

Differentiation of *Escherichia coli* Pathotypes by Oligonucleotide Spotted Array†

Raghavan U. M. Palaniappan,¹ Yu Zhang,¹ David Chiu,¹ Alfonso Torres,¹ Chobi DebRoy,² Thomas S. Whittam,³ and Yung-Fu Chang^{1*}

Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853¹; Gastroenteric Disease Center, Department of Veterinary Science, The Pennsylvania State University, University Park, Pennsylvania 16802²; and Microbial Evolution Laboratory, National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan 48824³

Received 27 July 2005/Returned for modification 28 September 2005/Accepted 11 January 2006

To accurately determine the pathotypes of *Escherichia coli* strains, a comprehensive assessment of each strain that targets multiple genes is required. A new approach to the identification and characterization of *E. coli* pathotypes was developed by constructing gene-specific probes (70-mers) for not only the virulence genes associated with each *E. coli* pathotype but also the O157-, CFT073-, and K-12-specific and common genes of each pathotype. Analysis of oligonucleotide probes with reference and clinical isolates of *E. coli* pathotypes indicated that the array could differentiate the pathotypes on the basis of their virulence and specific gene patterns. Probes targeting common genes of *E. coli* were present in all the reference and clinical strains. *Salmonella enterica* subsp. *enterica*-specific genes and *Salmonella* core genes were used as negative controls. The entire *E. coli* pathotype showed reactivity to only 4 of the 81 *Salmonella*-specific gene probes. Characterization of the genetic and virulence profiles of a single strain by using probes for virulence factors and specific and common genes in the spotted array is an ideal diagnostic tool for determination of *E. coli* pathotypes and could also have a significant impact on the epidemiological analysis of *E. coli* infections.

Escherichia coli is a normal commensal gram-negative rod-shaped bacterium that lives inside the intestinal tracts of humans and warm-blooded animals. However, some *E. coli* strains cause urinary tract infections, bacteremia, and bacterium-related diarrhea and are also the main cause of neonatal meningitis in human and animals. In 1999, the Centers for Disease Control and Prevention estimated that there were 269,060 cases of gastroenteritis caused by *E. coli* in the United States alone (15). Pathogenic *E. coli* strains can be distinguished from their nonpathogenic counterparts by the presence of virulence genes, which code for adherence and colonization, invasion, cell surface molecules, secretion, transport, and siderophore formation (9). These virulence genes are generally organized as large blocks in chromosomes, plasmids, or phages and are often transmissible between *E. coli* strains. Based on the type of virulence factor present and host clinical symptoms, *E. coli* strains are categorized into pathotypes: (i) enteropathogenic *E. coli* (EPEC), which causes diarrhea in children and animals; (ii) enterohemorrhagic *E. coli* (EHEC), which is responsible for hemorrhagic colitis and hemolytic-uremic syndrome; (iii) enterotoxigenic *E. coli* (ETEC), which causes traveler's diarrhea and porcine and bovine diarrhea; (iv) enteroaggregative *E. coli* (EAEC), which causes persistent diarrhea in humans, and diffusely adherent *E. coli* (DAEC), a subclass of enteroaggregative *E. coli* which causes diarrhea in children; (v) enteroinvasive *E. coli* (EIEC), which causes watery diarrhea and dys-

entery; (vi) uropathogenic *E. coli* (UPEC), which causes urinary tract infections in humans and animals; and (vii) neonatal meningitis *E. coli* (NMEC), which is responsible for meningitis and sepsis (12). EIEC and EAEC strains were reported to be found only in humans and not in animals. Even though *E. coli* is a significant human and animal pathogen, there is currently no rapid and efficient method for identifying the different pathotypes of *E. coli*.

In order to ensure public health and to monitor biological threats, a rapid, sensitive, and specific diagnostic assay is necessary for the identification of *E. coli* pathotypes. However, the bacterial genomes are extremely dynamic, and the ability of the organisms to acquire genetic elements, such as pathogenicity islands and virulence factors, from one another in the environment makes it difficult to identify the pathogens (16). Currently employed diagnostic assays, such as biochemical and immunological marker assays, PCR, reverse transcription-PCR, nucleic acid hybridization assays, and other bioassays are not comprehensive because they focus on the specific detection of a single target rather than multiple indicators of the pathogen. DNA microarrays provide the obvious method for exploring the genome at the molecular level. Screening of multiple markers makes it possible to determine the genetic and virulence profiles of a single strain or to distinguish one strain from others. Increasing the number of genetic regions examined will increase the confidence of correct identification and is especially important for *E. coli*, in which virulence and genetic profiles are pertinent since they may change due to lateral gene transfer.

The recently emerging DNA microarray or gene chip technology allows us to comprehensively screen thousands of genes arrayed on a single glass microscopic slide, making microarrays potentially useful for the typing of bacterial pathogens. Mi-

* Corresponding author. Mailing address: Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853. Phone: (607) 253-3675. Fax: (607) 253-3943. E-mail: yc42@cornell.edu.

† Supplemental material for this article may be found at <http://jcm.asm.org/>.

