INFECTION AND IMMUNITY, Jan. 1999, p. 230–236 Vol. 19, No. 1
0019-9567/99/$04.00+0

Type-Specific Contributions to Chromosome Size Differences in *Escherichia coli*

CHRISTOPHER K. RODE,[1] LYLA J. MELKERSON-WATSON,[1] AMANDA T. JOHNSON,[1] AND CRAIG A. BLOCH[1,2]*

Department of Pediatrics and Communicable Diseases[1] and Department of Biological Chemistry,[2] University of Michigan School of Medicine, Ann Arbor, Michigan 49109-0656

Received 10 August 1998/Returned for modification 17 September 1998/Accepted 8 October 1998

The *Escherichia coli* genome varies in size from 4.5 to 5.5 Mb. It is unclear whether this variation may be distributed finely throughout the genome or is concentrated at just a few chromosomal loci or on plasmids. Further, the functional correlates of size variation in different genome copies are largely unexplored. We carried out comparative macrorestriction mapping using rare-restriction-site alleles (made with the Tn*10*dRCP2 family of elements, containing the *Not*I, *Bln*I, I-*Ceu*I, and ultra-rare-cutting I-*Sce*I sites) among the chromosomes of laboratory *E. coli* K-12, newborn-sepsis-associated *E. coli* RS218, and uropathogenic *E. coli* J96. These comparisons showed just a few large accessory chromosomal segments accounting for nearly all strain-to-strain size differences. Of 10 sepsis-associated and urovirulence genes, previously isolated from the two pathogens by scoring for function, all were colocalized exclusively with one or more of the accessory chromosomal segments. The accessory chromosomal segments detected in the pathogenic strains from physical, macrorestriction comparisons may be a source of new virulence genes, not yet isolated by function.

The gram-negative bacterium *Escherichia coli* occurs commonly as a benign enteric commensal of mammals. Additionally, different types of *E. coli* characteristically cause different diseases (46), including the hemolytic-uremic syndrome (15, 19, 29, 49), urosepsis (1), and newborn sepsis/meningitis (20). Although recent determination of the entire nucleotide sequence from laboratory strain K-12 indicated 4,639 kb (6), estimates for natural isolates range from 4,660 to 5,300 kb (3). This indicates substantial size differences among genome copies of the various *E. coli* strains. How these differences originated and have persisted is unclear.

Genes for some enterobacterial virulence traits, especially those essential to one or another major pathogenic life cycle, may reside on specialized chromosomal elements, i.e., pathogenicity islands (7, 21, 23, 29, 35, 47); in contrast, others, notably antibiotic resistances, typically reside on plasmids (18). Chromosomal virulence traits may be both difficult to isolate by functional means (e.g., if their phenotypes can be scored only in interactions with mammalian hosts) and impossible to isolate by straightforward physical means (i.e., by plasmid preparations, given that genes for them occur integrated on the chromosome). They could be identified by the positional approach to gene discovery (13, 17), however, if the genes conferring them could be distinguished as local alterations to chromosome structure prior to functional analysis. To investigate the applicability of positional gene discovery for finding genes that contribute to *E. coli* pathogenesis, we mapped the components of chromosomal size differences among laboratory strain K-12 and two pathogenic strains, the uropathogen J96 and the newborn-sepsis-associated strain RS218. Further, we compared the locations of these large, accessory chromosomal segments with the locations of known virulence genes.

* Corresponding author. Mailing address: Department of Pediatrics and Communicable Diseases, University of Michigan School of Medicine, MSRBI, Room A520, 1150 West Medical Center Dr., Ann Arbor, MI 49109-0656. Phone: (734) 763-2005. Fax: (734) 647-9703. E-mail: cbloch@umich.edu.

MATERIALS AND METHODS

Bacterial genetics techniques. Bacterial strains were grown in LB with aeration or on solid LB or M9-glucose (34). Media were supplemented with kanamycin (50 μg/ml), spectinomycin (100 μg/ml), and/or chloramphenicol (15 μg/ml) as required. Cultures were incubated at 37°C, or at 30°C for P1 infections of RS218 and RS218-chimera cultures (10). Cells were stored long term by being suspended in LB-glycerol (80%/20%, vol/vol) and cooled to −80°C. Bacteriophage stocks were grown and stored as described by Sternberg and Maurer (45). Double-insertion mutants of strain MG1655 and single- and double-insertion mutants of strains RS218 and J96 were generated by transducing recipient strains with P1Δdam*rev6* lysates of MG1655 insertion mutants (40). Genome structure, assessed by pulsed-field gel electrophoresis (PFGE) in at least six independent isolates from each transduction, was used to confirm P1 transduction fidelity and lack of transduction-associated rearrangements.

