGTCGTAAAACCCGGGG TGTTCGATTTTGCCTGTTTCA	Metabolism (carbohydrate)	*B. thetaiotaomicron*	3E-37	162 (44)	Chicken
CP1-40	438/244	TATTTCTGGGTGCGGTTGTA CTGACCGGAATGACTCCCA	General function prediction only	*B. thetaiotaomicron*	6E-6	114 (30)	Chicken
CP1-55	391/289	GTGCGACCGATATGGACCA GAGACATCACCGGAAACAAGA	Metabolism (amino acid transport)	*B. fragilis*	2E-17	103 (56)	Chicken
CP1-74	493/295	AGACATCACCGGGAATAACTA CAAGGAGCTATGCCGCTTA	General function prediction only	*B. fragilis*	1E-19	115 (43)	Chicken

CP2-9	251/245	GTAAGAGAGCAAGCCCATGTA ACCTATGGTTCAACACGCTTTA	Metabolism (inorganic ion)	*B. fragilis*	2E-22	83 (59)	Chicken
CP2-10	424/276	CTTTGCTGCAAGCTCCTTGA TACGGAAGCGGAGGAAAG	Metabolism (nucleotide transport)	*B. fragilis*	8E-27	91 (61)	Chicken and turkey, geese, and pigeon
CP2-17	413/377	GATCTGGGTCATTTGGATTGA GTTGAAGGCGCAACTGTAAA	Cellular processes (wall/membrane)	*Lactobacillus acidophilus*	2E-40	135 (62)	Chicken, Canadian geese, and pigeon
CP2-24	456/277	GACAGTCCTATGGATGCCCA AAAACGGCAGCGCAAAGA	Information storage and processing	*Clostridiaceae*	6E-45	111 (98)	Most domestic animals
CP2-57	514/307	CGCGTGCGTTCCCTTTA AATGGGCGCAAGCCTGA	Cellular processes	*B. fragilis*	1E-05	106 (31)	Chicken
CP2-66	487/407	ATCGGCTACGATTTCCGTTA TGTTCGTCGCATGGCTCA	Cellular processes (defense)	*B. fragilis*	4E-19	157 (31)	Chicken and turkey
CP3-1	402/332	GAACAGGGAGGCGTCTTGA GCGTGCAGGCCCAGACCCGTA	Information storage and processing	*Bifidobacterium sp.*	6E-66	133 (99)	Most domestic animals
CP3-46	587/556	GGAAATCACAGTTTTGGGGA CGCATGGAGGACGATGGTA	Cellular processes (wall/membrane)	*B. fragilis*	5E-65	196 (63)	Chicken
CP3-48	445/412	GGCTGCCTGGTCGTCTACA AGCGGCCTCTTGAGTCCA	Information storage and processing	*C. thermocellum*	6E-45	146 (56)	Chicken and turkey
CP3-49	367/329	GTCCAGCGCCTCATTGAT TGGTGATCGACTTTTCCAAT	Metabolism (amino acid transport)	*C. tetani*	5E-29	122 (53)	Chicken
Cp3-73	395/354	ACCATTTTGCTTGTCACTGCCA AATGTAAGCCGCAAAGATGA	Cellular processes (wall/membrane)	*L. gasseri*	1E-19	131 (38)	Chicken and other birds

Several enriched gene fragments have phylogenetic relevance, like CP1-25, which is similar to the elongation factor G of *Clostridium* spp. The latter gene is a homolog of elongation factor Tu, a GTP-binding protein that plays a central role in protein synthesis (Ke et al., 2000) and that has been used as a phylogenetic marker in microbial systematics (Ludwig, 1993). This is the first bacterial phylogenetic gene besides the rRNA gene to be potentially useful for the development of host-specific markers. Other phylogenetic sequences recovered in our study but not used for primer development (i.e., CB-R1-61, CB-R1-50, and Cp1-49) are related to DNA-dependent RNA polymerase β-subunit (Mollet et al., 1997; Dahllof et al., 2000) and tRNA synthetase (Woese et al., 2000). The potential host-specific nature of these phylogenetic genes was not tested. One of the sequences (i.e., CB-R2-80) was similar to the putative lipoprotein that participates in the translocation of lipoproteins from the inner membrane to the outer membrane. This protein specifically catalyzes the removal of signal peptides from prolipoproteins (Hayashi and Wu, 1990).