TABLE 1. Sources, categories, and genotypes of the strains used in this study^a

| <i>E. coli</i> strain | Pathotype | Source | O type | H type | Genotype or phenotype |
|-----------------------|---------------|-----------------------|--------|--------|---------------------------|
| K-12 MG1655 | Nonpathogenic | U | U | U | |
| O42 | EAEC | Human (child) | 44 | 18 | |
| J96 | UPEC | U | 4 | U | |
| CFT073 | UPEC | U | 6 | 1 | |
| E2348/69 | EPEC | Human (child) | 127 | 6 | |
| EDL933 | EHEC | Food (hamburger) | 157 | 7 | |
| Dec1A (C2816) | EPEC | Human (infant) | 55 | 6 | α-Intimin |
| Dec1B (C2814) | EPEC | Human | 55 | 6 | α-Intimin |
| 0143 (C2822) | EIEC | Cantaloupe | Neg | 45 | IpaH |
| 2-3555 (C2834) | EIEC | Standard strain (WHO) | 143 | U | IpaH |
| 1.2887 (C2828) | UPEC | Dog (lung) | 4 | Neg | CNF1 |
| 1.2922 (C2832) | UPEC | Dog | 6 | Neg | CNF1 |
| 2.2943 (C2824) | UPEC | Dog (colon) | 4 | U | CNF1 |
| 4.0532 (C2836) | ETEC | Cow (feces) | U | U | Sta |
| 4.0663 (C2838) | ETEC | Pig (small intestine) | U | U | LT Sta STb K88 |
| 4.0954 (C2842) | ETEC | Goat | U | U | Sta, STb |
| 99.0489 (C2852) | EAEC | Human | U | U | EAEC |
| 99.0490 (C2854) | EAEC | Human | U | U | EAEC |
| 99.1781 (C2856) | EHEC | Human | 157 | 7 | SLT1 SLT2 <i>eae</i> HlyA |
| 99.1782 (C2858) | EHEC | Human | 157 | 7 | SLT1 SLT2 <i>eae</i> HlyA |
| 99.1783 (C2860) | EHEC | Human | 157 | 7 | SLT1 SLT2 <i>eae</i> HlyA |

^a Abbreviations: U, unknown; Neg, negative.

croarrays have been used for the differentiation of bacterial and viral pathogens and the identification of virulence factors (2–4, 10, 18, 20–23, 25, 26). However, a drawback with the current research is that the typing of bacterial species by the use of DNA microarrays is based purely on a few virulence genes, some of which have been shown in many studies to be shared between many pathotypes and cannot be conclusive determinants for the differentiation of pathotypes.

In this study, we used a new approach by taking advantage of the genomic sequences of *E. coli* and have developed an oligonucleotide spotted array (70-mers) representing the known *E. coli* pathotype virulence genes (those of EHEC, EPEC, UPEC, ETEC, EAEC, and EIEC), specific genes (those of *E. coli* O157 EDL933, *E. coli* K-12 MG1655, and *E. coli* CFT073), common genes (those of *E. coli* O157 EDL933, *E. coli* K-12 MG1655, and *E. coli* CFT073), and negative controls (core genes of *Salmonella* and dimethyl sulfoxide buffer without oligonucleotides). Standardization of the DNA microarray was done with reference strains of *E. coli*, and then the validity of the array was assessed with known clinical pathotypes of *E. coli*. The pathotype category, the virulence profile, and its relation with other categories were determined; the results indicated that the oligonucleotide DNA microarray can be widely applicable for clinical diagnosis and epidemiological surveys.

MATERIALS AND METHODS

Bacterial strains and growth conditions. All the reference strains and field isolates of *E. coli* pathotypes were collected in our laboratory. The sources, categories, and phenotype patterns of the *E. coli* pathotypes are listed in Table 1. All these strains were streaked onto Luria agar plates, and a single colony was selected and propagated in Luria broth at 37°C for 12 h. Genomic DNA was isolated by using a DNeasy kit (QIAGEN, Valencia, Calif.).

Construction of virulence, common, and specific gene probes. The DNA sequences of the virulence factor genes from each *E. coli* pathotype were obtained from the NCBI database, and the probes (70-mers) were designed by using Arrayologoselector (<http://arrayologoselector.sourceforge.net/>). The program op-

timizes the oligonucleotide selection based upon several parameters, including uniqueness in the genome, sequence complexity, lack of self-binding, G+C content, and proximity to the 3' end of the gene. The virulence gene probes that were designed were subjected to a search of the NCBI database with the BLAST program to confirm their uniqueness to a particular pathotype. The gene name, gene accession number, source, probe sequence, and its homology to other organisms are listed in the supplemental material (Appendix S1). The pathotype-specific genes were identified based on the annotation and comparison of the genomic sequences of strains EDL933, K-12, and CFT073 (24). We randomly selected 150 pathotype-specific genes for each pathotype and designed oligonucleotides (70-mers). These oligonucleotides that were designed were again subjected to a search of the NCBI database with the BLAST program by use of the genome sequences of EDL933, K-12, and CFT073 prior to synthesis of the probes (data not shown). Of these, 60, 40, and 61 gene probes of EDL933, K-12, and CFT073, respectively, showed no significant homology to other organisms in the NCBI database. A list of *E. coli* common genes based on *E. coli* O157 EDL933, *E. coli* K-12 MG1655, and *E. coli* CFT073 were obtained as described previously (24). Of these, 48 housekeeping genes were randomly selected, and probes were designed. Core genes of *Salmonella* were also included as negative controls for the microarray analysis (14). A list of the 443 gene probes obtained from *E. coli* and *Salmonella* is provided in the supplemental material (Appendix S1).