Genomic DNA biophysical techniques. Genomic DNAs were purified from 5-ml overnight cultures of wild-type and insertionally mutagenized *E. coli* in a manner suitable for yielding macrorestriction fragments (50 to 1000 kb), as described previously (40). After digestion of agarose-embedded DNAs with I-*Sce*I (Boehringer Mannheim, Indianapolis, Ind.) for 1 h, *Not*I (New England Biolabs, Beverly, Mass.) for 4 to 5 h, or *Bln*I (Panvera, Madison, Wis.) overnight, according to the manufacturers' directions, and after reaction buffer decanting, agarose dots were melted (70°C) and gently pipetted with plastic 200-μl tips into sample wells in 1.2% agarose (PFGE approved; FastLane; FMC, Portland, Maine) gels for electrophoresis in 0.5× TBE buffer (0.045 M Tris borate, 0.045 M boric acid, 0.001 M EDTA) in a PFGE apparatus (Bio-Rad DR-III) according to the manufacturer's instructions. Ramping of PFGE pulse times was determined as described elsewhere (5); ramping from 11 to 21 s over 13 h and from 50 to 56 s over 7 h was used to approximate log-linear separations between 150 and 350 kb. After electrophoresis of samples with Megabase I and/or II DNA standards (Gibco/BRL, Bethesda, Md.), fragments sizes were quantitated as described elsewhere (25).

RESULTS

Structural and functional correlates to genome size variation within species have only recently been attempted (3, 4). The components of genome size were investigated in three strains from different genealogical branches of *E. coli* (43). The strains analyzed were the nonpathogenic laboratory K-12 strain MG1655 (2), the newborn-sepsis strain RS218 (44), and the uropathogenic strain J96 (28). The chromosomes of these strains vary in length from 4,673 kb for strain MG1655 to 5,195 kb for strain RS218 to 5064 kb for strain J96 (Fig. 1). Also, strains RS218 and J96 carry plasmids of 110 and 113 kb, respectively. The additional ~556 kb of chromosomal DNA in

pathogenic strain RS218 and ~455 kb of chromosomal DNA in pathogenic strain J96 relative to the nonpathogenic strain MG1655 may reside within pathogenicity islands (7, 23, 29, 35), i.e., chromosomal segments on which genes contributing to the virulence of the pathogenic strains reside. Additionally, "black hole" genomic deletions that enhance pathogenicity (32) also need to be considered. Macrorestriction maps of these chromosomes by *Not*I (22, 23, and 25 fragments in strains MG1655, RS218, and J96, respectively), *Bln*I (13, 17, and 13 fragments in strains MG1655, RS218, and J96, respectively), and I-*Ceu*I (7 fragments in all strains) digestions are shown in Fig. 1. Further, through macrorestriction analyses we were able to map the positions in the different strains for a set of 20 rare-restriction-site alleles, made with Tn*10*dRCP2 insertion elements, which carry the rare-cutting polylinker 2 of rare restriction sites including *Not*I, *Bln*I, I-*Ceu*I, and I-*Sce*I.

Integrated macrorestriction mapping with transposons that carry rare restriction sites can distinguish accessory chromosomal segments from conserved chromosomal segments in the physical maps of different chromosomal copies (40); this requires determination of reference loci and of the physical distances separating those loci. Reference loci were determined, and gene order conservation was assessed in these three strains by introduction of Tn*10*dRCP2 insertions carrying the I-*Sce*I restriction site (9). I-*Sce*I is an ultra-rare-cutting megaendonuclease which recognizes an 18-bp nucleotide, TAGGGATA A↓CAGGGTAAT, generating 3′ cohesive ends (36). Statistically, the I-*Sce*I restriction site occurs once in $\sim 6.9 \times 10^{10}$ bp. Therefore, it is not surprising that this sequence does not occur in *E. coli* sequences of ~5 Mb. The Tn*10*dRCP2 family of insertion elements occurs in three antibiotic resistance varieties. Previously, MG1655::Tn*10*dKanRCP2, MG1655::Tn*10*d SpcRCP2, and MG1655::Tn*10*dCamRCP2 insertion mutants were isolated (9). From this strain collection, eight MG1655:: Tn*10*dKanRCP2, nine MG1655::Tn*10*dSpcRCP2, and three MG1655::Tn*10*dCamRCP2 strains (Table 1) were chosen to facilitate comparisons among the chromosomes of *E. coli* MG1655, RS218, and J96 and to localize chromosomal additions/deletions. These MG1655::Tn*10*dRCP2 mutants were chosen to give 20 I-*Sce*I insertions separated from one another by approximately ~250 kb (i.e., ~5,000 kb of *E. coli* genome/20). By this attention to spacing, the ability to resolve chromosomal segment size in the three strain backgrounds between neighboring pairs of I-*Sce*I fragments was optimized. This was because all of the I-*Sce*I fragments generated by adjacent pairs of these evenly spaced insertions could be determined to equivalent accuracy with a single set of PFGE parameters designed to afford log-linear separations between 150 and 350 kb (5). The 20 Tn*10*dRCP2, I-*Sce*I cleavage site landmarks were introduced around the chromosome within either the RS218 or the J96 strain background by P1 transduction (Table 1). The locations of the I-*Sce*I insertions within each strain were mapped relative to that strain's macrorestriction map based on the artificial *Not*I, *Bln*I, and/or I-*Ceu*I site introduced on the Tn*10*dRCP2 element. Locations of the Tn*10*dRCP2 inserts in all three strain backgrounds are shown on a linearized schematic of the *E. coli* chromosome opened at the *thrA* gene at 0 min (Fig. 1). The clockwise order of the 20 insertions was maintained in all three backgrounds, indicating a lack of detectable inversions or translocations. This was despite the potential for inversion between even the different laboratory derivatives of strain K-12 (26, 38) but was indeed expected from general conservation of the *E. coli* genetic map throughout the species and the family *Enterobacteriaceae* (39).

