Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle

Jaehyoung Kim, Joseph Nietfeldt, and Andrew K. Benson*

Department of Food Science and Technology, University of Nebraska, 330 Food Industry Complex, Lincoln, NE, 68583-0919

Communicated by Allan Campbell, Stanford University, Stanford, CA, August 26, 1999 (received for review May 5, 1999)

Multilocus-genotyping methods have shown that *Escherichia coli* O157:H7 is a geographically disseminated clone. However, high-resolution methods such as pulse-field gel electrophoresis demonstrate significant genomic diversity among different isolates. To assess the genetic relationship of human and bovine isolates of *E. coli* O157:H7 in detail, we have developed an octamer-based genome-scanning methodology, which compares the distance between over-represented, strand-biased octamers that occur in the genome. Comparison of octamer-based genome-scanning products derived from >1 megabase of the genome demonstrated the existence of two distinct lineages of *E. coli* O157:H7 that are disseminated within the United States. Human and bovine isolates are nonrandomly distributed among the lineages, suggesting that one of these lineages may be less virulent for humans or may not be efficiently transmitted to humans from bovine sources. Restriction fragment length polymorphism analysis with lambdoid phage genomes indicates that phage-mediated events are associated with divergence of the lineages, thereby providing one explanation for the degree of diversity that is observed among *E. coli* O157:H7 by other molecular-fingerprinting methods.

The enterohemorragic *Escherichia coli* (EHEC) have emerged as a leading cause of bloody diarrhea (hemorragic colitis) in the United States (1). The O157:H7 serotype is the predominant serotype of EHEC isolated from patients in the United States (2); however, other EHEC serotypes have been also identified and are more prevalent among hemorragic colitis patients in other countries (3, 4). Of the EHEC serotypes studied thus far, each shares a common set of genes that have a demonstrated or implicated role in pathogenesis. These include the *stx1* and *stx2* genes encoding the shiga toxins (5, 6), several genes located within the locus of enterocyte effacement (7–9), and the *ehxA* gene encoding a hemolysin (10).

Population genetic analyses of EHEC strains and shiga toxin-producing *E. coli* by multi-locus enzyme electrophoresis indicate that extant EHEC strains are derived from two distinct lineages (11). The EHEC 1 lineage consists solely of a few closely related and geographically disseminated multi-locus genotypes each bearing the O157:H7 serotype. In contrast, the EHEC 2 lineage is serotypically more diverse, displays a much higher degree of genomic diversity, and is only distantly related to the EHEC 1 lineage (12, 13). It was therefore hypothesized that the EHEC lineages diverged from a common ancestor bearing the locus of enterocyte effacement pathogenicity island, and they were independently lysogenized by Stx phage (12, 14).

Despite the multi-locus similarity of *E. coli* O157:H7 strains, considerable genome diversity has been observed by high-resolution methods such as pulse-field gel electrophoresis (PFGE) (15–18). The detection of restriction fragment length polymorphisms (RFLPs) by hybridization with phage λ (19, 20) suggests that movement of prophage into or out of the genome may be partly responsible. Diversity has also been proposed to occur in the O157:H7 genome through transient appearance of mutator strains, which enhance the rate at which foreign DNA can be recombined into the genome (21). Experimentally, however, it is unclear what accounts for the degree of diversity observed by PFGE.

To assess the genetic relatedness of *E. coli* O157:H7 isolates in detail, we used compositional bias of the *E. coli* chromosome to develop a method termed octamer-based genome scanning (OBGS). OBGS is based on PCR amplification of genomic segments that lie between over-represented, strand-biased octamers in the genome. In the *E. coli* K-12 genome, there are 150 different over-represented oligomers whose occurrence is skewed to one strand of the genome (23). Of these, 23 are octamers that occur from 515 to 867 times (24). Each of the octamers is biased to the leading strand of each replichore, and collectively their complements are believed to provide priming sites for discontinuous DNA replication (23, 24). To perform OBGS, fluorescent PCR primers based upon octamers biased to the leading strand are combined with unlabeled octamers biased to the lagging strand (complements of leading strand octamers). The size distribution of fluorescently labeled OBGS products is then measured by fragment analysis on automated sequencers followed by creation of binary files from the presence and absence of bands (Fig. 1). Because the distribution of OBGS products is measured on automated DNA sequencers, products that differ in length by as little as a single nucleotide between samples can be resolved. Thus, the occurrence of large-scale genome alterations as well as single nucleotide insertions or deletions can be detected among segments amplified from different strains.

In this paper, we show that OBGS analysis distinguishes two lineages of *E. coli* O157:H7. Human and bovine isolates are nonrandomly distributed among the lineages. RFLP analysis with lambdoid phage probes further demonstrate that phage-mediated events are associated with divergence.

Materials and Methods

Bacterial Strains. *E. coli* O157:H7 strains were obtained from Charles Kaspar (Food Research Institute, University of Wisconsin, Madison, WI), Thomas Cebula (U.S. Food and Drug Administration, Washington, DC), and Paul Fey (University of Nebraska Medical Center, Omaha, NE). Human isolates (FRIK523–FRIK856) and dairy cattle isolates (FRIK920–FRIK1641) from Wisconsin were previously described (18, 25). Human isolates from locations across the United States were derived from documented sporadic cases of hemorragic colitis in Nebraska (NE007–112), California (FDA513, FDA518, and ATCC43890), Michigan (FDA516), North Carolina

Abbreviations: EHEC, enterohemorragic *E. coli*; OBGS, octamer-based genome scanning; RFLP, restriction fragment length polymorphism; NJ, neighbor joining; PFGE, pulse-field gel electrophoresis; ECOR, *E. coli* Reference.