Microarray printing. The probes (70-mers) were synthesized (Illumina Inc.), suspended in 50% dimethyl sulfoxide, and spotted in triplicate onto Ultra-GAPS glass slides (Corning Inc., Corning, N.Y.) at the Cornell Microarray Core Facility (www.bigredspots.cornell.edu). Autoblank was also used as a negative control.

DNA preparation and labeling. The genomic DNA of the *E. coli* pathotypes was prepared according to the manufacturer's instructions by using DNeasy kits (QIAGEN). The harvested genomic DNA was digested with Sau3AI (New England Biolab, Beverly, Mass.) and was purified by using a QIAquick PCR purification kit (QIAGEN). The purified fragments were labeled according to the protocol of P. Brown (http://cmgm.stanford.edu/pbrown/protocols/4_genomic.html). The purified fragments were mixed with 15 µg of random hexamers (Amersham, Piscataway, N.J.), boiled for 5 min, and immediately cooled on ice. Deoxynucleoside triphosphates (6 nmol each of dATP, dGTP, and dTTP and 3 nmol of dCTP [Amersham]), 10 U of Klenow enzyme (New England Biolabs), and 3 nmol of Cy3-dCTP (Amersham) were added, and the mixture was incubated for 2 h at 37°C. The labeled probes were purified and concentrated with Microcon YM-30 (Millipore) to a volume of 12 µl or less. To the concentrated probe, 1 µl of 10 mg/ml salmon sperm DNA, 1 µl of 4 mg/ml yeast tRNA, 3.5 µl of 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), and 1.2 µl of 5% sodium dodecyl sulfate (SDS) were added, and the mixture was made up to 20 µl with

water. The hybridization mixture was boiled for 2 min, and the denatured probe was kept at 37°C for 30 min prior to hybridization.

Microarray experiments and data analysis. The microarray chip contained oligonucleotides (70-mers) representing open reading frames (ORFs) of the virulence genes from each *E. coli* pathotype (EHEC, EPEC, UPEC, ETEC, EAEC, and EIEC), specific genes (those of *E. coli* O157 EDL933, *E. coli* K-12 MG1655, and *E. coli* CFT073), common genes (those of *E. coli* O157 EDL933, *E. coli* K-12 MG1655, and *E. coli* CFT073), and negative controls (core genes of *Salmonella* and autoblanks). Each slide had triplicate spots of each ORF. Immediately before use, the slide was incubated in prehybridization solution (5× SSC, 0.1% SDS) at 55°C for 1 h and was then rinsed with 1× SSC, 0.2× SSC, and finally, Milli-Q water. The slides were dried either by blowing nitrogen gas on them or by centrifugation (600 rpm for 5 min). Denatured probe was slowly applied to the slide through the edges of a LifterSlip coverslip (Erie Scientific, Portsmouth, N.H.), the slide was immediately placed in a hybridization chamber (Corning, Elmira, N.Y.), and hybridization was allowed to occur overnight by submerging the slide in a 55°C water bath. The slides were successively washed with the following three solutions: solution I, 2× SSC–0.1% SDS at 60°C for 30 min; solution II, 1× SSC at 37°C for 10 min; solution III, 0.2× SSC at 37°C for 10 min; and solution IV, Milli-Q water at 37°C for 5 s. Finally, the slides were dried by centrifugation or with nitrogen gas, as described above, and scanned on a GenePix 4000A scanner (Axon Instruments, Union City, Calif.).

Each microarray experiment with reference strains was repeated five times, and the experiments with the clinical isolates of the *E. coli* pathotypes were repeated twice. Since the array was spotted in triplicate, a single hybridization yielded three sets of data for each gene. Fluorescence data for triplicate spots on one slide were collected separately by the use of Genepix pro 6.0 software. The data were then analyzed by using Avadis 3.3 prophetic software (<http://avadis.strandgenomics.com/>). The pixel intensity value of each spot was corrected for the background and averaged for triplicate spots. The log value for the average pixel intensity was taken and was subjected to descriptive statistical analysis. The data were filtered in such a way that genes which showed a mean value of more than two times the standard deviation ($P < 0.05$) were considered positive. Hierarchical clustering was performed by using the Pearson absolute method as the distance matrix.

RESULTS

Analysis of virulence, pathotype-specific, and common gene probes with reference strains of *E. coli* pathotypes. In order to assess the virulence, pathotype-specific, and common gene probes in the microarray, reference strains of *E. coli* pathotypes, which included *E. coli* strains with known genomic sequence (nonpathogenic strain K-12 MG1655, EHEC O157 EDL933, and UPEC CFT073) and well-characterized strains (EAEC O42 and EPEC E2348/69), were subjected to microarray analysis. As indicated in the Table 2, all the reference strains were almost 100% positive to the common gene probes of the *E. coli* pathotypes.

E. coli K-12 did not show any signal for the virulence genes of *E. coli* pathotypes except *afaD* and *ibeB* of UPEC and meningitis-causing *E. coli*, respectively. K-12 also showed no signals for *Salmonella* core genes or CFT073-specific genes except *eco-306*, *yijF*, and *eco-347*. However, K-12 reacted with 100% of its specific genes, as listed in Table 2.