The sequences similar to *Bacteroidetes* proteins used for PCR assays showed mainly chicken-specific or chicken/turkey-specific signals, while the *Clostridia*- and *Lactobacillales*-based PCR assays generated positive signals with fecal DNA of other birds as well. *Bacteroidetes* 16S rDNA-based methods have been previously shown to discriminate between human and cattle fecal microbial communities (Bernhard and Field, 2000b; Field et al., 2003); however, 16S rDNA sequences have not been useful in the development of chicken-specific assays. In this study, non-ribosomal *Bacteroidetes*-like genes were identified as specific to chicken and avian microbial communities using a metagenomic approach. These results are not surprising as several studies have shown that *Bacteroides* spp. have developed a host-specific relationship with their host. For example, *Bacteroides* spp. are beneficial to the human immune system and human metabolism and can also obtain specific nutritional benefits from the host gut cells in the form of a diversity of available glycans (Bäckhed et al., 2005). Genomic and proteomic data have provided relevant insights on the nature of the interactions between this commensal bacterial group and the human and mouse gut (Hooper and Gordon, 2001). Our results suggest that some *Bacteroidetes*-like populations might also develop host-specific interactions with non-mammalian host types, although it is yet to be determined whether the interactions between *Bacteroidetes*-like and chicken gut cells are different from other gut systems. The results also suggested that *Clostridia*- and *Bacilli*-like proteins might be involved in broad symbiotic interactions as evidenced by the presence of similar host-specific genetic markers in different avian species.

The geographic distribution of all 12 chicken-specific assays was determined by first challenging each assay against fecal DNA composites of chickens from Delaware, West Virginia, Ohio, and Georgia. Differences in the geographic distribution of host-specific markers were observed for most of the PCR assays (Table 3). For example, all PCR assays produced a positive signal with WV composite chicken fecal samples, which is not surprising as WV chicken fecal DNA extracts were used as the tester pool in the GFE experiments. In contrast, several PCR assays did not amplify DNA from fecal samples collected at other sample sites (e.g., CP1-74, CP1-40, and CP3-46). The host distribution of PCR assays that produced positive signals with two or more of the chicken composite fecal DNA templates were challenged against 40 individual chicken feces from the aforementioned geographic locations and 30 individual chicken feces collected in China. Different levels of host distribution were obtained for each of the assays (Table 3). Three of the PCR assays (CP2-9, CP3-49, and CB-R2-42) amplified at least a third of the individual fecal samples tested, while the other four assays produced positive signals in approximately 6–21% of the individual fecal samples. Interestingly, nearly half of the chicken samples from China tested positive with three assays, suggesting that these markers are globally distributed and that these PCR assays may be useful for water monitoring in other countries. The PCR assays developed thus far were generated by examining a small fraction of the cloned fragments. It is reasonable to speculate that the pool of host-specific assays will increase as the number of clones examined increases. Our results indicate that metagenomic enrichments approach allow for the rapid development of multiple genetic markers to confirm the presence of chicken fecal pollution in waters.

The detection limit of the CP2-9, CP3-49, and CB-R2-42 assays was determined using fecal DNA extracts from three different locations (WV, OH, and China). The results show different levels of sensitivity for each of the assays

Table 3 – Estimated host distribution of chicken-specific PCR assays

Animal type	Sampling locations	Number of fecal samples	CP2-9[a]	CP3-49[a]	CB-R2-42	CP1-74	CB-R2-80	CP3-46	CP1-40
Chicken	DE, US	12	0	1	1	1	0	1	1
Chicken	WV, US	15	3	3	2	2	5	3	2
Chicken	OH, US	9	8	6	5	3	5	2	1
Chicken	GA, US	4	3	2	1	2	1	1	0
Chicken	Shandong, China	30	14	13	15	2	0	2	0
No. of positive signals[b]		70	28	25	24	15	11	9	4
			(40%)	(36%)	(34%)	(21%)	(16%)	(13%)	(6%)

[a] Refers to specific PCR assay.
[b] Percentage is the number of positive signals to total fecal samples tested in a certain location.

ranging from 0.001 to 1 ng μl^{-1} of fecal DNA (Table 4), with the CP2-9-based assay consistently having the highest detection limit. Each assay showed different levels of detection for different fecal samples, suggesting that the density of the populations carrying the host-specific markers can vary even among individual hosts (Table 4). In most cases, the sensitivity of the host-specific assays was lower than assays targeting *Bacteroides*- and *Clostridia*-like sequences (Table 4), suggesting that the markers are found in a subset of the most predominate fecal bacteria, a phenomenon also observed with 16S rDNA-based markers.