Data deposition: These sequences reported in this paper have been deposited in the GenBank database (accession nos. AF177208, AF177209, and AF177210).

*To whom reprint requests should be addressed: E-mail: abenson1@unl.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

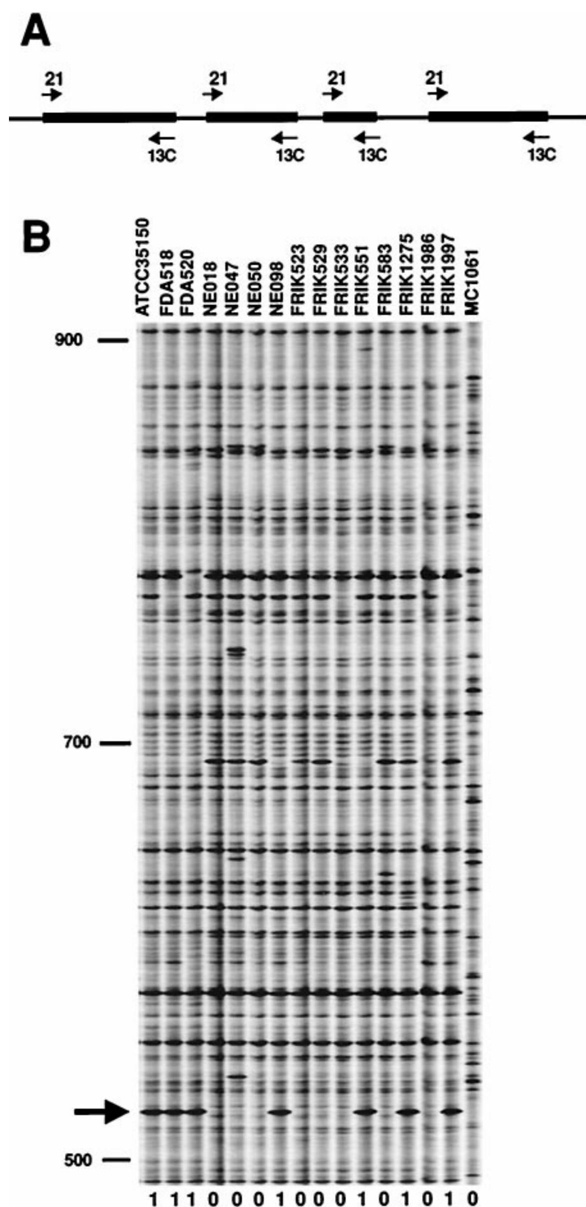

Fig. 1. OBGS strategy. (*A*) The occurrence of a set of leading strand-biased (OCT21) and lagging strand-biased (OCT13C) octamers, indicated by arrows, on a contiguous segment of the chromosome. The black rectangles between the arrows illustrate the segments of the genome that would be amplified by OBGS reactions with the OCT21 and OCT13C primer pair at primer saturation. (*B*) A section of the image file from OBGS reaction products of several *E. coli* O157:H7 strains and the K-12 strain MC1061 is shown after data collection. Molecular sizes are indicated in base length on the left. The image was compacted 5× vertically. The large arrow indicates an OBGS reaction product whose binary character scores for the different strains is indicated immediately below the image (1 = presence; 0 = absence).

(ATCC43889), Washington state (FDA517), Alaska (FDA519), and Oregon (FDA520) and outbreaks in Washington state (86–24, 93–001 through 93–066, and 95–003) and Michigan (ATCC483894 and ATCC43895). Dairy cattle isolates from geographically unlinked locations (FRIK1985–FRIK2004) were obtained from a collection of isolates generated during the U.S. Department of Agriculture, National Animal Health Monitoring System survey (26). The *E. coli* Reference (ECOR) collection (27) was obtained from Thomas Whittam (Pennsylvania State University, State College, PA). Strain MC1061 (a K-12 derivative) was obtained from A. Newton (Princeton University, Princeton, NJ).

Primer Design. Primers for OBGS were based on frequently occurring octamers in the *E. coli* K-12 genome (24). For OBGS reactions, one primer in each combination was labeled with a fluorophore (IRD800) that emits in the near-infrared region of the spectrum (Li-Cor, Lincoln, NE). Each OBGS primer combination consisted of a fluorescent primer based on an octamer biased to the leading strand and an unlabeled primer based on a second octamer biased to the lagging strand (complement of a leading strand octamer). Because of the effect of 5′-fluorochromes on annealing of the octamers, an AT dinucleotide was added to the 5′-end of each primer. The fluorescently labeled primers used in this study are K12OCT3B (ATGCTGGTGG), K12OCT4B (ATGCTGGCGG), K12OCT7 (ATTGGCGCTG), K12OCT14 (ATGCTGGCGA), K12OCT21 (ATGCGCTGGA), and K12OCT22 (ATCTGGCGAA). Unlabeled primers were purchased from Genosys (Woodlands, TX). Unlabeled primers used in this study are K12OCT1C (ATCGCCAGCG), K12OCT4C (ATCCGCCAGC), K12OCT6C (ATGCCAGCGC), K12OCT8C (ATGCGCCAGC), K12OCT13C (ATTGCGCCAG), and K12OCT19C (ATCTTCCAGC).