EHEC EDL933 showed 100% reactivity to all of its virulence and specific genes. Although the EPEC E2348/69 genome has not been completely sequenced, some of its virulence phenotypes are well characterized. Typical EPEC strains are known to have a locus of enterocyte effacement (LEE) region and an EAF plasmid. The presence of *tir* and genes encoding the type III secretion system suggest the presence of LEE, whereas the presence of *bfp* and *per* indicates the presence of the EAF plasmid. EPEC also shares a common set of virulence factors with EHEC, which includes the genes of the LEE region; these genes were designated EHEC-EPEC viru-

TABLE 2. Percent reactivity (positive) of gene-specific probes to reference and clinical isolates of *E. coli* pathotypes^a

| Gene ^b | % Reactivity ^c | | | | | | | | | | | | | | | | | | | | | |
|--------------------------|---------------------------|-----------|------------|------------|------------|------------|------------|------------|--------------|----------|------------|------------|------------|------------|------------|-------------|------------|------------|-----------|------------|------------|-----|
| | K12 | EHEC EDL* | EHEC C2856 | EHEC C2858 | EHEC C2860 | ETEC C2836 | ETEC C2842 | ETEC C2838 | UPEC CFT073* | UPEC J96 | UPEC C2824 | UPEC C2828 | UPEC C2832 | EIEC C2822 | EIEC C2834 | EPEC E2348* | EPEC C2816 | EPEC C2814 | EAEC O42* | EAEC C2852 | EAEC C2854 | |
| K-12 specific (40) | 100 | 7.5 | 18 | 7.5 | 10 | 4.5 | 4.3 | 4.0 | 4.0 | 4.0 | 4.0 | 3.8 | 13 | 4.0 | 2.5 | 3.3 | 3.5 | 3.8 | 3.3 | 4.0 | 3.3 | 3.3 |
| EDL933 specific (60) | 0 | 100 | 97 | 77 | 98 | 15 | 12 | 12 | 8.3 | 8.3 | 13 | 15 | 12 | 22 | 2.5 | 4.3 | 4.3 | 3.2 | 1.2 | 1.3 | 1.7 | 1.7 |
| CFT073 specific (61) | 4.9 | 1.6 | 21 | 13 | 18 | 4.6 | 3.9 | 2.0 | 100 | 90 | 9.8 | 9.7 | 8.9 | 3.3 | 1.3 | 2.5 | 2.6 | 3.1 | 3.4 | 1.8 | 5.9 | |
| E2348/69 virulence (36) | 3 | 1.2 | 1.2 | 1.2 | 9.1 | 3.9 | 3.0 | 1.2 | 100 | 81 | 8.6 | 8.6 | 7.0 | 1.2 | 9.1 | 6.1 | 3 | 6.1 | 6.1 | 6.1 | 6.1 | |
| E2348/69 virulence (17) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| EAEC virulence (11) | 0 | 9.1 | 9.1 | 9.1 | 9.1 | 9.1 | 9.1 | 9.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9.1 | 9.1 | 27 | 36 | |
| EIEC virulence (13) | 0 | 0 | 7.7 | 7.7 | 7.7 | 7.7 | 7.7 | 0 | 0 | 3.1 | 0 | 0 | 0 | 100 | 100 | 0 | 0 | 0 | 7.7 | 0 | 7.7 | |
| ETEC virulence (18) | 0 | 0 | 0 | 0 | 0 | 11 | 1.1 | 2.8 | 0 | 0 | 0 | 0 | 0 | 5.6 | 0 | 0 | 0 | 0 | 0 | 0 | 5.6 | |
| EHEC-EPEC virulence (23) | 0 | 7.4 | 7.0 | 5.2 | 6.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4.4 | 8.7 | 100 | 9.6 | 100 | 0 | 0 | 4.4 | |
| Common (48) | 96 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 98 | 100 | 98 | 100 | 98 | 100 | 100 | 100 | 100 | 100 | 100 |
| EDL933 virulence (15) | 0 | 100 | 100 | 100 | 100 | 6.7 | 6.7 | 6.7 | 0 | 0 | 0 | 0 | 0 | 0 | 2.0 | 6.7 | 6.7 | 6.7 | 0 | 0 | 0 | 0 |

^a The sources and categories of the strains are indicated in Table 1. Gene probes that show no signals for other UPEC virulence ($n = 15$), MNEC virulence ($n = 4$), and *Salmonella* core ($n = 81$) gene probes to reference strains are not included in this table. The overall positive criterion for the identification of *E. coli* pathotypes except ETEC and EAEC is 70 to 100% reactivity. ETEC and EAEC virulence genes are heterogeneous, and the representative strains for these categories have not been sequenced. The description of gene is represented in the supplemental material (Appendixes S1 and S2).

^b The number of genes is indicated in parentheses.

^c Asterisks indicate the reference strain of each pathotype.

lence genes, although the probe sequences were derived from EPEC E2348/69. On the basis of these criteria, EPEC E2348/69 showed 100% reactivity to its specific genes and the EHEC-EPEC virulence gene probes, whereas EHEC EDL933 showed only 74% reactivity to the EHEC-EPEC virulence gene probes. The nonreactive probes of EDL933 include *orf19*, *tir*, *eae*, *espA*, *espB*, and *espD*. These genes are already known to be variable on the basis of their sequences in EDL933. Since the gene probes constructed were based on the sequence of E2348/69, EDL933 was not reactive. However, the probes for the *tir-2*, *espA2*, and *espB2* genes, which were designed on the basis of the EDL933 genome sequence, were reactive only to EDL933 and not E2348/69, indicating the uniqueness of the probes used.