When the three best PCR assays (CP2-9, CP3-49, and CB-R2-42, based on the host distribution results) were challenged against water samples presumed to be contaminated with non-target sources (i.e., cattle, human, pigs, and geese), none of the DNA extracts produced PCR signals, further suggesting the host specificity of these markers. In contrast, when DNA extracts from water samples obtained from GA and DE watersheds

Table 4 – PCR assay detection sensitivity (ng DNA μl^{-1}) of chicken fecal DNA and environmental sample DNA possibly polluted by chicken manure

Sample	Source (DNA extract)	CP2-9	Cp3-49	CB-R2-42	*Clostridia*[a]	*Bacteroidetes*[a]
WV8	Chicken feces	1	1	0.01	1	1
WV9	Chicken feces	1	1	0.1	0.0001	1
WV11	Chicken feces	1	0.01	0.01	1	1
WV12	Chicken feces	0.1	1	0.01	0.0001	0.01
OH4	Chicken feces	0.01	1	1	0.01	0.01
OH5	Chicken feces	1	1	0.1	0.01	1
OH6	Chicken feces	0.1	0.01	1	0.01	0.01
OH10	Chicken feces	0.1	1	0.1	0.01	0.01
China1	Chicken feces	0.1	0.01	0.01	0.001	0.001
China3	Chicken feces	0.01	1	0.1	0.0001	0.0001
China10	Chicken feces	0.01	0.001	0.001	0.0001	0.001
China13	Chicken feces	0.01	0.001	0.001	0.0001	0.0001
DH2, GA	Water filtrate sample	0.1	–	–	0.001	0.001
DH4, GA	Water filtrate sample	1.0	–	–	0.01	0.01
Whitleyburg, DE	Water filtrate sample	–	0.1	0.1	0.01	0.01
Brownsville, DE	Water filtrate sample	–	0.1	1	0.01	0.01

[a] General PCR assays targeting members of *C. coccoides and Bacteroidetes* species.

Table 5 – Detection of chicken fecal pollution from environmental samples

Sample location	Possible fecal source contaminants	Source	No. of samples	CP2-9	CP3-49	CB-R2-42
Plum Creek, NE	Cow	Creek	20[a]	–	–	–
OH	Human	Effluents/lagoon	5[a]	–	–	–
OH	Pig	Ponds	2[a]	–	–	–
NJ	Canadian goose	Ponds	9[a]	–	–	+
GA	Chicken	Creek upstream adjacent to a farm	1	+	–	–
GA	Chicken	Creek downstream adjacent to a lagoon	1	+	–	–
Bullock Prong, DE	Chicken	River	1	+	–	–
Polly Branch, DE	Chicken	River	1	–	–	–
Bucks Branch, DE	Chicken	River	1	–	–	–
Swan Creek, DE	Chicken	River	1	–	–	–
Whitleyburg, DE	Chicken	River	1	–	+	+
Brownsville, DE	Chicken	River	1	–	+	+

[a] None of the individual samples tested showed a positive signal.