The crossing of pairs of rare-restriction-site alleles between different *E. coli* strain backgrounds allows physical distance

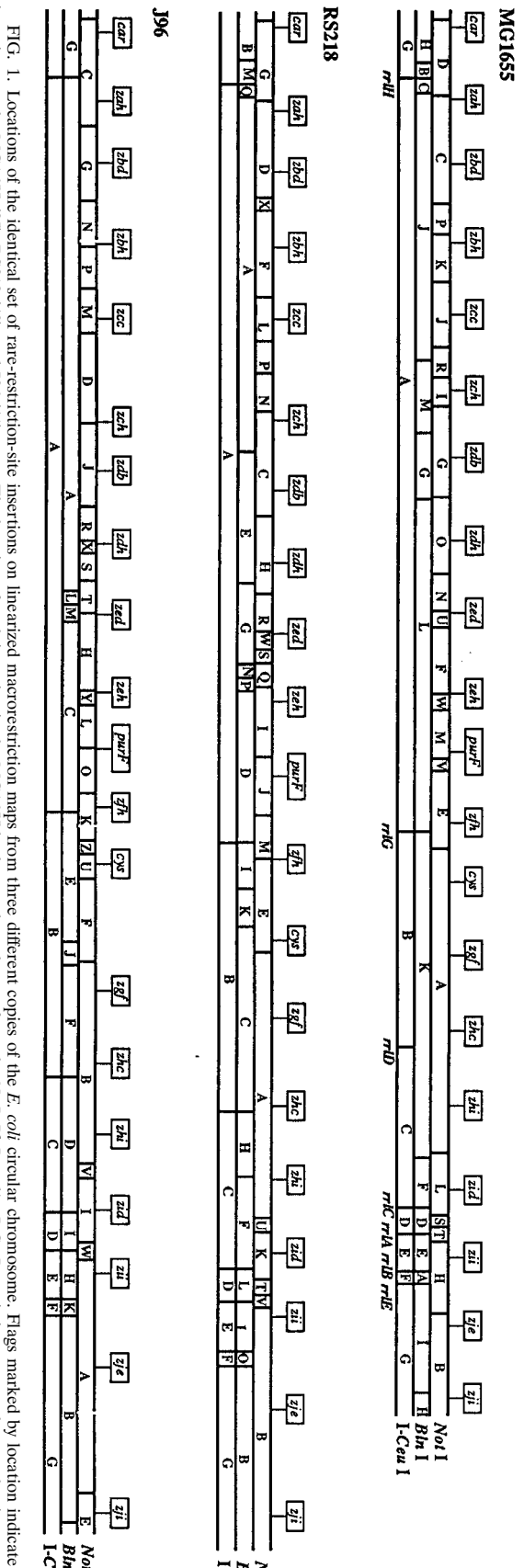

FIG. 1. Locations of the identical set of rare-restriction-site insertions on linearized macrorestriction maps from three different copies of the *E. coli* circular chromosome. Flags marked by location indicate all 20 insertions on the MG1655 (25), RS218 (40), and J96 (41) maps. The location of each insertion in the MG1655 background was known from the *Not*I, *Bln*I, and/or I-*Ceu*I restriction pattern changes that it caused (40); locations of insertions in the RS218 and J96 backgrounds were determined by the same procedure. For each insertion, the allele number and host strains from the three different backgrounds are given in Table 1; note that the actual host strains in Table 1 each carry only one insertion rather than all 20 as depicted here for conciseness.

TABLE 1. *E. coli* strains used in this study[a]