CFT073, the prototype strain of UPEC, showed 100% reactivity to CFT073 virulence and specific genes.

EAEC is characterized to have the following virulence genes: *aaf*, *agg*, *astA*, and *pet* (12). Our array detected *astA*, *aafA*, *aafB*, *aggR*, and *pet* by use of the reference strain of EAEC (O42). It also showed reactivity to all the EAEC virulence genes except *aggA* (91%). The *aggA*-specific gene probe was derived from EAEC isolate (O3:H2) and not EAEC O42, which might be the reason for the lack of reactivity.

Salmonella core genes were considered the negative control. These genes were derived by comparison of the genomes of different isolates of *Salmonella* and were absent from *E. coli* K-12 and *E. coli* O157:H7. As expected, 100% reactivity of the *Salmonella* core gene probes with isolates of *Salmonella enterica* serovar Typhimurium was observed (data not shown). *E. coli* pathotypes showed reactivity to 4 of 81 oligonucleotide probes specific for *Salmonella*. This might be because the *Salmonella* core genes established by McClelland et al. were compared with only two *E. coli* genomes (14).

The autoblank, which consisted of spots of dimethyl sulfoxide without any probes, was created as another negative control to make sure that nonspecific hybridization to the slide was not occurring. None of the autoblanks showed a fluorescent signal.

The specificity of the array for the differentiation of reference *E. coli* pathotypes was 98%. This specificity was estimated on the basis of the numbers of true-negative spots (autoblanks) and false-negative spots (*afaD*, *ibeB*).

Validation of virulence, pathotype-specific, and common gene probes with clinical isolates of *E. coli* pathotypes. All three clinical isolates of EHEC showed 100% reactivity to EHEC-specific virulence gene probes. With regard to the EDL933-specific gene probes, two clinical isolates (C2856 and C2860) showed reactivities of 96.5 and 98%, respectively, and one clinical isolate (C2860) showed only 77% reactivity. Similar to reference strain EDL933, clinical strains C2856, C2858, and C2860 showed less reactivity to EHEC-EPEC-specific gene probes (69.5, 52, and 65%, respectively), indicating that the genomes of clinical strains are more closely related to the EDL933 genome. All the clinical strains of EHEC showed lesser reactivities to K-12-specific genes and CFT073-specific genes (Table 2).

EPEC clinical isolate C2816 was 82 and 96% positive for the EPEC and EPEC-EHEC virulence genes probes, respectively, while EPEC clinical isolate C2814 was 88 and 100% positive,

respectively, indicating that these two strains belong to the EPEC pathotype.

J96, another prototype strain of UPEC, showed 81 and 90% reactivities to CFT073 virulence genes and specific genes, respectively. Probes for the *papA*, *sfaA*, *fsoE*, *papGII*, *sat*, and *iucD* genes failed to react with the J96 strain. Probes for CFT073-specific genes *kpsE*, *kpsD*, and *kpsM* (K15 capsule) and probes for two hypothetical genes also showed no reactivity to strain J96. The K15 capsule locus of strain J96, however, has not been characterized so far.

UPEC clinical strains C2824, C2828, and C2832 showed 86, 86, and 68% reactivities to reference strain CFT073-specific virulence gene probes, respectively, and 98, 97, and 89% reactivities to CFT073-specific genes, respectively. It is well known that UPEC strains exhibit greater diversity in their virulence genes, but screening with typical UPEC CFT073-specific genes indicates that these clinical isolates belong to the UPEC pathotype with a divergence in their virulence genes.

EAEC contains heterogeneous virulence markers. In comparison with the reference strain (O42), both EAEC clinical strains showed reactivity to *aggR*, but only one of them was positive for *astA* and one of them showed reactivity to *aggD*, *aggB*, and *aggA*. The clinical isolates of EAEC showed 32 to 40% and 12 to 16% reactivities to the K-12- and EDL933-specific genes, respectively.

The clinical isolates with a known phenotype of EIEC and ETEC showed 100% and 28% reactivities to the virulence gene probes for these pathotypes, respectively, in the microarray analysis. Since we used gene probes for all the toxins and various colonization factors synthesized by different ETEC isolates, the overall percentage of reactivity to ETEC virulence genes is low. However, ETEC isolates showed reactivity to their respective genotypes (toxins) in the array. ETEC clinical isolates showed marked differences in reactivity with ETEC virulence genes, like the EAEC isolates did (Table 2), indicating that these categories contain heterogeneous virulence markers.

The reactivities of the individual gene probes for each pathotype are represented in the supplemental material (Appendix S2). All the clinical isolates clustered with their respective reference *E. coli* pathotypes, indicating the potential for the use of microarray analysis for differentiation of the pathotypes. The diversity of pathotypes and their genetic relatedness are illustrated in Fig. 1.