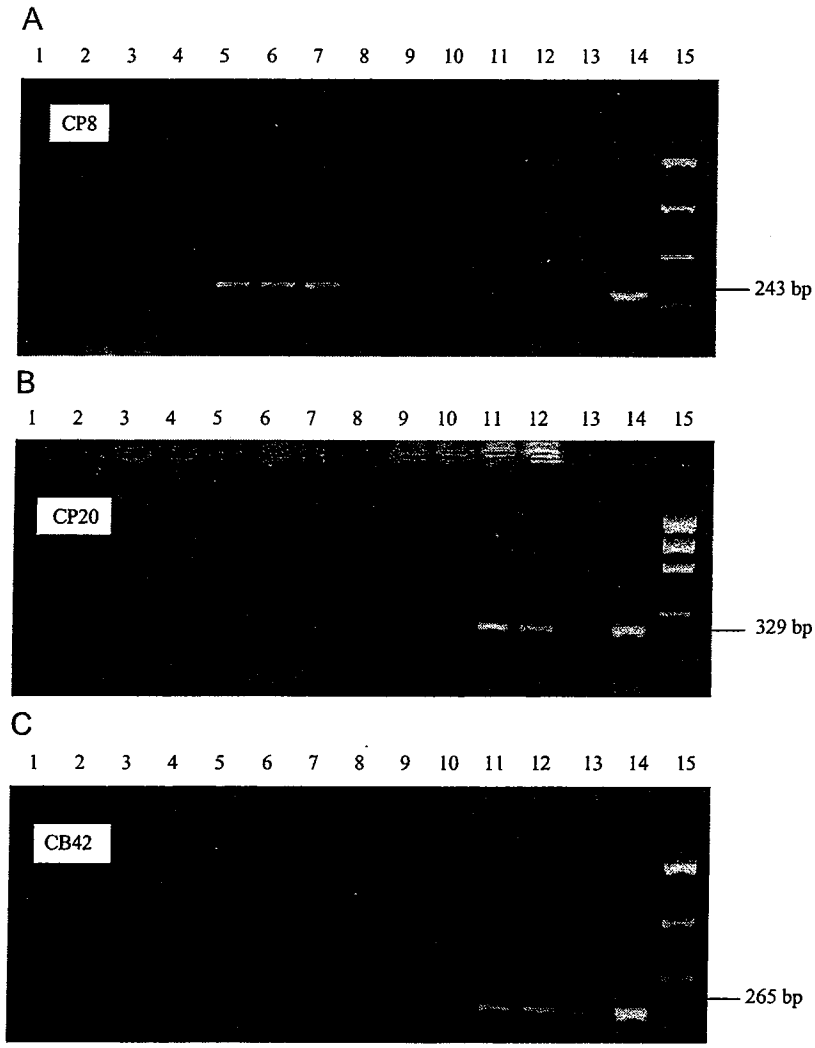

Fig. 4 – Gel electrophoresis of PCR products from reactions with chicken-specific PCR assays CP8, CP20, and CB42 (panels A, B, and C, respectively). Each PCR assay was tested against an individual DNA extract from possibly fecal-contaminated water by human (waste water, OH, lane 1), cow (creak, NE; lane 2), pigs (pond, OH; lane 3), geese (ponds, NJ; lane 4), and chickens (creek and lagoon, GA; lanes 5 and 6. river, DE; lanes 7–12). Lanes 13 and 14 are the PCR reaction controls for negative and positive.

possibly contaminated with chicken feces (Table 5; Fig. 4) were used as templates, the CP2-9 assay showed positive signals in one DE and two GA water samples, while the other two assays detected chicken-associated signals in two of the DE DNA extracts. None of the markers detected chicken contamination in two DE samples. In water samples, the minimal detection limit was 0.1 ng (Table 4). These results suggest that it might be necessary to use multiple markers when trying to detect any particular source of fecal contamination. Before using any assay in fecal source tracking studies, extensive field testing is required to determine the efficacy of the assays and the geographic distribution of the host-specific markers.

4. Conclusions

In this study, we used a competitive hybridization method to enrich for DNA metagenomic fragments specific to chicken fecal DNA. Most metagenomic clones were predicted to be similar to bacteria normally present in the chicken fecal microbial community. The cloned fragments were used to develop chicken-specific assays that were challenged against chicken and non-chicken DNA fecal extracts. Several assays were found to be chicken-specific while others produced amplification signals with a number of avian species. These results suggest that some genetic markers are conserved in bacteria present in the avian gastrointestinal tract. Each of the different assays showed different detection limits and different levels of host distribution. The detection limits were also different when the assays were challenged against water DNA extracts, suggesting that multiple markers are needed to track chicken fecal sources when using metagenomic-based assays. The approach used in this study can rapidly generate host-specific markers without the need for culturing bacteria and previous knowledge of sequence information. The assays discussed in this study represent the first chicken-specific assays available for source tracking studies.

Acknowledgments

This research was funded in part by a New Start Award from the National Center for Computational Toxicology of the US EPA, Office of Research and Development, and in part by the National Research Council. We are grateful to James Graham for early discussions on GFE and to Shichun Sun, Samuel Myoda, Katherine Painter, Marirosa Molina, Donald Stoeckel, Cathy Kelty, Margie Lee, and Regina Lamendella for providing samples. Any opinions expressed in this paper are those of the author(s) and do not, necessarily, reflect the official positions and policies of the United States Environmental Protection Agency and any mention of products or trade names does not constitute recommendation for use.

REFERENCES

Altekruse, S.F., Cohen, M.L., Swerdlow, D.L., 1997. Emerging foodborne diseases. Emerg. Infect. Dis. 3, 285–293.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., Gordon, J.I., 2005. Host-bacterial mutualism in the human intestine. Science 307, 1915–1920.

Bernhard, A.E., Field, K.G., 2000a. Identification of nonpoint sources of fecal pollution in coastal water by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl. Environ. Microbiol. 66, 1587–1594.

Bernhard, A.E., Field, K.G., 2000b. 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.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F., 2003. Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. 185, 6220–6223.

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

Dahllof, I., Baillie, H., Kjelleberg, S., 2000. *rpo* B-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. Appl. Environ. Microbiol. 66, 3376–3380.

Field, K.G., Chern, E.C., Dick, L.K., Fuhrman, J., Griffith, J., Holden, P.A., LaMontagne, M.G., Le, J., Olson, B., Simonich, M.T., 2003. A comparative study of culture-independent, library-independent genotypic methods of fecal source tracking. J. Water Health 1, 181–194.

