INFECTION AND IMMUNITY, Sept. 2009, p. 3713–3721 Vol. 77, No. 9
0019-9567/09/$08.00+0 doi:10.1128/IAI.00198-09

Analysis of the Genome of the *Escherichia coli* O157:H7 2006 Spinach-Associated Outbreak Isolate Indicates Candidate Genes That May Enhance Virulence▽†

Bridget R. Kulasekara,[1,2] Michael Jacobs,[3] Yang Zhou,[3] Zaining Wu,[3] Elizabeth Sims,[3] Channakhone Saenphimmachak,[3] Laurence Rohmer,[2] Jennifer M. Ritchie,[4] Matthew Radey,[2] Matthew McKevitt,[2] Theodore Larson Freeman,[2] Hillary Hayden,[3] Eric Haugen,[3] Will Gillett,[3] Christine Fong,[2] Jean Chang,[3] Viktoriya Beskhlebnaya,[5] Matthew K. Waldor,[4] Mansour Samadpour,[5] Thomas S. Whittam,[6] Rajinder Kaul,[3] Mitchell Brittnacher,[2] and Samuel I. Miller[1,2,7,8,9]*

Molecular and Cellular Biology Program, University of Washington, Seattle, Washington[1]; Department of Immunology, University of Washington, Seattle, Washington[2]; University of Washington Genome Center, University of Washington, Seattle, Washington[3]; Howard Hughes Medical Institute, Channing Laboratories, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts[4]; Institute for Environmental Health, Lake Forest Park, Washington[5]; Microbial Evolution Laboratory, National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan[6]; Department of Genome Sciences, University of Washington, Seattle, Washington[7]; Department of Microbiology, University of Washington, Seattle, Washington[8]; and Department of Medicine, University of Washington, Seattle, Washington[9]

Received 19 February 2009/Returned for modification 17 April 2009/Accepted 18 June 2009

In addition to causing diarrhea, *Escherichia coli* O157:H7 infection can lead to hemolytic-uremic syndrome (HUS), a severe disease characterized by hemolysis and renal failure. Differences in HUS frequency among *E. coli* O157:H7 outbreaks have been noted, but our understanding of bacterial factors that promote HUS is incomplete. In 2006, in an outbreak of *E. coli* O157:H7 caused by consumption of contaminated spinach, there was a notably high frequency of HUS. We sequenced the genome of the strain responsible (TW14359) with the goal of identifying candidate genetic factors that contribute to an enhanced ability to cause HUS. The TW14359 genome contains 70 kb of DNA segments not present in either of the two reference O157:H7 genomes. We identified seven putative virulence determinants, including two putative type III secretion system effector proteins, candidate genes that could result in increased pathogenicity or, alternatively, adaptation to plants, and an intact anaerobic nitric oxide reductase gene, *norV*. We surveyed 100 O157:H7 isolates for the presence of these putative virulence determinants. A *norV* deletion was found in over one-half of the strains surveyed and correlated strikingly with the absence of stx_1. The other putative virulence factors were found in 8 to 35% of the O157:H7 isolates surveyed, and their presence also correlated with the presence of *norV* and the absence of stx_1, indicating that the presence of *norV* may serve as a marker of a greater propensity for HUS, similar to the correlation between the absence of stx_1 and a propensity for HUS.

Escherichia coli O157:H7 is a human pathogen that infects more than 73,000 North Americans per year (39). Although infection by this organism typically causes symptoms such as watery or bloody diarrhea, it may also lead to the development of hemolytic-uremic syndrome (HUS), an infection sequela characterized by hemolysis and renal failure that can result in long-lasting kidney damage. Variables that contribute to the development of HUS include host factors, such as age (51), and the genetic background of the enterohemorrhagic *E. coli* (EHEC) isolate. Currently, no effective prophylaxis exists for HUS (45). Antibiotic treatment of *E. coli* O157:H7 infections is contraindicated as it is associated with increased infection sequelae (45, 58).

Humans become infected with EHEC by consuming contaminated food. EHEC are noninvasive pathogens that primarily colonize the human colon. Serotype O157:H7 is the predominant EHEC serotype in North America. The other commonly isolated EHEC serotypes include O26:H11, O103:H2, O111:NM, and O113:H21 (34). The systemic absorption of Shiga toxins produced by intestinal EHEC is thought to damage endothelial cells and to cause HUS (31). Shiga toxins are A-B-type toxins that inhibit protein synthesis. The genes encoding these potent toxins are borne on prophages that are related to phage λ. There are two main variants of Shiga toxin, Stx1 and Stx2. Stx2 is more cytotoxic than Stx1 in cell culture and animal models (27, 46, 48), and epidemiologic observations have revealed that the risk of developing HUS following an EHEC infection is heightened if the isolate produces Stx2 (4). Several variants of Stx2 exist, and Stx2c is the variant most commonly found in O157:H7 strains. Stx2 and Stx2c have the same biological function and possess identical A subunits and B subunits that share at least 97% identity (10).

Although important for virulence, Stx2 does not appear to be the only EHEC factor that significantly influences whether

* Corresponding author. Mailing address: Department of Microbiology, University of Washington, Box 357710, Seattle, WA 98195. Phone: (206) 616-5107. Fax: (206) 616-4295. E-mail: millersi@u.washington.edu.

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

▽ Published ahead of print on 29 June 2009.

patients infected with EHEC develop HUS. A comparison of statistics for several outbreaks caused by Stx2-producing O157:H7 strains showed that the rate of HUS can vary from less than 1% to 26% (23), indicating that strain-specific factors of stx_2-carrying O157:H7 strains are involved in determining clinical outcomes. To date, the most significant factor identified as a factor contributing to the variability is the presence of the stx_1 gene. O157:H7 strains that lack stx_1 but carry one or two stx_2 alleles are more likely to cause infections resulting in HUS (11, 35, 36).

A comparison of the genome sequences of O157:H7 outbreak isolates that have resulted in different HUS rates may provide further insight into genetic factors that contribute to this severe sequela of EHEC infection. The genome sequences of two O157:H7 strains that caused low frequencies of HUS are available. The Sakai strain, the cause of the 1996 outbreak in Japan, caused ~8,000 infections in people, the majority of whom were children, and the rate of HUS was 1.2% (32). In 1982, EDL933 caused the first diarrhea outbreak linked to the O157:H7 serotype and involved 44 individuals but no recorded HUS cases (41).

Sakai shares 4.1 Mb of DNA with the commensal *E. coli* K-12 strain MG1655 and has 296 novel DNA segments more than 19 bp long, termed S-loops, that account for 1.39 Mb. EDL933 shares 4.1 Mb with *E. coli* K-12 strain MG1655 and has 177 unique sequence segments more than 50 bp long, termed O-islands, that account for 1.34 Mb (19). For both the Sakai and EDL933 genomes there is significant evidence of horizontal transfer due to the presence of numerous prophage-related elements and the pO157 virulence plasmid. The virulence factors carried on the O157:H7-specific DNA segments, as well as pO157, include stx_1, stx_2, the locus of enterocyte effacement (LEE), which confers the ability to cause attaching and effacing lesions on enterocytes and, notably, encodes a type III secretion system (TTSS) (22), at least 39 TTSS effectors encoded either on the LEE or at other chromosomal locations (49), numerous fimbrial and nonfimbrial adhesins, and more than one hemolysin (56).