OBGS. OBGS reactions were conducted in a volume of 20 μl. Each reaction contained 50 ng of chromosomal DNA, 0.5 pmol of each primer, 750 μM dNTPs, 50 mM Tris (pH 9.3), 4 mM $MgCl_2$, and 0.5 units of Sequitherm (Epicentre, Madison, WI). The reactions were denatured for 2 min at 95°C followed by 30 cycles of 10 sec at 25°C, ramp to 40°C at 1°C/4 sec, 15 sec at 40°C, 30 sec at 72°C, and 15 sec at 95°C. After completion of the cycling, an equal volume of loading dye (0.012% bromphenol blue/0.1 mM EDTA, pH 8.0, in 100% formamide) was added. A portion (1 μl) of each reaction was then loaded onto 4% denaturing polyacrylamide gels (66-cm gel length) in a Li-Cor 4000L automated sequencer. Fragment length was calculated from molecular weight markers that were run alongside the reactions by using RFLPSCAN software (Scanalytics, Fairfax, VA). OBGS profiles were stable on repeat testing after three successive passages.

Data Analysis. Binary files were created in Microsoft EXCEL 97 from printed copies of the Images produced by an Alden 9315CTP photographic quality thermal printer (Alden Electronics, Westborough, MA). The files were generated from the presence/absence of bands (Fig. 1) between 200 and 1,500 bases in length and binary files from each primer combination were combined head to tail in Microsoft WORD 97. Phylogenetic relationships based on the OBGS characters were assessed through maximum parsimony methods in PAUP 4.0 (28) and by Neighbor Joining (NJ) analysis (29), both of which yielded similar results. The *E. coli* K-12 strain MC1061, which is phylogenetically distant from *E. coli* O157:H7, was included in each analysis as an outgroup to assess the ancestral state of the characters as described (13).

Isolation of Bacteriophage HB4. Bacteriophages were isolated from culture supernatants of the human isolate FRIK551 ($stx1^+$, $stx2^+$) after incubation for 24 hr at 37°C. Supernatants were filtered through a 0.45-μm filter and spotted onto lawns of *E. coli* HB101. Plaques were then excised and incubated in Luria broth for 30 min, and dilutions were plated onto lawns of HB101. Eleven different plaques that displayed unique sizes and turbidity were chosen and plaque purified on lawns of HB101. Genomic DNA from the phage isolates was prepared as described (30). Restriction analysis with *Bam*HI indicated that all but one band were conserved among the isolates. One isolate, designated HB4, was chosen as a representative. DNA sequence

GENETICS

Table 1. Estimates of chromosome coverage from OBGS reactions

Primer combination†	No. of bands‡	Length, bases§
OCT3B*–OCT4C	186	165,155
OCT4B*–OCT19C	245	227,178
OCT7*–OCT8C	235	215,356
OCT14*–OCT1C	167	174,872
OCT21*–OCT13C	189	161,173
OCT22*–OCT6C	165	154,268
Total	1,187	1,098,002

†Labeled primers are marked with an asterisk.
‡Total number of bands resolved between 200 and 1,500 bases in length from *E. coli* O157:H7 strain FRIK 1641.
§Combined sizes, in bases, of the bands resolved between 200 and 1,500 bases in length.

analyses of three cloned segments from HB4 were conducted for further characterization of the phage (accession nos. AF177208–AF177210).

RFLP analysis was conducted on genomic DNA of *E. coli* O157:H7 isolates digested individually with *Eco*RV, *Bam*HI, and *Hin*dIII. Hybridization probes were prepared from total HB4 genomic DNA by random priming. Hybridization was performed at 65°C for 16 hr in 0.5 M sodium phosphate (pH 7.4), 1% BSA, 7% SDS, and 1 mM EDTA. The blots were washed in 40 mM sodium phosphate at 65°C and subjected to autoradiography. From three different digests, 19 informative RFLPs were observed. These were converted to binary characters and the taxa were subjected to analysis by the Unweighted Pair Group Method with Arithmetic means (UPGMA) algorithm in PAUP 4.0 (28). Based on the 19 polymorphic characters, six prominent clades were observed and denoted subtypes A–F.

Results

OBGS Parameters. From 506 possible combinations of the 23 most over-represented octamers and their complements (23 leading strand octamers $\times$ 23 lagging strand complements $-$ 23 complementary pairs), 70 combinations were arbitrarily chosen and tested on a set of two *E. coli* O157:H7 strains (FRIK1641 and FRIK533) and strains from the ECOR collection (ECOR20 and ECOR50).