DISCUSSION

Numerous bioassays targeting *E. coli* virulence genes have been developed for the detection and the typing of pathogenic *E. coli*. The sheer number of virulence genes and the fact that the virulence genes in the pathogenicity islands could potentially be transferred to other bacteria in the environment (16) make it hard to differentiate between different *E. coli* pathotypes. Using data from the completed genomic sequences of EHEC *E. coli* O157 EDL933, nonpathogenic *E. coli* K-12 MG1655, and UPEC (*E. coli* CFT073) and also virulence factors from *E. coli* pathotypes, we have developed a microarray containing not only virulence gene probes associated with each *E. coli* pathotype but also probes for pathotype-specific genes developed for nonpathogenic K-12, EHEC EDL933, and

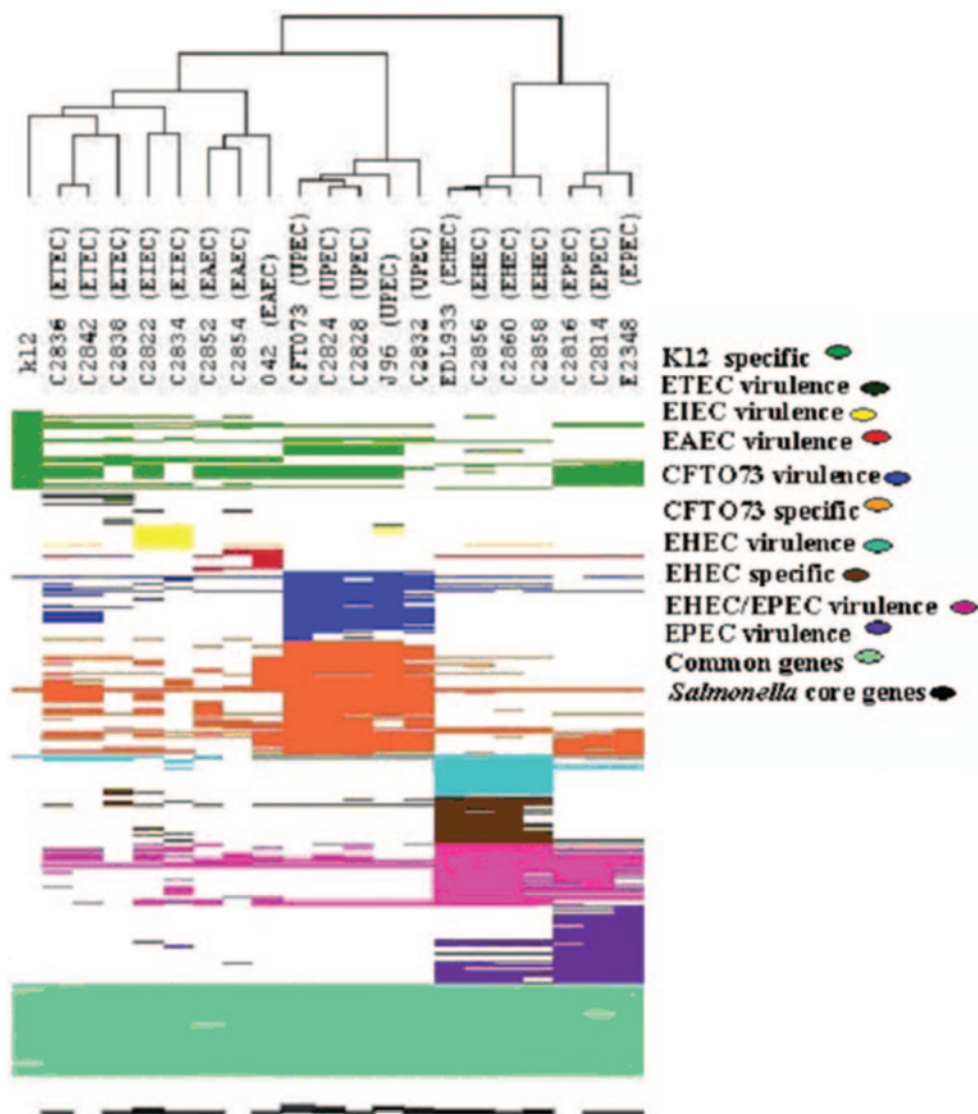


FIG. 1. Dendrogram representation of virulence, specific, and common gene patterns for *E. coli* pathotypes. Hierarchical clustering was obtained by using Avadis software. The pathotype virulence and specific genes are indicated with different colors to denote the specificity of pathotypes in the dendrogram. Fifteen EHEC, 17 EPEC, 36 UPEC, and 11 EAEC virulence gene probes were derived from strains EDL933, E2348/69, CFT073, and O42, respectively. Twenty-three EHEC-EPEC virulence gene probes were obtained from EPEC E2348/69. Sixty EHEC-specific genes, 61 UPEC-specific genes, and 40 nonpathogenic *E. coli*-specific genes were from strains EDL933, CFT073, and K-12, respectively. Gene probes that show no signals for 24 UPEC-specific gene probes and also 81 *Salmonella* core genes (negative control) were also included in the dendrogram. The results for four nonreactive virulence gene probes of meningitis-causing *E. coli* are not included in the dendrogram. The diversity between EHEC and EPEC in certain regions of EHEC-EPEC is also represented as a pattern profile in the dendrogram. All the reference and clinical isolates except the ETEC and EAEC isolates contained conserved virulence and specific genes; the ETEC and EAEC isolates contained heterogeneous virulence genes. However, the uniqueness of ETEC and EAEC toxins and virulence genes is evident in the pattern profile of the dendrogram.

UPEC CFT073. *E. coli* common genes and *Salmonella* core genes were used as positive and negative controls, respectively.