No genome sequence is available yet for an O157:H7 outbreak isolate that has caused an outbreak resulting in a significantly higher HUS rate. One O157:H7 isolate, TW14359, caused an outbreak associated with contaminated spinach that sickened 205 individuals in September and October of 2006. A total of 15% of the afflicted individuals developed HUS (5, 28). This rate is significantly higher than the average annual rate of 4.1% for O157:H7 cases that develop HUS (39). The relatively high percentage of adults, ~8%, who developed HUS in the TW14359 outbreak also likely reflects the greater virulence of this strain (6). Furthermore, Manning et al. performed a phylogenetic analysis of TW14359 utilizing 96 single-nucleotide polymorphisms (SNPs) and demonstrated that this isolate belongs to a more virulent clade of O157:H7 strains (clade 8); the majority of these isolates lack stx_1 and carry stx_2 (28). A partial genome sequence consisting of 200 contigs of the TW14359 genome was also reported by Manning et al., which was found to contain stx_2 and stx_{2c}. While an analysis of these sequence data identified the genes of the two reference isolates that were also present in TW14359 and identified backbone SNPs, it did not provide a list of novel genetic features or provide assembled DNA segments containing repetitive DNA elements, such as phage-like elements. Here we describe the entire genome sequence of this isolate and, focusing on novel genetic material, identify potential genetic features of TW14359 that may promote this strain's outstanding pathogenicity.

MATERIALS AND METHODS

Bacterial strains. TW14359 is a clinical isolate obtained from a patient in Michigan and was described by Manning et al. (28). Of the 100 O157:H7 isolates surveyed by PCR, 54 were obtained from the Institute for Environmental Health in Lake Forest Park, WA. All of these isolates had unique pulsed-field gel electrophoresis patterns obtained using an XbaI restriction digest. All of them were isolated between April 2005 and December 2007. Nineteen isolates were human isolates, 33 isolates were food isolates, and 2 isolates had a canine origin. An additional two isolates had a bovine origin. Thirty-nine isolates were obtained from the Washington State Public Health Laboratory and had a human origin. The remaining five isolates were obtained from patients at Children's Hospital in Seattle, WA, in 1986.

Genomic DNA extraction. For isolation of genomic DNA for PCR, bacterial cultures were grown on LB agar at 37°C overnight. A small amount of cells was resuspended in 0.5 ml phosphate-buffered saline and pelleted, after which a Qiagen QIAamp DNA mini kit was used to purify the genomic DNA. To isolate TW14359 genomic DNA for sequencing, this strain was grown overnight at 37°C with shaking in 40 ml of LB broth in 100-ml flasks. Genomic DNA was isolated using Qiagen maxi kits supplemented with 500/G genomic tips. A single round with two parallel preparations produced approximately 1 mg of total genomic DNA, which was used for all downstream genome-sequencing applications (shotgun plasmid library construction, fosmid library construction, and PCR-based gap-closing finishing experiments).

Genome sequencing. Whole-genome shotgun sequencing was carried out using a small insert plasmid library cloned into the pUC19 vector, as described previously (59). Independent plasmid clones were sequenced utilizing BigDye terminator chemistry and ABI 3730 capillary DNA sequencers. A large insert fosmid library (average insert size, 38 kb) was also constructed using the Epicentre pCC1Fos vector. Fosmid clones were paired-end sequenced to a depth of 8× clone coverage using BigDye terminator chemistry and standard sequencing protocols (40). A total of 40,704 plasmid and 1,920 fosmid paired-end reads were generated, and the sequence data were assembled and viewed using the phred/phrap/consed software package (13, 14, 17). The genome assembly procedure following shotgun sequencing resulted in 641 total contigs with an average Q20 read length of 715 bases and provided about 5× sequence coverage. The genome sequence quality and contiguity were improved further by completing the experiments suggested by successive rounds of the autofinish tool in consed (18). Following six rounds of autofinish, manual finishing was used, including (i) use of specialized sequencing chemistries to sequence difficult regions, (ii) PCR amplification and sequencing of specific targeted regions, (iii) transposon mutagenesis of small insert clones followed by sequencing to fix misassembled or difficult-to-assemble regions, and (iv) shotgun sequencing of the targeted fosmid clones to fix long-range misassemblies in the assembled genome. The consensus sequences for transposon-mutagenized small insert clones and the shotgun-sequenced fosmid clones were used as backbones in the main genome assembly to resolve misassembled regions.

Genome assembly validation. The final genome assembly was validated by using two independent methods. Fingerprint data were generated using 1,152 of the 1,920 paired-end-sequenced fosmid clones by digestion with three independent restriction enzymes, DraI, EcoRV, and PvuII. The fosmid paired-end sequence and experimentally derived fingerprint data were used for assembly validation in combination with the SeqTile software tools (20, 21, 59). Complete correspondence between the virtual and experimentally derived fingerprint patterns of the genome for the restriction enzyme domains of DraI, EcoRV, and PvuII was observed. The finished genome was also validated by comparing the optical map (44) of the genome (carried out at the OpGen Inc. facility in Madison, WI) with the virtual map of the finished genome assembly for BglII and NcoI restriction domains. Complete correspondence at a resolution level of 3 kb or higher was observed.

Genome annotation and analyses. Annotation of genome features for strain TW14359 was based on the annotation of the *E. coli* O157:H7 EDL933 strain at the Enteropathogen Resource Integration Center (http://www.ericbrc.org/). The annotation was completed using the Prokaryotic Genome Annotation Tool developed by us at the University of Washington and using methods described by Rohmer et al. (43). Additional predictions made by RAST were also incorporated (3). Genome comparisons were conducted using the default parameters of

TABLE 1. Notable TW14359-specific polymorphisms

EDL933 locus tag	Sakai locus tag	Gene	Function	Mutation type
Z0865	ECs0739	*nei*	Endonuclease VIII (formamidopyrimidine-DNA glycosylase)	Frameshift
Z0907	ECs0774	*tolA*	Inner membrane protein	90-Nucleotide insertion
Z2516	ECs1858	*yciR*	GGDEF/EAL protein	Frameshift
Z2730	ECs2408	*ydiD*	Short-chain acyl-coenzyme A synthase	Frameshift
Z3469	ECs3100	*yojI*	Multidrug/microcin transporter component	Frameshift
Z4018	ECs3566	*norV*	Anaerobic nitric oxide reductase	204-Nucleotide insertion
Z4947	ECs4412	*bcsB*	Cellulose synthase subunit	18-Nucleotide insertion
Z5317	ECs4733	*hemX*	Putative uroporphyrinogen III C-methyltransferase	36-Nucleotide insertion

MEGABLAST (2, 63). Strain-specific segments were identified as the segments larger than 100 bp that were not represented in MEGABLAST-produced alignments.

PCR. For each 50-ml PCR mixture, 10 ng of genomic DNA and 0.5 ml of Qiagen *Taq* polymerase were used. The reaction mixture was assembled according to the manufacturer's instructions, and the conditions for the PCR were as follows: 94°C for 3 min and 30 cycles of 94°C for 30 s, a primer-specific temperature for 30 s, and 1 min per kb at 72°C. The primer sequences and annealing temperatures are as follows: for ECSP_0242, CCGATTTATGGAGGAAGCC AATG and GCATTACACCCAGGCTTATTCAG were annealed at 59°C; for ECSP_1773, TCTTTAAATTTCATAACAAGGGCA and TGTACGCATCTG TAATCGTCG were annealed at 55°C; for ECSP_2687, AAGAACCCTACTA CCTATTAGCGCC and GGGTTGAGTTCTACCCAAAGTG were annealed at 59°C; for ECSP_2870, TAGTTTGATTCTTGTTGGCGTTCG and TAAAC CTAAAGGCAAACCGTCCTC were annealed at 58°C; for ECSP_2872, GAG CATGGTTGAATGGATAAGCC and CTTCATGATTACCTCGCCGT were annealed at 60°C; for stx_2, AAATGGGTACTGTGCCTGTTACTG and CTTA ACTCCTTTATTTACCCGTTGT were annealed at 51°C; for ECSP_3286, AA CCGATAAGAAACAGTATCCCAG and TGCATGGTGTAACTTGCGGC were annealed at 57°C; for ECSP_3620, CGTGACTGGGAAGTACGAGATT and GAAGTTATCCGGGACTTCACTC were annealed at 60°C; and for stx_1, CCCGGATCCATGAAAAAAACATTATTAATAGC and CCCGAATTCAG CTATTCTGAGTCAACG were annealed at 52°C. As ECSP_2872 was found to be present with ECSP_2870 in the first 50 isolates tested, we chose not to test the remaining 50 isolates for ECSP_2872 or to report the ECSP_2872 results and assumed that the presence of ECSP_2870 reflected the presence of ECSP_2872.