Within the 200- to 1,500-base window, 18 combinations produced >200 bands, 43 produced between 150 and 200 bands, and 9 produced <150 bands. Six of 61 combinations that generate 150–250 bands were chosen for further characterization based on maximum discrimination of the *E. coli* O157:H7 strains from the ECOR strains. To estimate the degree of genome coverage, the combined length of amplification products from these six combinations was measured within the 200- to 1,500-bp window. The combined length of amplicons from *E. coli* O157:H7 strain FRIK1641 is shown in Table 1. The six primer combinations provided a range of coverage from 154,268 bases (K12OCT3B–K12OCT4C) to 227,178 bases (K12OCT4B–K12OCT19C), and together the OBGS amplicons from all six primer combinations provided an estimated 1,098,002 bases of comparative information. This estimate does not account for redundancy resulting from overlapping OBGS products in regions of closely spaced octamers. Among the products from the two O157:H7 strains, 61 polymorphisms were observed, indicating that OBGS can detect significant diversity between isolates of the O157:H7 clone.

Application of OBGS Analysis to *E. coli* O157:H7 Isolates from a Limited Geographic Region. Epidemiological studies of *E. coli* O157:H7 in humans and cattle have generated large sets of strains with corresponding biochemical and genetic profiles. Except for instances in which human isolates have been epidemiologically linked to isolates from food, little is known about the detailed genetic relationship of these isolates. We therefore initiated our studies on a characterized strain set derived from dairy cattle and humans within a three-county region of Wisconsin. Human isolates in this set were derived from sporadic cases (FRIK523–FRIK579) and an outbreak of hemorragic colitis at a daycare center (FRIK583–FRIK856) that occurred during 1994 (25). Cattle isolates (FRIK920–FRIK1641) were derived from a 1995–1996 longitudinal study of three dairy cattle farms in this same region (18).

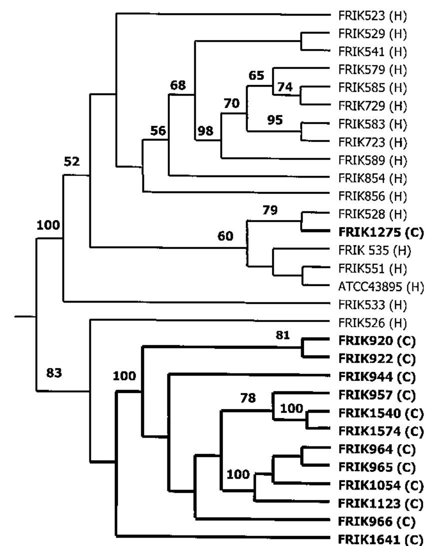

Fig. 2. Genetic relationships of human and cattle isolates of *E. coli* O157:H7 from a three-county region in Wisconsin. The dendrogram was produced by NJ analysis of binary files representing presence and absence of 1,251 OBGS segments from each strain with six different OBGS-primer combinations (tree length = 960, consistency index = 0.74, retention index = 0.83 with 168 characters). The tree is rooted with the K-12 derivative MC1061 as an outgroup. Among the O157:H7 strains, 1,060 of 1,251 characters were conserved. Human isolates (H) and cattle (C) isolates were previously described (18, 25). Bootstrap percentages, derived from 1,000 replicates of a NJ search, are indicated at nodes of clusters scoring >50%.

OBGS analysis was performed on each isolate by using the six different primer combinations characterized above. The O157:H7 strain ATCC43895 (EDL933) was included as a standard O157:H7 strain, and the K-12 derivative MC1061 was included as an outgroup. Of 1,251 OBGS reaction products scored from this strain set, 191 segments were variable among the O157:H7 strains, indicating that within this limited geographic region, a considerable amount of genomic diversity can be observed by either PFGE (18, 25) or by OBGS. Of these variable segments, 140 were parsimony-informative.

A dendrogram obtained by NJ analysis (Fig. 2) demonstrated that the strains constituted a monophyletic lineage that has diverged into two distinct populations, one containing primarily human isolates and the other containing primarily cattle isolates. These results suggest that a substantial portion of the *E. coli* O157:H7 strains isolated from cattle may comprise a genetically distinct population that is not commonly isolated from human cases of hemorragic colitis. However, because the cattle herds are regionally confined and human isolates could have originated from food sources outside of this region, it remained possible that the limited geography of strain sampling influenced the distribution.

The Two Subpopulations of *E. coli* O157:H7 Are Geographically Distributed. We next tested human and bovine isolates from a broader geographic region to determine whether geography

biased our data. We reasoned that if the subpopulation of bovine isolates from Wisconsin were unique to that region, then these strains would remain genetically unrelated to cattle or human isolates from other states. Alternatively, if the subpopulation were unique to cattle, we would anticipate that cattle isolates from geographically unlinked locations would be genetically related to the Wisconsin cattle strains. In the case of the latter, we would also anticipate that most human isolates would be distinct from this population.

Additional cattle isolates for these experiments were obtained from a collection of isolates from a 1991–1992 prevalence study of *E. coli* O157:H7 in dairy cattle herds across the nation (26). The isolates were derived from 16 different states, including Minnesota, Idaho, Maryland, New York, Oregon, Tennessee, New Hampshire, Colorado, Wisconsin, Washington, Ohio, Nebraska, Florida, Vermont, Georgia, and California. Additional human isolates, collected from outbreaks and sporadic cases of hemorragic colitis cases in Alaska, Oregon, California, Montana, Michigan, Nebraska, North Carolina, Washington, and Wisconsin during the years 1982–1998 were also included. By using the same primer combinations described above, 1,250 OBGS products from this strain set were evaluated (Fig. 3). In this geographically and temporally dispersed strain set, only 892 OBGS products were conserved among the O157:H7 strains. Of the 358 variable segments, 223 were parsimony-informative. Thus, the genomic diversity of the strains increased with geographic and temporal distribution as would be expected; however, the relationship of the strains was more informative.