Recently, microarrays targeting multiple virulence factors have been developed for the detection of pathogenic *E. coli* (1). However, the microarray analysis was evaluated as positive or negative by the color of the fluorescence intensity. Each spot in the array is subjected to variability in intensity, and the evaluation of the array by the color of the fluorescence intensity is difficult, especially when the fluorescence intensity is compared with the background, which may eventually lead to

artifacts. We have calculated the background locally for each spot rather than globally for the entire image, which might have enhanced the quality of each spot. The microarray data were analyzed by taking the log value of the median without the background, and the data were further filtered statistically by taking two times the standard deviation ($P < 0.05$) of the median value. Pathogenic *E. coli* strains associated with human and animal diseases are remarkably diverse and show variability in the genes encoding surface antigens or virulence factors (5). These variable genes may not bind or may bind poorly to

the probes in the array. Seventy-mer oligonucleotide probes rather than cDNA PCR probes were used in this multiple-target assay, since oligonucleotide probes are a cost-effective alternative to cDNA PCR probes. Furthermore, the oligonucleotide-based arrays provide a reduction in cross-hybridization and an increase in the differentiation of highly homologous regions (6, 11, 23).

The microarray's ability to detect virulence, pathotype-specific, and common genes was primarily analyzed with reference strains of *E. coli* pathotypes, such as nonpathogenic K-12 MG1655, EHEC O157 EDL933, EAEC O42, EPEC E2348/69, and UPEC CFT073. To validate the array further, clinical isolates representing each *E. coli* pathotype were used. The results indicate that microarray analysis could potentially differentiate *E. coli* pathotypes not only on the basis of virulence genes but also on the basis of specific and common genes. Even the potentially powerful microarray showed false-positive results for *afaD*, *ibeB*, and *traT*. Bekal et al. developed a microarray targeting virulence factors for the detection of pathogenic *E. coli*, and that microarray also showed false-positive results for *aagA* and *cdt* (1).

Comparison of the LEE regions of EPEC E2348/69 and EHEC *E. coli* O157 showed that *tir*, *eae*, *espA*, *espB*, and *espD* are more diverse (7, 13). The diversity is evident in our study with EHEC reference and clinical isolates. UPEC isolates also contained heterogeneous virulence genes. UPEC prototype strain J96 showed reactivity to *papC*, *papE*, *papG1*, *papGII*, *papX*, *papK*, *papJ*, *papH*, and *papI* but not to *papA* or *papA2*. None of the clinical isolates showed reactivity to *papA* or *papA2*, supporting the *papA* subunit diversity in UPEC isolates (8, 19). EIEC isolates also contained different virulence genes (1, 17), but the clinical isolates of EIEC evaluated in this study had similar virulence profiles.

E. coli pathotypes are categorized by the presence of virulence factors. The virulence factors are acquired from numerous sources, including bacteriophages, plasmids, and the genomes of other bacteria (5). It is apparent that the virulence genes in the pathogenic island are transferable and could contribute to the heterogeneity of *E. coli* strains (5). Therefore, it is ideal to track the pathotypes on the basis of the virulence genes. However, it would be better to include nontransferable and specific genes of *E. coli* pathotypes, as the categories of pathotypes keep on increasing. Screening with multiple markers such as virulence, specific, and core genes of *E. coli* help us to identify the emergence of the new pathotypes and would also allow us to assess the relative genetic and virulence profiles of a single strain in comparison with particular pathotypes.

An unknown pathotype can be identified primarily by its percentage of reactivity to the virulence genes of EHEC, UPEC, EPEC, EAEC, EIEC, and ETEC. The pathotype-specific genes constructed on the basis of the available genome sequences of strains EDL933, CFT073, and K-12 enhance the accuracy of identification of EHEC and UPEC. However, ETEC and EAEC are not well characterized compared to the other pathotypes, and their virulence genes are heterogeneous among isolates. Therefore, the reactivities of certain genes, especially with toxins and fimbrial types (ETEC *sta*, *stb*, and *lt*; EAEC *astA* and *aagR*) can be considered important criteria that can be used to denote the pathotype of an isolate. Although the clinical isolates of the *E. coli* pathotypes showed

gene contents similar to those of reference strains, variations in gene content between pathotypes and between each strain were also observed by DNA microarray analysis. This indicates that the oligonucleotide array not only allows us to differentiate between pathotypes but also is useful for rapid strain characterization. The ability to characterize genome deletions of clinical isolates relative to the sequences of reference strains may allow us to reconstruct the phylogeny. The microarray that was developed could differentiate the major pathotypes, but the analysis of reference and clinical isolates of EHEC 2, EPEC 2, and DAEC strains may enhance its validity for wider application for the diagnosis of *E. coli* infections. Since the virulence genes are heterogeneous within each category, especially for ETEC, EAEC, and UPEC, the establishment of core genes of each pathotype would help us to identify the pathotype. With the completion of more *E. coli* genomic sequences, the inclusion of more specific genes from each pathotype may enhance the identification of pathogenic *E. coli* strains and their pathotypes.

ACKNOWLEDGMENT

This project was supported with federal funds from the National Institute of Allergy and Infectious Diseases, National Institute of Health, U.S. Department of Health and Human Services, under contract N01-AI-30054, project ZC002-03.