Phage analyses. Coding sequences in the Stx2 and Stx2c phages were compared to a collection of all complete Stx-encoding phage sequences using the BLAST program with an expectation cutoff value of 1E-10. Sequences were scored as homologous if the score was found to be greater than or equal to the number of nucleotides in the gene. The Stx-encoding phage genomes included the genomes of CP-933V (accession no. NC_002655), CP-1639 (AJ304858), YYZ-2008 (FJ184280), VT1-Sakai (AP000400), Stx1-converting phage (AP005153), BP-4795 (AJ556162), VT2-Sakai (AP000363), Stx2-86 (AB255436), Stx2-converting phage II (AP005154), Stx2-converting phage I (AP004402), Min27 (EU311208), BP-933W (AF125520), P27 (AJ298298), 2851 (FM180578), and 1717 (FJ188381.1).

RNA extraction. LB medium in flasks was inoculated using cultures grown overnight in LB to an optical density at 600 nm (OD_{600}) of 0.1. Cultures were grown at 37°C with shaking to an OD_{600} of ~0.6, and then mitomycin C was added at a final concentration of 0.5 μg/ml. To harvest RNA, 1E9 cells were collected per time point (assuming that an OD_{600} of 1 was equivalent to 5E8 cells) and immediately vortexed in 1 ml of TRI reagent (Ambion). Then 0.2 ml of chloroform was added, and the mixture was vortexed and then centrifuged for 15 min. Isopropanol (0.5 ml) was added to the previously collected aqueous phase, and the mixture was incubated for 10 min at room temperature and then centrifuged for 15 min at 4°C. The pellet was washed once with 1 ml of 75% ethanol and then resuspended in 75 ml water. Residual DNA was destroyed by addition of 10 μl of recombinant DNAse (Ambion) and incubated for 30 min at 37°C. RNA was then purified further using a Qiagen RNeasy mini kit.

Quantitative PCR. Quantitative PCR was performed using RNA collected from three independent biological replicates. Each quantitative PCR was performed with an Mx4000 multiplex quantitative PCR system (Stratagene) using samples loaded in triplicate and 2 ng of total RNA. The 10-μl reaction mixture contained SYBR green PCR master mixture (Applied Biosystems) (5 μl 2× master mixture, 400 nM of each primer, 0.5 U of StrataScript reverse transcriptase, 0.5 U of RNase Block), and the following PCR cycling conditions were used: 48°C for 30 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. After each assay, a dissociation curve was constructed to confirm the specificity of all PCR amplicons. Primers were designed using Applied Biosystems Primer Express 2.0 software. For amplification of stx_2, the primers used were Stx2-1072F (AGGATGACACATTTACAGTGAAGGTT) and Stx2-1197R (CACAGGTACTGGATTTGATTGTGAC), and the amplicon size was 126 nucleotides. For detection of stx_{2c} transcripts, the primers used were Stx2c-1072F (AGAATGATACATTCACAGTAAAAGTGGC) and Stx2c-1201R (GATTCA CAGGTACTGGATTTGATTGT), and the amplicon size was 130 nucleotides. PCR amplicons were used for standard curves as 1:4 serial dilutions. The standard curves showed the following reaction efficiencies: for stx_2: 97.5% (R^2 = 0.992); and for stx_{2c}, 91.9% (R^2 = 0.996). The resulting cycle threshold values were converted to numbers of copies, normalized to total RNA, and expressed as averages ± standard deviations for triplicate samples. Total RNA was quantitated with the Mx4000 multiplex quantitative PCR system using triplicate wells with a RiboGreen RNA quantitation kit (Molecular Probes) and standards supplied by the manufacturer. PCR amplicons (used to convert data to copy numbers) were also quantitated with the Mx4000 multiplex quantitative PCR system using triplicate wells with a PicoGreen RNA quantitation kit (Molecular Probes) and standards supplied by the manufacturer.

Stx2 ELISA. An enzyme-linked immunosorbent assay was performed as described by Ritchie et al. (42).

Nucleotide sequence accession numbers. The final finished genome sequence and pO157 sequences have been deposited in the GenBank database under accession numbers CP001368 and CP001369, respectively.

RESULTS AND DISCUSSION

TW14359 nucleotide diversity. The G+C contents of TW14359, Sakai, and EDL933 genomes range from 50.4% to 50.5%. This homogeneity was not unexpected since the SNP diversity of *E. coli* O157:H7 isolates has been shown to be quite low. The TW14359 pO157 plasmid is virtually indistinguishable from that of Sakai, as the two plasmids differ by three nonsynonymous SNPs. A previous analysis of the SNP content in 1,199 O157:H7 chromosomal genes from 11 O157 isolates identified a total of 906 SNPs in 523 genes (62). Integration of TW14359 polymorphisms into this analysis showed that TW14359 has 158 of the 523 characterized SNPs and has an additional 151 novel polymorphisms, 125 of which are located in the 523 genes containing the previously characterized 906 SNPs and 26 of which are located in genes previously identified as invariant (see Tables S1 and S2 in the supplemental material). SNPs specific to this isolate were not identified when the partial genome sequence of TW14359 was published. Table 1 shows a summary of selected TW14359-specific frameshift and insertion mutations compared to reference isolates EDL933 and Sakai (for a complete list see Table S2 in the supplemental material). These TW14359-specific polymorphisms affect the propanoate metabolism pathway, transport and membrane proteins, and recombination-related processes. Two mutations

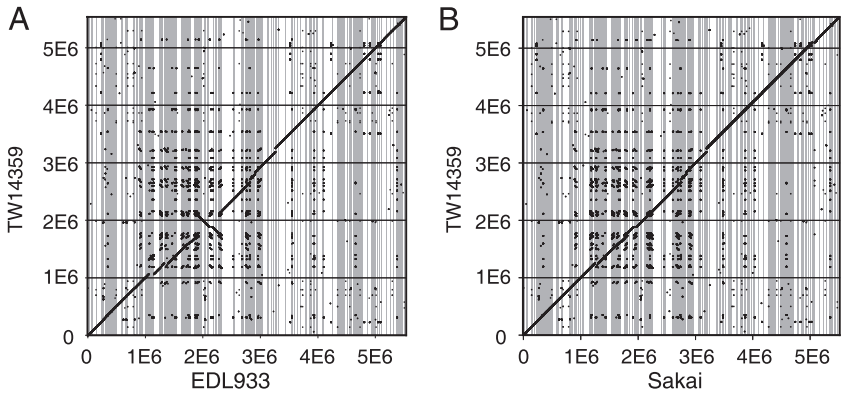

FIG. 1. Whole-genome comparisons. (A) Comparison of TW14359 and EDL933. DNA segments more than 125 bp long that share homology as determined by a MEGABLAST search are compared. TW14359 coordinates are shown on the *y* axis, and EDL933 coordinates are shown on the *x* axis. O-islands are indicated by gray lines. (B) Comparison of TW14359 and Sakai. S-loops are indicated by gray lines.

are involved in cellulose and curli production, traits important for *E. coli* K-12 biofilm formation on plastic and for *E. coli* O157:H7 colonization of alfalfa sprouts (29). TW14359 has an 18-nucleotide insertion in the *bcsB* gene encoding a cellulose synthase component. This insertion preserves the *bcsB* reading frame and encodes a twice-repeated amino acid triplet consisting of leucine, alanine, and valine (LAV). Examination of the sequence surrounding the insertion shows that the LAV triplet is repeated in tandem a total of five times. In the BcsB orthologs that share >97% amino acid identity, the number of LAV triplets varies from zero to six. TW14359 also has a frameshift mutation in *yciR*, a gene involved in modulating the amounts of the bacterial second messenger cyclic diguanylate. Although YciR contains both cyclase and phosphodiesterase domains, an in vitro enzymatic assay of YciR activity detected only phosphodiesterase activity (55). The frameshift mutation in the TW14359 *yciR* allele is located in the phosphodiesterase domain, indicating that expression of the mutant protein would result in a variant of YciR that lacks this activity. Given that the *Salmonella yciR* deletion mutant has a cellulose-overproducing phenotype (15), the *yciR* nonsense mutation in TW14359 may also augment the ability of this isolate to produce cellulose and hence promote colonization of spinach.