Consistent with our findings with the Wisconsin set, NJ analysis of this geographically and temporally dispersed strain set (Fig. 3) demonstrated the existence of two distinct lineages (I and II). Bovine isolates from geographically unlinked locations were highly related to the large set of bovine isolates from Wisconsin and constituted the majority of strains in lineage II, indicating that this clade is not regionally specific. The absence of human isolates within the clade suggests that this population is not readily transmitted to humans or is not readily capable of causing disease. In fact, the scant representation of human isolates among lineage II suggests that the entire lineage may be impaired in transmissibility to or virulence in humans.

In contrast to lineage II strains, most bovine isolates in lineage I were dispersed among the lineage and related to human isolates, as would be expected because contaminated beef is one source for transmission of *E. coli* O157:H7 to humans. Assuming the bovine isolates from the National Animal Health Monitoring Systems survey (FRIK1985–FRIK2004) are representative of distribution of the two lineages in the United States, our results predict that only one-third (7/20) of bovine isolates belong to the apparently more virulent lineage I.

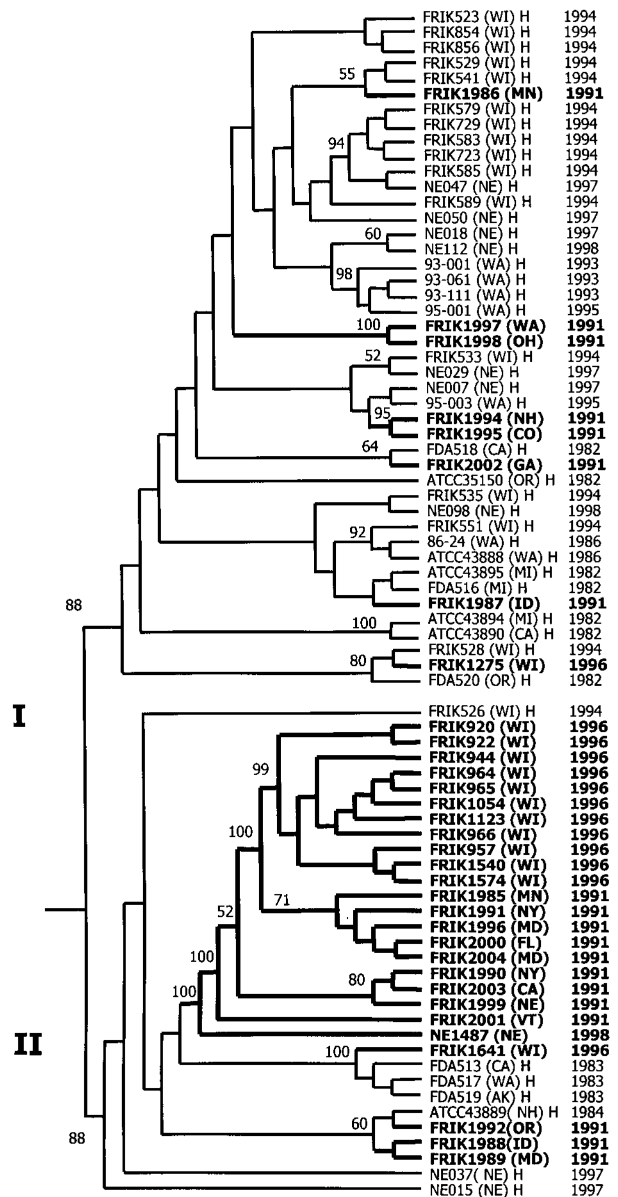

Fig. 3. Genetic relationships of cattle and human isolates of *E. coli* strains from across the United States. The dendrogram was produced by NJ analysis of binary files representing the presence and absence of 1,250 OBGS products from each strain derived from six different OBGS-primer combinations (tree length = 1398, consistency index = 0.61, retention index = 0.81 with 235 characters). Among the O157:H7 strains, 892 of 1,250 bands were conserved. The tree is rooted using the K-12 derivative MC1061 as an outgroup. The state from which each strain originated is indicated by the two-letter abbreviation in parentheses. Human isolates have an H at the end of the designation and bovine isolates are in bold. The year of isolation is indicated to the right. Bootstrap percentages, derived from 1,000 replications of a NJ search, are indicated at nodes scoring >50%.

Lambdoid Prophage and Divergence of the *E. coli* O157:H7 Lineages. Among the OBGS products from each octamer combination, we noted the occurrence of at least three polymorphic OBGS products that were exclusive to most isolates of lineage I or II. The frequency in which such dimorphic segments of the genome would be detected by OBGS is determined by dispersion of the polymorphisms relative to the nearest octamers and/or by the likelihood of polymorphisms resulting from insertions that are large enough to contain octamers. The latter provides a much simpler explanation for divergence because movement of large octamer-containing elements such as prophage or pathogenicity islands, would require fewer events to generate the two subpopulations than acquisition of multiple polymorphisms.