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

Grothus, D., Cantor, C.R., Smith, C.L., 1993. PCR amplification of megabase DNA with tagged random primers (T-PCR). Nucleic Acids Res. 21, 1321–1322.

Guan, Y., Shortridge, K.F., Krauss, S., Chin, P.S., Dyrting, K.C., Ellis, T.M., Webster, R.G., Peiris, M., 2000. H9N2 influenza viruses possessing H5N1-like internal genomes continue to circulate in poultry in southeastern China. J. Virol. 74, 9372–9380.

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

Hayashi, S., Wu, H.C., 1990. Lipoproteins in bacteria. J. Bioenergy. Biomembr. 22, 451–471.

Hooper, L.V., Midtvedt, T., Gordon, J.I., 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu. Rev. Nutr. 22, 283–307.

Hooper, L.V., Gordon, J.I., 2001. Commensal host-bacterial relationships in the gut. Science 292, 1115–1118.

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

Ke, D., Boissinot, M., Huletsky, A., Picard, F.J., Frenette, J., Ouellette, M., Roy, P.H., Bergeron, M.G., 2000. Evidence for horizontal gene transfer in evolution of elongation factor Tu in enterococci. J. Bacteriol. 182, 6913–6920.

Lamendella, R., Santo Domingo, J.W., Oerther, D.B., Vogel, J.R., Stoeckel, D.M., 2007. Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of *Bacteroidetes* 16S rRNA gene. FEMS Microbiol. Ecol. 59, 651–660.

Lu, J., Sanchez, S., Hofacre, C., Maurer, J.J., Harmon, B.G., Lee, M.D., 2003a. Evaluation of broiler litter with reference to the microbial composition as assessed by using 16S rRNA and functional gene markers. Appl. Environ. Microbiol. 69, 901–908.

Lu, J., Idris, J., Harmon, B.G., Hofacre, C., Maurer, J.J., Lee, M.D., 2003b. Diversity and succession of the intestinal bacterial community of the maturing broiler chicken. Appl. Environ. Microbiol. 69, 6816–6824.

Ludwig, W., Neumaier, J., Klugbauer, N., Brockmann, E., Roller, C., Jilg, S., Reetz, K., Schachtner, I., Ludvigsen, A., Bachleitner, M., Fischer, U., Schleifer, K.H., 1993. Phylogenetic relations of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthetase beta-subunit genes. Antonie van Leeuwenhoek 64, 285–305.

Martin, S.A., McCann, M.A., 1998. Microbiological survey of Georgia poultry litter. J. Appl. Poult. Res. 7, 90–98.

Matsuki, T., Watanabe, K., Fujimoto, J., Miyamoto, Y., Takada, T., Matsumoto, K., Oyaizu, H., Tanaka, R., 2002. Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. Appl. Environ. Microbiol. 68, 5445–5451.

Mollet, C., Drancourt, M., Raoult, D., 1997. *rpo* B sequence analysis as a novel basis for bacterial identification. Mol. Microbiol. 26, 1005–1011.

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

Pearl, F., Lee, D., Bray, J., Sillitoe, I., Todd, A., Harrison, A., Thornton, J., Orengo, C., 2000. Assigning genomic sequences to CATH. Nucleic Acids Res. 28, 277–282.

Scott, T.M., Jenkins, T.M., Lukasik, J., Rose, J.B., 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.

Shanks, O.C., Santo Domingo, J., Graham, J.E., 2006a. Use of competitive DNA hybridization to identify differences in the genomes of bacteria. J. Microb. Methods 66, 321–330.

Shanks, O.C., Santo Domingo, J., Lamendella, R., Kelty, C.A., Graham, J.E., 2006b. Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cattle fecal samples. Appl. Environ. Microbiol. 72, 4054–4060.

Stuven, R., Bock, E., 2001. Nitrification and denitrification as a source for N and NO_2 production in high-strength wastewater. Water Res. 35, 1905–1914.

Tatusov, R.L., Koonin, E.V., Lipman, D.J., 1997. A genomic perspective on protein families. Science 278, 631–637.

Woese, C.R., Olsen, G.J., Ibba, M., Söll, D., 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. 64, 202–236.

Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V., Gordon, J.I., 2003. A genomic view of the human–*Bacteroides thetaiotaomicron* symbiosis. Science 299, 2074–2076.

Zhu, X.Y., Zhong, T., Panya, Y., Joerger, R.D., 2002. 16S rRNA-based analysis of microbiota from the cecum of broiler chickens. Appl. Environ. Microbiol. 68, 124–137.