Identification of strain-specific DNA. The size of the TW14359 genome is 5,528,136 bp, which is similar to the size of the EDL933 genome (5,528,445 bp) and approximately 29 kb greater than the size of the Sakai genome (5,498,450 bp). We compared the contents of the three sequenced genomes to identify homologous segments in isolates. As shown in Fig. 1, both reference isolates share a significant amount of genomic content with TW14359. The locations of O-islands and S-loops were mapped for each comparison. Although the content of O-islands and S-loops, in general, is preserved, the specific locations of numerous elements within these segments are not preserved. DNA segments within O-islands and S-loops are repeated significantly more often than the segments that are not located within O-islands and S-loops. Upon further examination, many of these elements proved to be phage- and transposon-related genes known to promote genetic diversity.

We identified segments of DNA unique to each isolate (see Table S3 in the supplemental material). As shown in Fig. 2A, the majority of contiguous unique DNA segments are less than 3 kb long. Comparisons not including TW14359 resulted in the smallest number of unique segments per category, while the comparisons including TW14359 produced the largest numbers of unique segments. Data for strain-specific sequences (Table 2) show that TW14359 has the largest amount (70 kb) of unique sequences and that, by this measure, EDL933 and Sakai appear to be more related to each other than to TW14359. The total amount of strain-specific DNA is not correlated with genome size, as shown by the observation that TW14359 has the largest amount of strain-specific DNA while EDL933 has the smallest amount of strain-specific DNA.

Identification of putative virulence determinants. We further examined the sequences specific to TW14359 for the presence of coding sequences that could be responsible for the enhanced virulence of this isolate. The criteria for selection included homology to known virulence factors or to genes involved in resistance against host defense mechanisms or the presence of domains common in eukaryotic proteins but not in prokaryotic proteins. The locations of all strain-specific regions, as well as the TW14359 putative virulence determinants that we have identified, are shown in Fig. 2B.

With the exception of one nearly contiguous 24-kb region made up of a P2 family prophage genome, all of the unique segments account for less than 6 kb. Two genes in the P2 family prophage genome are especially notable because of their potential as virulence factors. Located at adjacent positions, these genes share the greatest homology as well as genetic context with a pair of genes encoded by the gram-positive plant root-colonizing organism *Bacillus amyloliquefaciens* FZB42, which is known for its ability to stimulate plant growth. In the *B. amyloliquefaciens* genome, the homologs of the TW14359 putative virulence factor genes are in a region identified as a genomic island (genomic island 16) (8). ECSP_2870 encodes a protein that is 35% identical to the *B. amyloliquefaciens* RBAM_037120 gene product and carries an SMC-N domain found in eukaryotic proteins involved in the structural maintenance of chromosomes. ECSP_2872 encodes a protein that is 34% identical to the *B. amyloliquefaciens* locus tag RBAM_037130 product and contains signatures of a putative serine esterase domain that is common to eukaryotic proteins.

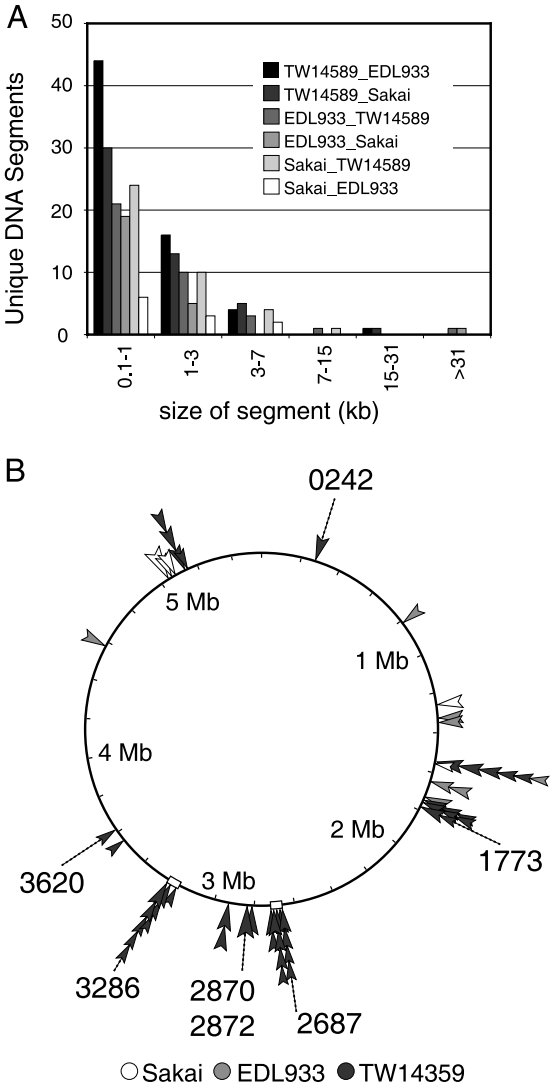

FIG. 2. Strain-specific segments. (A) Histogram for strain-specific DNA segments as determined by pairwise whole-genome comparisons. The strain pairs indicated refer to unique segments of the first strain compared to the second strain. Each pair of strains is represented at the same location in the categories. The categories indicated on the *x* axis include the lower number but not the higher number. (B) TW14359, EDL933, and Sakai strain-specific regions are indicated by arrowheads whose sizes are proportional to the sizes of the strain-specific regions. The smallest strain-specific segment is 104 bp long, and the largest is 34 kb long (both are in the Sakai genome). TW14359 putative virulence determinants are labeled with ECSP locus tag numbers. TW14359 Stx2- and Stx2c-encoding phages are indicated by open rectangles.

Because of the presence of domains commonly found in eukaryotes, these genes possibly encode virulence factors. Alternatively, because of the role of the organism that provided the source of the homology, it is conceivable that these genes could be utilized in some manner to promote adaptation of *E. coli* O157:H7 to the plant host.

Other factors that may contribute to TW14359 virulence include two genes that encode proteins similar to TTSS effector proteins encoded on the *Shigella* sp. virulence plasmid. Locus tag ECSP_2687 encodes a protein that is 32% identical to the *Shigella* protein OspB, which in *Shigella* functions as a virulence factor by reducing cytokine expression through alteration of chromatin remodeling, effectively diminishing the host immune response (64). An OspB homolog is also present in the Stx1-converting bacteriophage BP-4795 that encodes the non-LEE effector NleA, which was isolated from the non-O157:H7 EHEC isolate O84:H4 (9). Following our initial analysis, an OspB homolog was found in the Stx2c-encoding phage 2851, which was isolated from an O157:H7 strain (47). The second gene related to a *Shigella* TTSS effector gene, ECSP_1773, encodes a protein that is 91% identical to the *Shigella* protein OspG. The *Shigella* copy of OspG acts in eukaryotic cells as a virulence factor by binding ubiquitinylated ubiquitin-conjugating enzymes and is thought to interfere with the innate immunity pathway by preventing NF-κB activation (24). *E. coli* O157:H7 also encodes two TTSS effectors, NleH1 and NleH2, both of which share approximately 30% identity with OspG (49). Because both of these proteins are encoded by the three genomes sequenced, it does not appear that they contribute to the variation in virulence of O157:H7 isolates.