To test this hypothesis, we isolated lambdoid phages from supernatants of *E. coli* O157:H7 cultures. DNA sequence analyses of three cloned segments from one such phage (phage HB4) showed strong similarity (98% identity) to the *recT* gene that is part of the cryptic lambdoid phage Rac of K-12 (31) and 80% identity to the λ *J* gene encoding a component of the tail. The HB4 phage genome was then used as a probe in RFLP analyses of a subset of *E. coli* O157:H7 strains from each population. Phylogenetic analysis of RFLPs generated from different restriction digests grouped the strains into six different subtypes (A–F), which were then superimposed onto an NJ tree derived from OBGS characters (Fig. 4).

The confinement of subtypes A, B, and C to lineage I and of subtypes D, E, and F to lineage II demonstrates that phage-

GENETICS

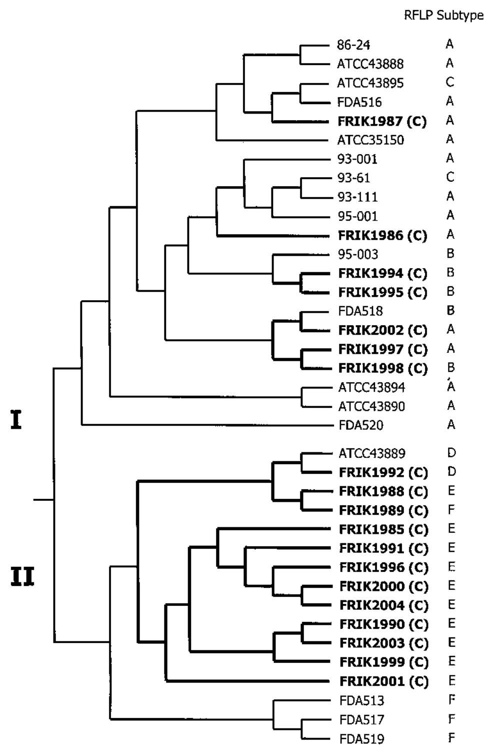

Fig. 4. Phylogenetic relationships of *E. coli* O157:H7 strains and the concordance of lambdoid phage-mediated polymorphisms. The dendrogram was generated by NJ analysis of binary files representing the presence/absence of 1,250 OBGS reaction products from six different OBGS combinations (tree length = 1,080, consistency index = 0.74, retention index = 0.79 with 187 characters). Strains isolated from cattle are in bold and have a (C) at the end of the designation. Of the O157:H7 strains, 953 characters were constant. The letters in the column HB4 RFLP represent subtypes obtained by RFLP analysis by using HB4-genomic DNA as a probe. The six different HB4 RFLP subtypes were distinguished by Unweighted Pair Group Method with Arithmetic means (UPGMA) analysis of 19 variable characters generated from three different restriction digests of genomic DNA from the strain set.

associated polymorphisms segregate precisely along lineages predicted by OBGS analysis. Genotypes A and E were the most prevalent among the lineage I and II strains, respectively, suggesting that polymorphisms distinguishing these genotypes may reflect the ancestral events. Such concordance implies that phage-mediated events caused or were a consequence of divergence of the two lineages. Because integration, excision, and recombination among HB4 and other lambdoid phage could generate the polymorphisms that were detected, the precise nature of the phage-mediated events remains to be determined.

Discussion

To further characterize the genetic relatedness of *E. coli* O157:H7 isolates, we have developed OBGS, a unique strategy for detailed genome comparison based on compositional bias of the genome. The abundance of specific oligomers and their random distribution among coding and noncoding regions of the chromosome (23, 24) provide ideal substrates for comparative genome analyses. Skewed oligomers have been detected in nearly all of the microbial genomes that have been completed (23). In each of the genomes, stand bias of the oligomers shifts abruptly very near the predicted or experimentally determined origin and terminus of replication, suggesting a role in DNA replication (23). In the case of *E. coli*, conservation of the CTG sequence within many of the leading strand oligomers suggests that their complements could serve as priming sites for discontinuous DNA synthesis on the lagging strand (24). However, the CTG motif is not conserved among oligomers across species (23), suggesting that primase specificity is not conserved or that the compositional bias is conserved for other reasons. Nevertheless, the pervasiveness of compositional bias among microbial genomes indicates that OBGS can be used widely for comparative genome analyses.

Based on our calculations (Table 1), single-pass genome coverage in *E. coli* can be accomplished by OBGS with ≈30 octamer combinations. Enough primer combinations can therefore be generated from the matrix of strand-biased octamers and their complements to provide the redundancy necessary to ensure sampling of the entire genome. Because OBGS generates PCR-amplified products, phylogenetically relevant products can be purified from discontinuous gels for reamplification and subsequent DNA sequence analysis.

OBGS Analysis of *E. coli* O157:H7 Populations. Phylogenetic analysis of OBGS products indicates that the O157:H7 clone has diverged into two genetically distinct lineages. Based on mean character differences among pairwise alignments of OBGS characters, strains in lineage I and II averaged 7.6% mismatches ,whereas alignments between strains within each lineage averaged only 2.4% mismatches, implying that the lineages are monophyletic but significantly divergent.