Strain	RS218 transductant	J96 transductant	Description[b]
MG1655			K-12 prototype
RS218			Newborn meningitis prototype
J96			Pyelonephritis prototype
			MG1655::Tn*10*dKanRCP2; Km^r
χM2008	χM3008	χM4008	*zah-108*, (316), [312], {319}
χM2020	χM3020	χM4020	*zbh-120*, (813), [836], {812}
χM2031	χM3031	χM4031	*zch-131*, (1267), [1439], {1385}
χM2064	χM3064	χM4064	*purF-164*, (2419), [2659], {2474}
χM2080	χM3080	χM4080	*cys-180*, (2875), [3158], {2922}
χM2095	χM3095	χM4095	*zhc-195*, (3363), [3701], {3640}
χM2107	χM3107	χM4107	*zid-207*, (3870), [4209], {4148}
χM2127	χM3127	χM4127	*zji-227*, (4504), [4941], {4895}
			MG1655::Tn*10*dSpcRCP2; Sp^r
χM2002	χM3002	χM4002	*car-102*, (35), [35], {35}
χM2014	χM3014	χM4014	*zbd-114*, (575), [597], {573}
χM2026	χM3026	χM4026	*zcc-126*, (1038), [1130], {1031}
χM2046	χM3046	χM4046	*zdh-146*, (1773), [1922], {1803}
χM2061	χM3061	χM4061	*zeh-161*, (2226), [2400], {2281}
χM2070	χM3070	χM4070	*zfh-170*, (2671), [2914], {2723}
χM2088	χM3088	χM4088	*zgf-188*, (3124), [3457], {3401}
χM2099	χM3099	χM4099	*zhi-199*, (3607), [3945], {3889}
χM2121	χM3121	χM4121	*zje-221*, (4315), [4682], {4596}
			MG1655::Tn*10*dCamRCP2; Cm^r
χM2038	χM3038	χM4038	*zdb-138*, (1499), [1668], {1564}
χM2053	χM3053	χM4053	*zed-153*, (1985), [2161], {2012}
χM2115	χM3115	χM4115	*zii-215*, (4080), [4443], {4357}
			MG1655 (Tn*10*dSpcRCP2) (Tn*10*dKanRCP2)
χM2207	χM3207	χM4207	*car-102* (Sp^r) *zah-108* (Km^r)
χM2208	χM3208	χM4208	*zbd-114* (Sp^r) *zbh-120* (Km^r)
χM2209	χM3209	χM4209	*zcc-126* (Sp^r) *zch-131* (Km^r)
χM2210	χM3210	χM4210	*zeh-161* (Sp^r) *purF-164* (Km^r)
χM2211	χM3211	χM4211	*zfh-170* (Sp^r) *cys-180* (Km^r)
χM2212	χM3212	χM4212	*zgf-188* (Sp^r) *zhc-195* (Km^r)
χM2213	χM3213	χM4213	*zhi-199* (Sp^r) *zid-207* (Km^r)
χM2214	χM3214	χM4214	*zje-221* (Sp^r) *zji-227* (Km^r)
			MG1655 (Tn*10*dKanRCP2) (Tn*10*dSpcRCP2)
χM2215	χM3215	χM4215	*zah-108* (Km^r) *zbd-114* (Sp^r)
χM2216	χM3216	χM4216	*zbh-120* (Km^r) *zcc-126* (Sp^r)
χM2217	χM3217	χM4217	*cys-180* (Km^r) *zgf-188* (Sp^r)
χM2216	χM3216	χM4216	*zbh-120* (Km^r) *zcc-126* (Sp^r)
χM2217	χM3217	χM4217	*cys-180* (Km^r) *zgf-188* (Sp^r)
χM2218	χM3218	χM4218	*purF-164* (Km^r) *zfh-170* (Sp^r)
χM2219	χM3219	χM4219	*zji-227* (Km^r) *car-102* (Sp^r)
χM2220	χM3220	χM4220	*zhc-195* (Km^r) *zhi-199* (Sp^r)
			MG1655 (Tn*10*dCamRCP2) (Tn*10*dSpcRCP2)
χM2221	χM3221	χM4221	*zdb-138* (Cm^r) *zdh-146* (Sp^r)
χM2222	χM3222	χM4222	*zii-215* (Cm^r) *zje-221* (Sp^r)
χM2223	χM3223	χM4223	*zed-153* (Cm^r) *zeh-161* (Sp^r)
			MG1655 (Tn*10*dSpcRCP2) (Tn*10*dCamRCP2)
χM2224	χM3224	χM4224	*zdh-146* (Sp^r) *zed-153* (Cm^r)
			MG1655 (Tn*10*dKanRCP2) (Tn*10*dCamRCP2)
χM2225	χM3225	χM4225	*zch-131* (Km^r) *zdb-138* (Cm^r)
χM2226	χM3226	χM4226	*zid-207* (Km^r) *zii-215* (Cm^r)

[a] This study was the source for all strains listed except strain MG1655 (2), strain RS218 (44), strain J96 (28), and χM strains 2002, 2008, 2014, 2020, 2026, 2031, 2038, 2046, 2053, 2061, 2064, 2070, 2080, 2088, 2095, 2099, 2107, 2115, 2121, and 2127 (9).