Locus tag ECSP_0242 encodes a putative virulence factor that contains five ankyrin repeats, a domain common in eukaryotes that greatly facilitates protein-protein interactions (26). Homologs of this protein with >98% identity have been found in other pathogenic *E. coli* isolates, including enteroinvasive *E. coli* and enteropathogenic *E. coli* isolates. Homologs are uncommon in other bacterial species. Recently, several ankyrin repeat-containing *Legionella pneumophila* and *Coxiella burnetii* virulence factors have been identified (1, 37).

The putative virulence determinant encoded by locus tag ECSP_3286 contains a cytochrome b_5 family domain associated with eukaryotic proteins that has heme binding properties. This protein has significant homology only to a protein produced by *Shigella flexneri* 5 str. 8401. Similar to *Serratia marcescens*, *Shigella dysenteriae*, and *Yersinia*, *Vibrio*, and *Pseudomonas* species, *E. coli* O157:H7 produces an outer membrane protein, ChuA, that binds to and facilitates the transport of extracellular heme (38, 50, 54, 60). The *Serratia*, *Yersinia*, and *Pseudomonas* orthologs of *chuA* encode a protein that also recognizes HasA, a secreted protein that binds heme with a much higher affinity than ChuA (7, 54). As *E. coli* O157:H7 does not possess a *hasA* ortholog, it is possible that the ECSP_3286-encoded product serves in place of HasA.

TABLE 2. Quantification of strain-specific DNA segments as determined by pairwise whole-genome comparisons

Isolate	Isolate(s) compared	Size (kb)
TW14359	EDL933	77
	Sakai	73
	Both isolates	70
EDL933	TW14359	60
	Sakai	15
	Both isolates	11
Sakai	TW14359	90
	EDL933	50
	Both isolates	41

The last unique region of interest is in the open reading frame ECSP_3620 encoding the anaerobic nitric oxide reductase NorV (16). The *norV* copies in EDL933 and Sakai have the same 204-bp deletion that preserves the frame of the coding sequence but results in a loss of 68 amino acids spanning the entire flavodoxin domain, presumably destroying the function of NorV. This suggests that the *norV* gene is functional in TW14359 but not in the less virulent O157:H7 strains sequenced.

Stx phage sequence analysis. The TW14359-specific sequence is overrepresented in the two Stx phages as 22% of this sequence is located in these two regions, which together make up 2.2% of the genome. To obtain an overall estimate of the prevalence of these TW14359-specific regions in other Stx phages, we compared the TW14359 Stx-encoding phages to the other sequenced Stx-encoding phages (Fig. 3). The TW14359-specific regions in the Stx2 phage are clustered near the right end of the phage. In general, the majority of the TW14359-specific regions are rarely found in the other Stx phages. The proteins encoded by these unique regions include proteins related to phage function, such as Roi and NinG family proteins, two putative phage antirepressors, DNA-modifying enzymes, such as a putative endonuclease and a putative DNA methylase, and several hypothetical proteins. Notable TW14359-specific regions include the regions that encode the putative virulence determinant with the probable heme binding protein and the phage integrase. This integrase appears to act on a site in $tRNA^{Arg}$, *argW*, as the 3′ end of this tRNA has been duplicated so that it flanks the Stx2 phage. The TW14359-specific regions in the Stx2c phage are scattered throughout the phage and also, with the exception of the other Stx2c phages, appear to be uncommon in other Stx phages. These regions encode the putative virulence factor OspB (ECSP_2687), an antirepressor-like protein, the antitermination Q family protein, the replication O family protein, a putative replicative DNA helicase, part of the putative phage protein KilA, and the phage integrase. This integrase has specificity for *sbcB*, as indicated by duplication of part of the coding sequence flanking the Stx2c phage (47).

Measurement of Stx2 and Stx2c expression. Strains belonging to the virulent clade identified by Manning et al. (28) are significantly more likely to possess both stx_{2c} and stx_2. This observation raises the question of how much each stx_2 variant contributes to the combined level of the stx_2 transcript or Stx2 protein and consequently the role of each toxin in pathogenesis. In general, expression of both Stx2 and Stx2c is inducible by addition of DNA-damaging agents, such as mitomycin C (10, 42, 47, 61). The stx_2 and stx_{2c} genes are usually located upstream of the phage lysis genes, resulting in coordination of expression of these stx_2 variants with the transition of the Stx phage from the lysogenic stage to the lytic stage. The regulatory pathway for stx_2 and stx_{2c} typically involves phage proteins cI, N, and Q (52). The orthologs of the genes encoding these three proteins in most of the sequenced Stx2 phages are highly conserved, suggesting that the nature of stx_2 transcription and the transition of the Stx2 phage to the lytic cycle in stx_2-carrying isolates, including TW14359, EDL933, and Sakai, should be similar. Additionally, as stx_{2c} phages also show sequence conservation, we expected conserved stx_{2c} phage gene expression upon induction with environmental stress. We designed a set of primers for use in quantitative reverse transcription-PCR that would discriminate between the two transcripts. Measurement of the stx_2 transcript for 3 h following the addition of mitomycin C and comparison with the data for EDL933 (Fig. 4) demonstrated that TW14359 stx_2 transcription is induced by mitomycin C and is induced to a greater extent than EDL933 stx_2 transcription. Conversely, stx_{2c} transcription was not increased significantly following mitomycin C addition, and the overall level of the stx_{2c} transcript was very low, comparable to the basal levels of the uninduced stx_2 transcript. Although strains belonging to the virulent clade identified by Manning et al. are significantly more likely to possess both stx_2 and stx_{2c}, in TW14359 transcription of stx_2 but not transcription of stx_{2c} is induced by addition of mitomycin C, which suggests that Stx2c production may not contribute significantly to the pathogenesis of TW14359 and also supports previous evidence suggesting that certain Stx2 phages exert regulatory control over

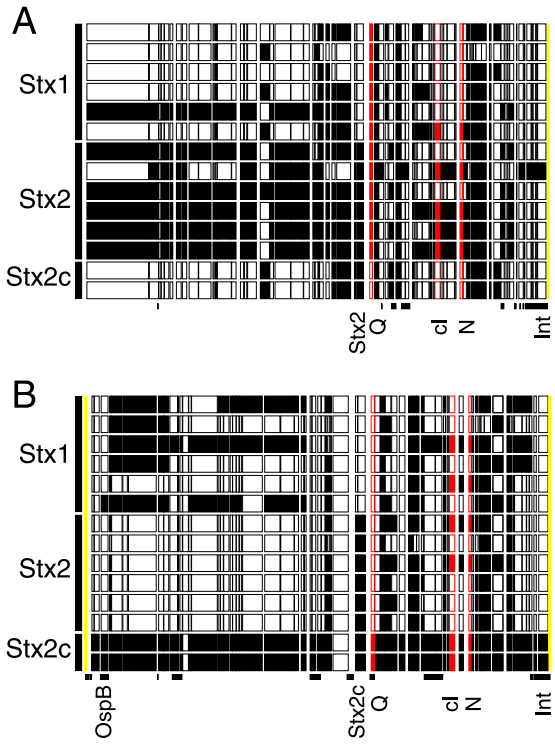

FIG. 3. TW14359 stx_2 and stx_{2c} phages. (A) Comparison of the TW14359 Stx2-encoding phage with other Stx-encoding phages. Each row indicates homology to a specific phage. The phages used are listed in order of *stx* genotype, as indicated on the left, and are (from top to bottom) CP-933V, CP-1639, YYZ-2008, VT1-Sakai, Stx1-converting phage, BP-4795, VT2-Sakai, Stx2-86, Stx2-converting phage II, Stx2-converting phage I, Min27, BP-933W, 2851, and 1717. The TW14359 stx_2 phage coding sequences are indicated by rectangles and are filled if a homologous sequence is present in the corresponding *stx* phage. A phage border is indicated by a vertical yellow line if it is not located at the edge of a coding sequence. The locations of TW14359-specific DNA segments are indicated by the bars below the coding sequence maps. The putative heme binding protein is located in the TW14359-specific sequence that is the fifth sequence from the left. Genes involved in *stx* regulation are labeled and indicated by red. Genes in unique regions are labeled if they are specifically noted in the text. (B) Comparison of TW14359 Stx2c-encoding phages. For details see the explanation for panel A.