Remarkably, most human and bovine isolates are found in separate lineages. Similar distribution of bovine and human isolates between the two lineages was observed in a geographically and temporally unlinked strain set (Fig. 3), indicating that the lineages are widely disseminated within the United States. The nonrandom distribution of human and bovine isolates among the lineages suggests that lineage II strains may not readily cause disease or may not be transmitted efficiently to humans from bovine sources. This finding could reflect acquisition of a new physiological or virulence characteristic in lineage I that facilitates pathogenesis or transmission to humans. Alternatively, the distribution may reflect a loss of characteristics in lineage II that are necessary for virulence in humans, perhaps as a consequence of adaptation to the bovine environment. In support of the differential virulence hypothesis, epidemiological surveys in central Europe and Australia have found a lower incidence of O157:H7 among human cases of hemorragic colitis in these countries despite prevalence of the organism in cattle similar to that in the U.S., leading to the prediction of a less virulent subpopulation of *E. coli* O157:H7 in animals (3, 4). We are currently evaluating virulence of a representative stain set through *in vitro* assays and animal models, and conducting epidemiological studies to further characterize transmission patterns of the lineages.

The Role of Prophage in Divergence. Although we observed considerable heterogeneity among all of the *E. coli* O157:H7 isolates that were examined (only 71% of OBGS amplicons from O157:H7 strains were monomorphic), the frequency of OBGS products distinguishing isolates in lineage I and II implicated an association of large segments of DNA in divergence of the populations. Concordance of the RFLP subtypes detected by hybridization to the lambdoid phage HB4 and phylogenetic analysis of the strains by OBGS analysis suggests lysogenization, excision, recombination, or other genetic events among prophage was associated with divergence. We have recently determined that presence/absence of HB4 itself is not responsible for the polymorphisms because Southern blot analysis demonstrates that all but a single HB4-genomic band comigrate with monomorphic bands from all O157:H7 strains examined to date (J.N. and A.K.B, unpublished data). However, we cannot yet rule out the possibility of HB4 integration in different regions of the chromosome.

The association of phage-mediated events with divergence suggests that the events either caused or were a consequence of divergence. If the consequence of divergence, these polymorphisms may be selectively neutral relics that serve as convenient markers for divergence. In contrast, lambdoid prophage could impart differential fitness through effects on virulence or physiological characteristics of the host cell. Examples of such characteristics include serum resistance conferred by λ *lom* and *bor* genes (32), restriction and modification systems associated with λ and Rac (33), exclusion of other phages by λ *rexAB* (34), virulence properties by shiga-toxin genes (5, 6), inhibition of chromosomally encoded programmed cell death by λ *rexAB* (35), and putative proteins that alter host signal transduction pathways (36).

Comparative Genomics and Microbial Ecology. Comparative genome analysis by methods such as OBGS is an emerging field of applied genomics that provides new insights into the ecological and physiological basis for microbial population structures. Coverage of the genome by OBGS and DNA sequence analysis of relevant polymorphisms will permit us to pinpoint genome alterations that may have contributed to population structure. Experimentally dissecting the impact of such alterations on gene function and physiology will ultimately provide a means to infer how changes in human practices such as agriculture and animal husbandry suppress and select populations of *E. coli* O157:H7 and other populations of emerging pathogens.

We thank Charles Kaspar and Thomas Cebula for strains and stimulating discussions and Thomas Whittam for assessing the frequency distribution of mismatches in our binary data. We thank Anna Oommen for isolating the HB4 phage and Abraham Oommen for advice on optimizing OBGS PCR conditions. We are grateful to Li-Cor, for the gifts of near-infrared-labeled oligonucleotides. We thank Robert Hutkins, Lawrence Harshman, and Thomas Cebula for critically reviewing this manuscript. This work was supported by Grant 98–35201-6215 from the U.S. Department of Agriculture National Research Initiative Competitive Grants Program to A.K.B. and by funding provided by Nebraska's legislative bill LB1206. This paper is number 12,614 of the Nebraska Agricultural Experiment Station.