[b] Locus designations were relative to established transposon insertions, determined as described previously (40), or were determined by auxanography (25). Physical map coordinates are clockwise from the position 25 kb counterclockwise to the (conserved) native *Not*I site nearest the car locus (25). The insertions' coordinates refer to kilobases in the MG1655 (parentheses), RS218 (brackets), and J96 (braces) strain backgrounds; these were determined from the distances between insertions by I-*Sce*I digestions from double mutants (Fig. 3B).

comparisons between pairs of corresponding points in different copies of the *E. coli* genome (9). Biophysical comparison of corresponding genome segments between strains MG1655 and pathogenic strains RS218 and J96 was carried out following introduction of pairs of Tn*10*dRCP2 insertion alleles. This was done by sequential P1 transductions; the distinct antibiotic resistances carried by the neighboring Tn*10*dRCP2 inserts in the set enabled construction of double mutants in the second

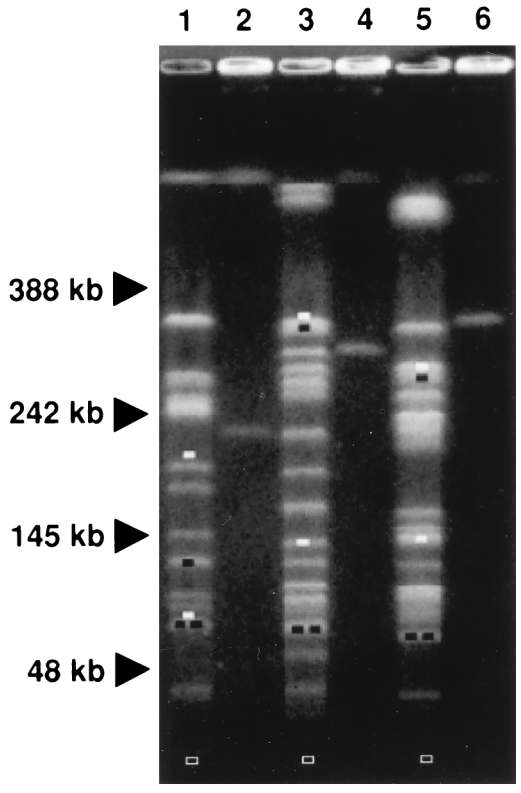

FIG. 2. Determination of macrorestriction fragment length polymorphism on the *E. coli* chromosome. The macrorestriction digestion patterns of genomic DNAs from different double mutants, all bearing the same insertions (*zcc-126* and *zch-131*) in the MG1655 background (lanes 1 and 2), the RS218 background (lanes 3 and 4), and the J96 background (lanes 5 and 6), are shown. Missing native fragments (white bars) and novel subfragments (black bars) generated by *Not*I restriction (lanes 1, 3, and 5) indicated the positions of the insertions in the different strain backgrounds relative to native sites (Fig. 1). (Open bars indicate the electrophoretic positions of 5-kb novel *Not*I subfragments from each background, invisible under the conditions shown.) Unique fragments generated by I-*Sce*I restriction (lanes 2, 4, and 6) indicate by contrast, in readily comparable units allowed by the absence of native sites, the distances between the insertions in the different strain backgrounds (Fig. 3).

step of this process (Table 1). In this way, the chromosome was divided into 20 contiguous and nonoverlapping intervals in each of the three strain backgrounds. A representative macrorestriction-PFGE analysis of corresponding double-insertion mutants, with strains χM2211, χM3211, and χM4211, is shown in Fig. 2. A genomic *Not*I digestion of strain χM2211 (MG1655 *zcc-126*::Tn*10*dSpcRCP2, *zch-131*::Tn*10*dKanRCP2) is shown (Fig. 2, lane 1). The *Not*I pattern serves to verify the Tn*10*dRCP2 insertions into native fragments J_N and I_N. This strain carrying a pair of Tn*10*dRCP2 inserts was detected by the loss of the J_N and I_N bands and the generation of four new subfragments (Fig. 2, lane 1). This same strain was also digested with I-*Sce*I, resulting in a single band of 229 kb (Fig. 2, lane 2). The corresponding double mutants χM3211 (RS218 *zcc-126*::Tn*10*dSpcRCP2 *zch-131*::Tn*10*dKanRCP2) and χM4211 (J96 *zcc-126*::Tn*10*dSpcRCP2 *zch-131*::Tn*10*dKanRCP2) were also digested with *Not*I and I-*Sce*I. For verification of the double insertion into the RS218 background, loss of C_N and M_N was sought (Fig. 2, lane 3); for verification in the J96 background, loss of D_N and M_N was sought (Fig. 2, lane 5). The different sizes of the genome interval, measured in unit fragments following I-*Sce*I digestion, were determined to be 309 kb within χM3211 for RS218 (Fig. 2, lane 4) and 354 kb within χM4211 for J96 (Fig. 2, lane 6). These data indicate that both pathogenic *E. coli* strains RS218 and J96 contain added chromosomal segments within this interval, potentially containing virulence factors (see below).