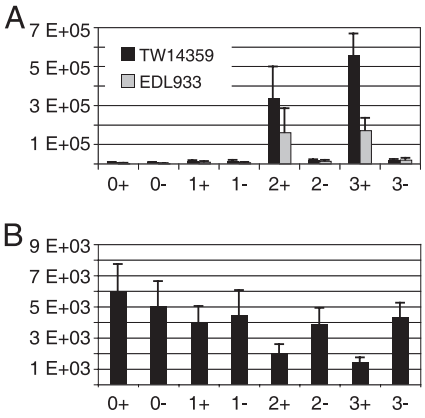

FIG. 4. Transcription of stx_2. (A) Measurements for TW14359 stx_2 and EDL933 stx_2 transcripts in response to mitomycin C. The numbers on the *x* axis indicate the time (in hours) following mitomycin C addition; a plus sign indicates a measurement for a culture containing mitomycin C, and a minus sign indicates a measurement for a culture without mitomycin C. The *y* axis indicates the copy number of a transcript per ng transcript per ng total RNA. The means of three independent experiments are shown. The error bars indicate the standard errors of the means. (B) Measurements for the TW14359 stx_{2c} transcript in response to mitomycin C. For details see the explanation for panel A.

stx_{2c}-carrying phages (33). While absolute levels of Stx2 production are not indicative of virulence, a high level of Stx2 induction in response to mitomycin C addition is characteristic of isolates that cause HUS (42). We measured the amounts of Stx2 in EDL933 supernatants and combined Stx2c and Stx2 protein levels in TW14359 supernatants. For the TW14359 measurements, we assumed that the majority of the toxin was Stx2 as the quantitative reverse transcription-PCR analysis showed that the amount of the stx_2 transcript was almost 300-fold greater than the amount of the stx_{2c} transcript. Three hours following addition of mitomycin C, TW14359 produced 45.4 μg ml^{-1} of Stx2 and Stx2c, while EDL933 produced 37.7 μg ml^{-1} of Stx2. Although similar amounts of Stx2 were produced, Stx2 induction was found to be greater for TW14359 than for EDL933. In the absence of mitomycin C, TW14359 produced 0.3 μg ml^{-1}, while EDL933 produced 0.8 μg ml^{-1}, meaning that the levels of Stx2 induction were 150-fold in TW14359 and approximately 50-fold in EDL933.

Prevalence of putative virulence determinants in *E. coli* O157:H7. We examined 100 *E. coli* O157:H7 isolates to determine the presence of the putative virulence factors, and the results are summarized in Table 3. We found that the majority of isolates that we surveyed, over half of which were clinical isolates, contained at least one of the stx_2 variants (96%), and 58% of the isolates had what appears to be a deletion in *norV* identical to the deletion that occurs in EDL933 and Sakai. It is conceivable that the effect of this deletion is a reduction in colonization or persistence within the large intestine, as this is a known anaerobic environment in which nitric oxide is produced. Nitric oxide also inhibits Stx2 expression (53), making it likely that isolates carrying the *norV*-inactivating mutation produce lower levels of Stx2 in the large intestine, thus potentially attenuating virulence. Interestingly, the *norV* deletion is associated with the presence of stx_1, as 93% of the isolates with the *norV* deletion contain stx_1, while only 10% of the isolates with an intact *norV* gene contain stx_1. The other putative TW14359 virulence factors were present in 8% to 35% of the isolates. An intact *norV* gene correlates strongly not only with the absence of stx_1 but also with the presence of the other putative virulence genes, with the exception of ECSP_1773. This striking association may be due to clonal expansion of an ancestral stx_1 strain with a *norV* deletion, as Manning et al. found that strains carrying stx_1 were not evenly distributed among all clades that have been defined and are overrepresented in clades 2 and 3 and underrepresented in the more virulent clade 8. However, other evidence suggests that selection, at least for deletion of *norV*, occurs in strains carrying stx_1. An intact *norV* gene is present in the genomes of the other *E. coli* pathotypes, as well as in species closely related to *E. coli*, none of which carry stx_1. *S. dysenteriae* is a species that commonly carries stx_1 but presumably acquired this gene by an event distinct from the event in the O157:H7 lineage. Of the two *S. dysenteriae* genome sequences available, that of *S. dysenteriae* 1012 shows an inactivation of *norV* by a transposon insertion, suggesting that the inactivation of *norV* in O157:H7 and in *S. dysenteriae* resulted from independent events, supporting the idea that there is selection for inactivation of *norV* in isolates carrying stx_1. The effect of genetic background on the presence of virulence factors has been observed previously. In *Shigella* spp., a specific genetic background resulting in expression of lysine decarboxylase has been shown to interfere with the activity of the virulence factor enterotoxin (30). Additionally, specific genetic backgrounds have been associated with the presence of certain virulence factors in several *E. coli* pathotypes (12) and in *Pseudomonas aeruginosa* (57), suggesting that the influence of genetic background on the presence of virulence factors may be a more common phenomenon than previously thought.

As the majority of the isolates tested also carry at least one variant of stx_2, the data are interesting in light of the observation that isolates that harbor stx_2 but lack stx_1 pose a higher risk for the development of HUS. Among the isolates surveyed by Manning et al., those that harbor stx_2 and lack stx_1 are also overrepresented in the more virulent clade 8, indicating that it

TABLE 3. Percentages of isolates that contain putative virulence determinants[a]

Genotype	% of isolates with gene or locus tag					
	stx_1	ECSP_ 1773	ECSP_ 0242	ECSP_ 2687	ECSP_ 2870	ECSP_ 3286
norV	10	11	81	64	66	64
Δ*norV*	93	5	2	7	0	0
Total	58	8	35	31	28	27

[a] Both overall percentages and percentages for different *norV* genotypes are indicated.

was perhaps these isolates belonging to clade 8 that were responsible for the observed increase in virulence. Several authors have postulated that it is not the absence of stx_1 that enhances virulence but the presence of associated genetic factors that are responsible for the increased pathogenicity. The presence of an intact *norV* gene combined with any of the virulence factors that we have identified may contribute to the increased virulence observed for stx_1-negative, stx_2-positive isolates and clade 8 strains.

ACKNOWLEDGMENTS

We thank Maynard V. Olson at the University of Washington for support and guidance, Xuan Qin at Seattle Children's Hospital for verifying the serotype of TW14359, Steve L. Moseley at the University of Washington Washington State Public Health Laboratory and Carolyn Hovde at the University of Idaho for providing strains, and John Kemner for coordinating the acquisition of strains. The Diabetes Endocrinology Research Center (supported by NIH NIDDK grant P30 DK-17047) performed the quantitative reverse transcription-PCR.

B.R.K. was a recipient of an NSF graduate research fellowship. This research was funded by NIH grant U54AI057141.