1. Griffin, P. M. (1998) in *Escherichia coli O157:H7 and Other Shiga Toxin-Producing E. coli Strains*, eds. Kaper, J. B. & O'Brien, A. D. (Am. Soc. Microbiol., Washington, DC), pp. 15–22.
2. Girffin, P. M. & Tauxe, R. V. (1991) *Epidemiol. Rev.* **13,** 60–98.
3. Caprioli, A. & Tozzi, A. E. (1998) in *Escherichia coli O157:H7 and Other Shiga Toxin-Producing E. coli Strains*, eds. Kaper, J. B. & O'Brien, A. D. (Am. Soc. Microbiol., Washington, DC), pp. 38–48.
4. Robins-Browne, R. M., Elliott, E. & Desmarchelier, P. (1998) in *Escherichia coli O157:H7 and Other Shiga Toxin-Producing E. coli Strains*, eds. Kaper, J. B. & O'Brien, A. D. (Am. Soc. Microbiol., Washington, DC), pp. 66–72.
5. Newland, J. W., Stockbrine, N. A., Miller, F. F., O'Brien, A. D. & Holmes, R. K. (1985) *Science* **230,** 179–181.
6. O'Brien, A. D., Marques, L. R., Kerry, C. F., Newland, J. W. & Holmes, R. K. (1989) *Microb. Pathog.* **6,** 381–390.
7. Donnenberg, M. S., Tzipori, S., McKee, M. L., O'Brien, A. D., Alroy, J. & Kaper, J. B. (1993) *J. Clin. Invest.* **92,** 1418–1424.
8. Dytoc, M., Soni, R., Cockerill, F., III, De Azavedo, J., Louie, M. L., Brunton, J. & Sherman, P. (1993) *Infect. Immun.* **61,** 3382–3391.
9. McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S. &. Kaper, J. B. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 1664–1668.
10. Schmidt, H., Beutin, L. & Karch, H. (1995) *Infect. Immun.* **63,** 1055–1061.
11. Whittam, T. S., Wolfe, M. L., Wachsmuth, I. K., Orskov, F., Orskov, I. & Wilson, R. A. (1993) *Infect. Immun.* **61,** 1619–1629.
12. Whittam, T. S. (1998) in *Escherichia coli and Other Shiga Toxin-Producing E. coli Strains*, eds. Kaper, J. B. & O'Brien, A. D. (Am. Soc. Microbiol., Washington, DC), pp. 195–209.
13. Boerlin, P., Chen, S., Colbourne, J. K., Johnson, R., De Grandis, S. & Gyles, C. (1998) *Infect. Immun.* **66,** 2553–2561.
14. Feng, P., Lampel, K. A., Karch, H. & Whittam, T. S. (1998) *J. Infect. Dis.* **177,** 1750–1753.
15. Faith, N. G., Shere, J. A., Brosch, R., Arnold, K. W., Ansay, S. E., Lee, M.-S., Luchansky, J. B. & Kaspar, C. W. (1996) *Appl. Environ. Microbiol.* **62,** 1519–1525.
16. Dargatz, D. A., Wells, S. J., Thomas, L., Hancock, D. D. & Garber, L. P. (1997) *J. Food Protect.* **60,** 466–470.
17. Hancock, D. D., Rice, D. H., Herriott, D. E., Besser, T. E., Ebel, E. E. & Carpenter, L. V. (1997) *Epidemiol. Infect.* **118,** 193–195.
18. Shere, J. A., Bartlett, K. J. & Kaspar, C. W. (1998) *Appl. Environ. Microbiol.* **64,** 1390–1399.
19. Samadpour, M., Grimm, L. M., Desai, B., Alfi, D., Ongerth, J. E. & Tarr, P. I. (1993) *J. Clin. Microbiol.* **31,** 3179–3183.
20. Datz, M., Janetzki-Milkman, C., Franke, S., Gunzer, F., Schmidt, H. & Karch, H. (1996) *Appl. Environ. Microbiol.* **62,** 791–797.
21. LeClerc, J. E., Baoguang, L., Payne, W. L. & Cebula, T. A. (1996) *Science* **274,** 1208–1211.
22. Besser, T., Hancock, D. D, Pritchett, L. C., McRae, E. M., Rice, D. H. & Tarr, P. I. (1997) *J. Infect. Dis.* **175,** 726–729.
23. Salzberg, S. L., Salzberg, A. J., Kerlavage, A. R. & Tomb, J.-F. (1998) *Gene* **217,** 57–67.
24. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997) *Science* **277,** 1453–1462.
25. Gouveia, S., Proctor, M. E., Lee, M.-S., Luchansky, J. B. & Kaspar, C. W. (1998) *J. Clin. Microbiol.* **36,** 727–733.
26. Lee, M.-S., Kaspar, C. W., Brosch, R., Shere, J. & Luchansky, J. B. (1996) *Vet. Microbiol.* **48,** 223–230.
27. Ochman, H. & Selander, R. K. (1984) *J. Bacteriol.* **157,** 690–693.
28. Swofford, D. L. (1998) PAUP (Sinauer, Sunderland, MA), Version 4.
29. Saitou, N. & Nei. M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
30. Wilshaw, G. A., Smith, H. R., Scotland, S. M., Field, A. M. & Rowe, B. (1985) *J. Gen. Microbiol.* **133,** 1309–1317.
31. Clark, A. J., Sharma, V., Brenowitz, S., Chu, C. C., Sandler, S., Templin, A., Berger, I. & Cohen, A. (1993) *J. Bacteriol.* **175,** 7673–7682.
32. Barondess, J. J. & Beckwith, J. (1990) *Nature (London)* **346,** 871–874.
33. King, G. & Murray, N. E. (1995) *Mol. Microbiol.* **16,** 769–777.
34. Parma, D. H., Snyder, M., Sobolevski, S., Nawroz, M., Brody, E. & Gold, L. (1992) *Genes Dev.* **6,** 497–510.
35. Engelberg-Kulka, H., Reches, M., Marasimhan, S., Schoulaker-Schwarz, R., Klemes, Y., Aizenman, E. & Glaser, G. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 15481–15486.
36. Plunkett, G., III, Rose, D. J., Durfee, T. J. & Blattner, F. R. (1999) *J. Bacteriol.* **181,** 1767–1778.