REFERENCES

- Bekal, S., R. Brousseau, L. Masson, G. Prefontaine, J. Fairbrother, and J. Harel. 2003. Rapid identification of *Escherichia coli* pathotypes by virulence gene detection with DNA microarrays. *J. Clin. Microbiol.* **41**:2113–2125.
- Call, D. R., M. K. Bakko, M. J. Krug, and M. C. Roberts. 2003. Identifying antimicrobial resistance genes with DNA microarrays. *Antimicrob. Agents Chemother.* **47**:3290–3295.
- Call, D. R., F. J. Brockman, and D. P. Chandler. 2001. Detecting and genotyping *Escherichia coli* O157:H7 using multiplexed PCR and nucleic acid microarrays. *Int. J. Food Microbiol.* **67**:71–80.
- Cho, J. C., and J. M. Tiedje. 2001. Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl. Environ. Microbiol.* **67**:3677–3682.
- Donnenberg, M. S., and T. S. Whittam. 2001. Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. *J. Clin. Invest.* **107**:539–548.
- Dorrell, N., S. J. Hinchliffe, and B. W. Wren. 2005. Comparative phylogenomics of pathogenic bacteria by microarray analysis. *Curr. Opin. Microbiol.* **8**:620–626.
- Elliott, S. J., L. A. Wainwright, T. K. McDaniel, K. G. Jarvis, Y. K. Deng, L. C. Lai, B. P. McNamara, M. S. Donnenberg, and J. B. Kaper. 1998. The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol. Microbiol.* **28**:1–4.
- Feria, C., J. Machado, J. D. Correia, J. Goncalves, and W. Gastra. 2001. Virulence genes and P fimbriae PapA subunit diversity in canine and feline uropathogenic *Escherichia coli*. *Vet. Microbiol.* **82**:81–89.
- Finlay, B. B., and S. Falkow. 1997. Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* **61**:136–169.
- Fukushima, M., K. Kakinuma, H. Hayashi, H. Nagai, K. Ito, and R. Kawaguchi. 2003. Detection and identification of *Mycobacterium* species isolates by DNA microarray. *J. Clin. Microbiol.* **41**:2605–2615.
- Hessner, M. J., V. K. Singh, X. Wang, S. Khan, M. R. Tschannen, and T. C. Zahrt. 2004. Utilization of a labeled tracking oligonucleotide for visualization and quality control of spotted 70-mer arrays. *BMC Genomics* **5**:12.
- Kaper, J. B., J. P. Nataro, and H. L. Mobley. 2004. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2**:123–140.
- Kresse, A. U., M. Rohde, and C. A. Guzman. 1999. The EspD protein of enterohemorrhagic *Escherichia coli* is required for the formation of bacterial surface appendages and is incorporated in the cytoplasmic membranes of target cells. *Infect. Immun.* **67**:4834–4842.
- McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**:852–856.
- Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro,

- P. M. Griffin, and R. V. Tauxe.** 1999. Food-related illness and death in the United States. *Emerg. Infect. Dis.* **5**:607–625.
16. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299–304.
17. **Smajs, D., and G. M. Weinstock.** 2001. The iron- and temperature-regulated *cjrBC* genes of *Shigella* and enteroinvasive *Escherichia coli* strains code for colicin Js uptake. *J. Bacteriol.* **183**:3958–3966.
18. **van Ijperen, C., P. Kuhnert, J. Frey, and J. P. Clewley.** 2002. Virulence typing of *Escherichia coli* using microarrays. *Mol. Cell. Probes* **16**:371–378.
19. **Verdonck, F., E. Cox, and B. M. Goddeeris.** 2004. F4 fimbriae expressed by porcine enterotoxigenic *Escherichia coli*, an example of an eccentric fimbrial system? *J. Mol. Microbiol. Biotechnol.* **7**:155–169.
20. **Volokhov, D., V. Chizhikov, K. Chumakov, and A. Rasooly.** 2003. Microarray-based identification of thermophilic *Campylobacter jejuni*, *C. coli*, *C. lari*, and *C. upsaliensis*. *J. Clin. Microbiol.* **41**:4071–4080.
21. **Volokhov, D., A. Rasooly, K. Chumakov, and V. Chizhikov.** 2002. Identification of *Listeria* species by microarray-based assay. *J. Clin. Microbiol.* **40**:4720–4728.
22. **Wang, D., L. Coscoy, M. Zylberberg, P. C. Avila, H. A. Boushey, D. Ganem, and J. L. DeRisi.** 2002. Microarray-based detection and genotyping of viral pathogens. *Proc. Natl. Acad. Sci. USA* **99**:15687–15692.
23. **Wang, H. Y., R. L. Malek, A. E. Kwitek, A. S. Greene, T. V. Luu, B. Behbahani, B. Frank, J. Quackenbush, and N. H. Lee.** 2003. Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. *Genome Biol.* **4**:R5.
24. **Welch, R. A., V. Burland, G. Plunkett III, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley, M. S. Donnenberg, and F. R. Blattner.** 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **99**:17020–17024.
25. **Wu, C. F., J. J. Valdes, W. E. Bentley, and J. W. Sekowski.** 2003. DNA microarray for discrimination between pathogenic O157:H7 EDL933 and non-pathogenic *Escherichia coli* strains. *Biosens. Bioelectron.* **19**:1–8.
26. **Wu, L., D. K. Thompson, G. Li, R. A. Hurt, J. M. Tiedje, and J. Zhou.** 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl. Environ. Microbiol.* **67**:5780–5790.