Similar comparative analyses were repeated for each of the 20 ~250-kb genomic intervals. The results are summarized schematically in Fig. 3A relative to the MG1655 background. The sizes of the I-*Sce*I intervals for the double-Tn*10*dRCP2 strains are given in Fig. 3B. An interval size difference of >7 kb between corresponding I-*Sce*I fragments was taken to indicate substantial addition or deletion. Detected differences of ≤7 kb were considered to have resulted from small rearrangements including transposon and insertion sequence migrations. It is conceivable, however, and a general shortcoming of comparative mapping with rare-restriction-site insertions, that the differences of ≤7 kb could have reflected larger additions or deletions canceling out each other's contributions to size within a given interval. The RS218 chromosome contained 10 unique segments relative to the MG1655 chromosome: *zah* to *zbd* (26 kb), *zbh* to *zcc* (69 kb), *zcc* to *zch* (80 kb), *zdh* to *zed* (27 kb), *zeh* to *purF* (66 kb), *zfh* to *cys* (40 kb); *cys* to *zgf* (50 kb), *zid* to *zii* (24 kb), *zje* to *zji* (70 kb), and *zji* to *thrA* (85 kb). The RS218 chromosome had one deletion of 20 kb relative to the MG1655 chromosome between *zdb* and *zdh*. The J96 chromosome had two deletions relative to strain MG1655: one of 35 kb within the same *zdb*-to-*zdh* interval (as in RS218), and a second of 53 kb between *zch* and *zdb*. The J96 chromosome had four unique segments relative to the MG1655 chromosome: *zcc* to *zch* (125 kb), *zed* to *zeh* (28 kb), *cys* to *zgf* (230 kb), and *zje* to *zii* (110 kb). Three of these J96 unique segments mapped to the same intervals as different unique segments in strain RS218: those at *zcc* to *zch* (22 to 27 min), *cys* to *zgf* (*cys* to 65 min), and *zje* to *zii* (94 to 98 min). These locations contain the virulence factors *sfa* (22 to 27 min in strains RS218 and J96) *kpsA* (64 min in strain RS218), *hlyB*/*D* and *pap* (64 min in strain J96), and *hlyB*/*D*, *prs*, and *cnf* (94 to 98 min in strain J96). Interestingly, some of the same virulence factors (*hlyB*/*D* and *prs*) were originally mapped to yet other chromosomal loci in uropathogenic strain 536 (12).

DISCUSSION

Previously we have shown that insertions containing rare restriction sites can facilitate integrated genome mapping (40). In subsequent work we have shown that pairs of insertions containing the unique I-*Sce*I restriction site allow purification of the genomic intervals that they flank (9). In this study, we combined integrated genome mapping with the isolation of genomic intervals between I-*Sce*I insertions to identify chromosomal segments that distinguish pathogenic from nonpathogenic *E. coli* strains. Through comparisons of corresponding chromosomal segments, we anticipated finding evidence of insertions and/or deletions that contributed to genome size variation and perhaps to pathogenic traits. We identified 11 such chromosomal differences between strains RS218 and MG1655 (10 additions and 1 deletion) and 6 between strains J96 and MG1655 (4 additions and 2 deletions) of genomic segments of 15 kb or larger. These relatively few, large additions and deletions accounted for nearly all genome size differences.

The *E. coli* strains that we examined, MG1655, RS218, and J96, exhibit distinct modes of interaction with mammalian hosts. Strain MG1655 is a nonpathogenic derivative of *E. coli* K-12. Strains RS218 and J96 are pathogenic, especially in targeted subpopulations of the host. These strains also exhibit extensive (up to ~500 kb) variation in chromosome size. De-