REFERENCES

1. **Al-Khodor, S., C. T. Price, F. Habyarimana, A. Kalia, and Y. Abu Kwaik.** 2008. A Dot/Icm-translocated ankyrin protein of *Legionella pneumophila* is required for intracellular proliferation within human macrophages and protozoa. Mol. Microbiol. **70:**908–923.
2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
3. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9:**75.
4. **Boerlin, P., S. A. McEwen, F. Boerlin-Petzold, J. B. Wilson, R. P. Johnson, and C. L. Gyles.** 1999. Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. J. Clin. Microbiol. **37:**497–503.
5. **California Food Emergency Response Team.** 2007. Investigation of an *Escherichia coli* O157:H7 outbreak associated with Dole pre-packaged spinach. California Department of Health Services, Sacramento. http://www.dhs.ca.gov/fdb/local/PDF/2006%20Spinach%20Report%20Final%20redacted%20no%20photosfigures.PDF.
6. **CDC.** 2006. Update on multi-state outbreak of E. coli O157:H7 infections from fresh spinach, October 6, 2006. Center for Disease Control, Atlanta, GA. http://www.cdc.gov/ecoli/2006/september/updates/100606.htm.
7. **Cescau, S., H. Cwerman, S. Letoffe, P. Delepelaire, C. Wandersman, and F. Biville.** 2007. Heme acquisition by hemophores. Biometals **20:**603–613.
8. **Chen, X. H., A. Koumoutsi, R. Scholz, A. Eisenreich, K. Schneider, I. Heinemeyer, B. Morgenstern, B. Voss, W. R. Hess, O. Reva, H. Junge, B. Voigt, P. R. Jungblut, J. Vater, R. Sussmuth, H. Liesegang, A. Strittmatter, G. Gottschalk, and R. Borriss.** 2007. Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. Nat. Biotechnol. **25:**1007–1014.
9. **Creuzburg, K., J. Recktenwald, V. Kuhle, S. Herold, M. Hensel, and H. Schmidt.** 2005. The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-like type III effector protein. J. Bacteriol. **187:**8494–8498.
10. **de Sablet, T., Y. Bertin, M. Vareille, J. P. Girardeau, A. Garrivier, A. P. Gobert, and C. Martin.** 2008. Differential expression of stx2 variants in Shiga toxin-producing *Escherichia coli* belonging to seropathotypes A and C. Microbiology **154:**176–186.
11. **Eklund, M., K. Leino, and A. Siitonen.** 2002. Clinical *Escherichia coli* strains carrying *stx* genes: *stx* variants and *stx*-positive virulence profiles. J. Clin. Microbiol. **40:**4585–4593.
12. **Escobar-Paramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec, and E. Denamur.** 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Mol. Biol. Evol. **21:**1085–1094.
13. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
14. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
15. **Garcia, B., C. Latasa, C. Solano, F. Garcia-del Portillo, C. Gamazo, and I. Lasa.** 2004. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. Mol. Microbiol. **54:**264–277.
16. **Gardner, A. M., R. A. Helmick, and P. R. Gardner.** 2002. Flavorubredoxin, an inducible catalyst for nitric oxide reduction and detoxification in *Escherichia coli*. J. Biol. Chem. **277:**8172–8177.
17. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
18. **Gordon, D., C. Desmarais, and P. Green.** 2001. Automated finishing with autofinish. Genome Res. **11:**614–625.
19. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
20. **Hayden, H. S., W. Gillett, C. Saenphimmachak, R. Lim, Y. Zhou, M. A. Jacobs, J. Chang, L. Rohmer, D. A. D'Argenio, A. Palmieri, R. Levy, E. Haugen, G. K. Wong, M. J. Brittnacher, J. L. Burns, S. I. Miller, M. V. Olson, and R. Kaul.** 2008. Large-insert genome analysis technology detects structural variation in *Pseudomonas aeruginosa* clinical strains from cystic fibrosis patients. Genomics **91:**530–537.
21. **Hendrickson, E. L., R. Kaul, Y. Zhou, D. Bovee, P. Chapman, J. Chung, E. Conway de Macario, J. A. Dodsworth, W. Gillett, D. E. Graham, M. Hackett, A. K. Haydock, A. Kang, M. L. Land, R. Levy, T. J. Lie, T. A. Major, B. C. Moore, I. Porat, A. Palmeiri, G. Rouse, C. Saenphimmachak, D. Soll, S. Van Dien, T. Wang, W. B. Whitman, Q. Xia, Y. Zhang, F. W. Larimer, M. V. Olson, and J. A. Leigh.** 2004. Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. J. Bacteriol. **186:**6956–6969.
22. **Jerse, A. E., J. Yu, B. D. Tall, and J. B. Kaper.** 1990. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. Proc. Natl. Acad. Sci. USA **87:**7839–7843.
23. **Karch, H., M. Bielaszewska, M. Bitzan, and H. Schmidt.** 1999. Epidemiology and diagnosis of Shiga toxin-producing *Escherichia coli* infections. Diagn. Microbiol. Infect. Dis. **34:**229–243.
24. **Kim, D. W., G. Lenzen, A. L. Page, P. Legrain, P. J. Sansonetti, and C. Parsot.** 2005. The *Shigella flexneri* effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. Proc. Natl. Acad. Sci. USA **102:**14046–14051.
25. Reference deleted.
26. **Li, J., A. Mahajan, and M. D. Tsai.** 2006. Ankyrin repeat: a unique motif mediating protein-protein interactions. Biochemistry **45:**15168–15178.
27. **Louise, C. B., and T. G. Obrig.** 1995. Specific interaction of *Escherichia coli* O157:H7-derived Shiga-like toxin II with human renal endothelial cells. J. Infect. Dis. **172:**1397–1401.
28. **Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. E. Dietrich, W. Zhang, B. Swaminathan, D. Alland, and T. S. Whittam.** 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. USA **105:**4868–4873.
29. **Matthysse, A. G., R. Deora, M. Mishra, and A. G. Torres.** 2008. Polysaccharides cellulose, poly-beta-1,6-*N*-acetyl-D-glucosamine, and colanic acid are required for optimal binding of *Escherichia coli* O157:H7 strains to alfalfa sprouts and K-12 strains to plastic but not for binding to epithelial cells. Appl. Environ. Microbiol. **74:**2384–2390.
30. **Maurelli, A. T., R. E. Fernandez, C. A. Bloch, C. K. Rode, and A. Fasano.** 1998. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. Proc. Natl. Acad. Sci. USA **95:**3943–3948.
31. **Meyers, K. E., and B. S. Kaplan.** 2000. Many cell types are Shiga toxin targets. Kidney Int. **57:**2650–2651.
32. **Michino, H., K. Araki, S. Minami, S. Takaya, N. Sakai, M. Miyazaki, A. Ono, and H. Yanagawa.** 1999. Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. Am. J. Epidemiol. **150:**787–796.
33. **Muniesa, M., M. de Simon, G. Prats, D. Ferrer, H. Panella, and J. Jofre.** 2003. Shiga toxin 2-converting bacteriophages associated with clonal variability in *Escherichia coli* O157:H7 strains of human origin isolated from a single outbreak. Infect. Immun. **71:**4554–4562.
34. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
35. **Orth, D., K. Grif, A. B. Khan, A. Naim, M. P. Dierich, and R. Wurzner.** 2007. The Shiga toxin genotype rather than the amount of Shiga toxin or the cytotoxicity of Shiga toxin in vitro correlates with the appearance of the hemolytic uremic syndrome. Diagn. Microbiol. Infect. Dis. **59:**235–242.
36. **Ostroff, S. M., P. I. Tarr, M. A. Neill, J. H. Lewis, N. Hargrett-Bean, and J. M. Kobayashi.** 1989. Toxin genotypes and plasmid profiles as determinants of systemic sequelae in *Escherichia coli* O157:H7 infections. J. Infect. Dis. **160:**994–998.
37. **Pan, X., A. Luhrmann, A. Satoh, M. A. Laskowski-Arce, and C. R. Roy.** 2008.

Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. Science **320:**1651–1654.
38. **Payne, S. M., E. E. Wyckoff, E. R. Murphy, A. G. Oglesby, M. L. Boulette, and N. M. Davies.** 2006. Iron and pathogenesis of *Shigella*: iron acquisition in the intracellular environment. Biometals **19:**173–180.
39. **Rangel, J. M., P. H. Sparling, C. Crowe, P. M. Griffin, and D. L. Swerdlow.** 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982-2002. Emerg. Infect. Dis. **11:**603–609.
40. **Raymond, C. K., S. Subramanian, M. Paddock, R. Qiu, C. Deodato, A. Palmieri, J. Chang, T. Radke, E. Haugen, A. Kas, D. Waring, D. Bovee, R. Stacy, R. Kaul, and M. V. Olson.** 2005. Targeted, haplotype-resolved resequencing of long segments of the human genome. Genomics **86:**759–766.
41. **Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, N. T. Hargrett, P. A. Blake, and M. L. Cohen.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N. Engl. J. Med. **308:**681–685.
42. **Ritchie, J. M., P. L. Wagner, D. W. Acheson, and M. K. Waldor.** 2003. Comparison of Shiga toxin production by hemolytic-uremic syndrome-associated and bovine-associated Shiga toxin-producing *Escherichia coli* isolates. Appl. Environ. Microbiol. **69:**1059–1066.
43. **Rohmer, L., C. Fong, S. Abmayr, M. Wasnick, T. J. Larson Freeman, M. Radey, T. Guina, K. Svensson, H. S. Hayden, M. Jacobs, L. A. Gallagher, C. Manoil, R. K. Ernst, B. Drees, D. Buckley, E. Haugen, D. Bovee, Y. Zhou, J. Chang, R. Levy, R. Lim, W. Gillett, D. Guenthener, A. Kang, S. A. Shaffer, G. Taylor, J. Chen, B. Gallis, D. A. D'Argenio, M. Forsman, M. V. Olson, D. R. Goodlett, R. Kaul, S. I. Miller, and M. J. Brittnacher.** 2007. Comparison of *Francisella tularensis* genomes reveals evolutionary events associated with the emergence of human pathogenic strains. Genome Biol. **8:**R102.
44. **Schwartz, D. C., and A. Samad.** 1997. Optical mapping approaches to molecular genomics. Curr. Opin. Biotechnol. **8:**70–74.
45. **Serna, A. T., and E. C. Boedeker.** 2008. Pathogenesis and treatment of Shiga toxin-producing *Escherichia coli* infections. Curr. Opin. Gastroenterol. **24:**38–47.
46. **Siegler, R. L., T. G. Obrig, T. J. Pysher, V. L. Tesh, N. D. Denkers, and F. B. Taylor.** 2003. Response to Shiga toxin 1 and 2 in a baboon model of hemolytic uremic syndrome. Pediatr. Nephrol. **18:**92–96.
47. **Strauch, E., J. A. Hammerl, A. Konietzny, S. Schneiker-Bekel, W. Arnold, A. Goesmann, A. Puhler, and L. Beutin.** 2008. Bacteriophage 2851 is a prototype phage for dissemination of the Shiga toxin variant gene 2c in *Escherichia coli* O157:H7. Infect. Immun. **76:**5466–5477.
48. **Tesh, V. L., J. A. Burris, J. W. Owens, V. M. Gordon, E. A. Wadolkowski, A. D. O'Brien, and J. E. Samuel.** 1993. Comparison of the relative toxicities of Shiga-like toxins type I and type II for mice. Infect. Immun. **61:**3392–3402.
49. **Tobe, T., S. A. Beatson, H. Taniguchi, H. Abe, C. M. Bailey, A. Fivian, R. Younis, S. Matthews, O. Marches, G. Frankel, T. Hayashi, and M. J. Pallen.** 2006. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. Proc. Natl. Acad. Sci. USA **103:**14941–14946.
50. **Torres, A. G., and S. M. Payne.** 1997. Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7. Mol. Microbiol. **23:**825–833.
51. **Tserenpuntsag, B., H. G. Chang, P. F. Smith, and D. L. Morse.** 2005. Hemolytic uremic syndrome risk and *Escherichia coli* O157:H7. Emerg. Infect. Dis. **11:**1955–1957.
52. **Tyler, J. S., M. J. Mills, and D. I. Friedman.** 2004. The operator and early promoter region of the Shiga toxin type 2-encoding bacteriophage 933W and control of toxin expression. J. Bacteriol. **186:**7670–7679.
53. **Vareille, M., T. de Sablet, T. Hindre, C. Martin, and A. P. Gobert.** 2007. Nitric oxide inhibits Shiga-toxin synthesis by enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. USA **104:**10199–10204.
54. **Wandersman, C., and P. Delepelaire.** 2004. Bacterial iron sources: from siderophores to hemophores. Annu. Rev. Microbiol. **58:**611–647.
55. **Weber, H., C. Pesavento, A. Possling, G. Tischendorf, and R. Hengge.** 2006. Cyclic-di-GMP-mediated signalling within the sigma network of *Escherichia coli*. Mol. Microbiol. **62:**1014–1034.
56. **Welinder-Olsson, C., and B. Kaijser.** 2005. Enterohemorrhagic *Escherichia coli* (EHEC). Scand. J. Infect. Dis. **37:**405–416.
57. **Wiehlmann, L., G. Wagner, N. Cramer, B. Siebert, P. Gudowius, G. Morales, T. Kohler, C. van Delden, C. Weinel, P. Slickers, and B. Tummler.** 2007. Population structure of *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. USA **104:**8101–8106.
58. **Wong, C. S., S. Jelacic, R. L. Habeeb, S. L. Watkins, and P. I. Tarr.** 2000. The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. N. Engl. J. Med. **342:**1930–1936.
59. **Wood, D. W., J. C. Setubal, R. Kaul, D. E. Monks, J. P. Kitajima, V. K. Okura, Y. Zhou, L. Chen, G. E. Wood, N. F. Almeida, Jr., L. Woo, Y. Chen, I. T. Paulsen, J. A. Eisen, P. D. Karp, D. Bovee, Sr., P. Chapman, J. Clendenning, G. Deatherage, W. Gillet, C. Grant, T. Kutyavin, R. Levy, M. J. Li, E. McClelland, A. Palmieri, C. Raymond, G. Rouse, C. Saenphimmachak, Z. Wu, P. Romero, D. Gordon, S. Zhang, H. Yoo, Y. Tao, P. Biddle, M. Jung, W. Krespan, M. Perry, B. Gordon-Kamm, L. Liao, S. Kim, C. Hendrick, Z. Y. Zhao, M. Dolan, F. Chumley, S. V. Tingey, J. F. Tomb, M. P. Gordon, M. V. Olson, and E. W. Nester.** 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science **294:**2317–2323.
60. **Wyckoff, E. E., A. R. Mey, and S. M. Payne.** 2007. Iron acquisition in *Vibrio cholerae*. Biometals **20:**405–416.
61. **Zhang, W., M. Bielaszewska, A. W. Friedrich, T. Kuczius, and H. Karch.** 2005. Transcriptional analysis of genes encoding Shiga toxin 2 and its variants in *Escherichia coli*. Appl. Environ. Microbiol. **71:**558–561.
62. **Zhang, W., W. Qi, T. J. Albert, A. S. Motiwala, D. Alland, E. K. Hyytia-Trees, E. M. Ribot, P. I. Fields, T. S. Whittam, and B. Swaminathan.** 2006. Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. Genome Res. **16:**757–767.
63. **Zhang, Z., S. Schwartz, L. Wagner, and W. Miller.** 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. **7:**203–214.
64. **Zurawski, D. V., K. L. Mumy, C. S. Faherty, B. A. McCormick, and A. T. Maurelli.** 2009. *Shigella flexneri* type III secretion system effectors OspB and OspF target the nucleus to downregulate the host inflammatory response via interactions with retinoblastoma protein. Mol. Microbiol. **71:**350–368.

Editor: V. J. DiRita