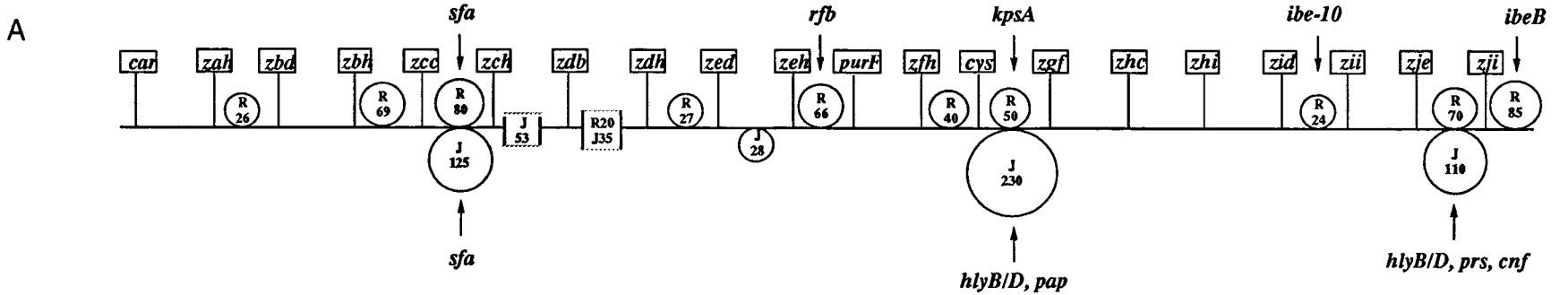

FIG. 3. Three-copy integrated macrorestriction map of the *E. coli* chromosome. Three identical sets of 20 rare-restriction-site insertions, each in a different copy of the chromosome (Fig. 1), were used; the different copies encoded genealogically distant *E. coli* types (i.e., strains MG1655, RS218, and J96). (A) Flags marked by location indicate the borders of the 20 chromosome intervals delimited by adjacent pairs of the insertions; the allele numbers of these insertions are given in Table 1. The distances between the insertions in each strain background were determined, as shown in Fig. 2, by PFGE of I-*Sce*I digests from double mutants constructed by P1 transduction. Intervals containing differences of >7 kb between copies from the different backgrounds are shown as carrying putative copy-specific additions (tangential circles) or putative copy-specific deletions (gaps). The additions and deletions are labeled by size in kilobases and by "R" or "J" for RS218 or J96, respectively. Intervals containing fragment-size differences of ≤7 kb (potentially caused by small rearrangements including insertion-sequence and transposon migrations) are depicted as essentially undisrupted (without circles or gaps). Arrows indicate the known positions of J96 virulence factors (*sfa* [41]; *hlyB/D* and *pap* [47]; *hlyB/D*, *prs*, and *cnf* [47]) and of RS218 virulence factors (*sfa* [41]; *kpsA* [16]; *ibe-10* [7]; *ibeB* [7]). (B) The sizes of I-*Sce*I fragments from MG1655, RS218, and J96 double mutants are indicated for the corresponding chromosome intervals in the gridwork below the map. The additions and deletions represented schematically are reflected in the I-*Sce*I fragment sizes labeled in bold.

spite their extensive structural and functional divergence, overall chromosomal gene order is conserved among these three strains; i.e., the 20 I-*Sce*I insertions are in the same order along all three chromosomes. In addition, the data indicated five different classes of genomic intervals: (i) seven intervals carrying genomic segments of the same length in all three strains, (ii) eight intervals carrying additional genomic segments in one or the other of the two pathogenic strains, (iii) three intervals carrying additional genomic segments in both pathogenic strains, (iv) one interval carrying a genomic deletion relative to strain MG1655 in one of the two pathogenic strains, and (v) one interval carrying different genomic deletions relative to strain MG1655 in both pathogenic strains (Fig. 3). The conservation of overall gene order and of many physical distances among the chromosomes of the three strains indicates a previously unimagined degree of structural identity-by-descent among them. They represent various instances of the *E. coli* chromosome in which different combinations of accessory components and/or deletions have been acquired.

Associated with a majority of the strain-specific chromosomal segments from the two pathogens were genes contributing to some of the key virulence traits distinguishing them. For instance, newborn-sepsis-associated strain RS218 carries genes putatively for penetration of epithelial basement membranes (*sfa*) (24), for immune evasion by molecular mimicry of a fetal brain antigen (*kpsA*) (22), and for penetration of the blood-brain barrier (*ibe-10*) (27). The coordinates for these virulence genes within the *E. coli* chromosome are as follows: *sfa*, 24 min (41); *kpsA*, 64 min (8, 16, 48); *ibe-10*, 87 min (7); and *ibeB*, 98 min (7). Herein we demonstrate the association of these virulence factors with the RS218-specific segments at *zcc* to *zch* (~24 min), *zeh* to *purF* (~47 min), *cys* to *zgf* (~64 min), *zid* to *zii* (~87 min), and *zji* to *thrA* (~98 min), respectively. The uropathogenic strain J96 carries genes putatively for penetration of epithelial basement membrane (*sfa*) (37, 42), for ascendance of the host's ureters (*pap*) (30), for disruption of eukaryotic cells by α-hemolysin (*hly*) and by cytotoxic necrotizing factor 1 (*cnf*) (11), and for adhesion to host tissues (*prs*) (31). The coordinates for these virulence genes within the *E. coli* chromosome are as follows: *sfa* at 24 min (41); *hlyB*/*D* and *pap* at 64 min (47); and *hlyB*/*D*, *prs*, and *cnf* at 94 min (47). Again, these J96 virulence factors are associated with J96-specific segments at *zcc* to *zch* (~24 min), *cys* to *zgf* (~64 min), and *zje* to *zji* (~94 min), respectively. The acquisition of different strain-specific pathogenicity islands within the same genomic regions indicates that these loci are potential hot spots for evolution of pathogenic traits. Insertions of many known pathogenicity islands into the *E. coli* chromosome are at tRNA genes: at the phenylalanine gene *pheV* for the *pap* gene of J96 (47) and the *kpsA* gene of strain RS218 (16); and at the selenocysteine gene *selC* for the locus of enterocyte effacement element (33), and at the phenylalanine gene *pheR* for the *prs* and *hly* genes, of strain J96 (11). Further, strains RS218 and J96 both have large genomic deletions relative to strain MG1655 at *zdb* to *zdh* (~31 min). Strain J96 has a second deletion occurring at *zch* to *zdb* (~27 min). These deletions may constitute virulence black holes (loss of genes enhancing a strain's virulence), like that recently reported for the evolution of *Shigella* spp. and enteroinvasive *E. coli* (32). This inverse complement to pathogenicity islands may also contribute to the evolution of these pathogens by enhancing the pathogen's survival through the loss of chromosomal sequences.

Our findings of large unique components to the *E. coli* chromosome in pathogenic strains are consistent with the work of others showing that virulence factors tend to be clustered both on pathogenicity islands (23) and within particular branches of the *E. coli* tree (14). Further, the segments uniquely absent from the chromosomes of pathogenic strains are consistent with Maurelli et al.'s black-hole concept that loss of chromosomal components may be important in the evolution of pathogenesis (32). In the two pathogenic strains, the unique regions identified by physical chromosomal alignments relative to strain MG1655 were colocalized with known virulence genes, and it is possible that the unaccounted-for coding capacity of these regions may contain new unidentified virulence factors. Also, the unique chromosomal segments in the three strains accounted for most of the genome size differences among them, which may suggest an explanation for the correlation between genome size variation and conventional genetic distance in *E. coli*.

ACKNOWLEDGMENTS

This work was supported by grants R29-AI31419 and R01-AI40074 from the National Institutes of Health to C.A.B.

We thank S. Hanash for kind and enthusiastic support of this work, J. Adams for critically reviewing the manuscript, and Erin McDaid-Kelly and Janice Hatch for technical assistance in preparation of the manuscript.

REFERENCES

1. **Bacheller, C. D., and J. M. Bernstein.** 1997. Urinary tract infections. Med. Clin. North Am. **81:**719–730.
2. **Bachmann, B. J.** 1987. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1190–1219. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
3. **Bergthorsson, U., and H. Ochman.** 1998. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol. Biol. Evol. **15:**6–16.
4. **Bergthorsson, U., and H. Ochman.** 1995. Heterogeneity of genome sizes among natural isolates of *Escherichia coli*. J. Bacteriol. **177:**5784–5789.
5. **Birren, B., and E. Lai.** 1993. Pulsed field gel electrophoresis: a practical guide. Academic Press, Inc., San Diego, Calif.
6. **Blattner, F., G. Plunkett III, C. A. Bloch, N. T. Perna, M. Riley, V. Burland, J. Collado-Vides, J. D. Glassner, C. K. Rode, G. Mayhew, J. Gregor, N. W. Davis, H. Kirkpatrick, M. Goeden, D. Rose, R. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli*. Science **277:**1453–1462.
7. **Bloch, C. A., S.-H. Huang, C. K. Rode, and K. S. Kim.** 1996. Mapping of noninvasion Tn*phoA* mutations on the *Escherichia coli* O18:K1:H7 chromosome. FEMS Microbiol. Lett. **144:**171–176.
8. **Bloch, C. A., and C. K. Rode.** 1996. Pathogenicity island evaluation in *Escherichia coli* K1 by crossing with strain K-12. Infect. Immun. **64:**3218–3223.
9. **Bloch, C. A., C. K. Rode, V. H. Obreque, and J. Mahillon.** 1996. Purification of *Escherichia coli* chromosomal segments without cloning. Biochem. Biophys. Res. Commun. **223:**104–111.
10. **Bloch, C. A., G. M. Thorne, and F. M. Ausubel.** 1989. General method for site-directed mutagenesis in *Escherichia coli* O18ac:K1:H7: deletion of the inducible superoxide dismutase gene, *sodA*, does not diminish bacteremia in neonatal rats. Infect. Immun. **57:**2141–2149.
11. **Blum, G., V. Falbo, A. Caprioli, and J. Hacker.** 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and α-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. FEMS Microbiol. Lett. **126:**189–196.
12. **Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschape, and J. Hacker.** 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606–614.
13. **Botstein, D., R. L. White, M. Skolnick, and R. W. Davis.** 1980. Construction of a genetic linkage map in Man using restriction fragment length polymorphism. Am. J. Hum. Genet. **32:**314–331.
14. **Boyd, E. F., and D. L. Hartl.** 1998. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. J. Bacteriol. **180:**1159–1165.
15. **Campos, L. C., T. S. Whittam, T. A. Gomes, J. R. Andrade, and L. R. Trabulsi.** 1994. *Escherichia coli* serogroup O111 includes several clones of diarrheagenic strains with different virulence properties. Infect. Immun. **62:**3282–3288.
16. **Cieslewicz, M., and E. Vimr.** 1997. Reduced polysialic acid capsule expression in Escherichia coli K1 mutants with chromosomal defects in *kpsF*. Mol. Microbiol. **26:**237–249.
17. **Collins, F., M. Guyer, and A. Chakravarti.** 1997. Variations on a theme: cataloging human DNA sequence variation. Science **278:**1580–1581.

18. **Falkow, S.** 1975. Infectious multiple drug resistance. Pion Limited, London, England.
19. **Feng, P., K. A. Lampel, H. Karch, and T. S. Whittam.** 1998. Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. J. Infect. Dis. **177:**1750–1753.
20. **Ferrieri, P.** 1990. Neonatal susceptibility and immunity to major bacterial pathogens. Rev. Infect. Dis. **12**(Suppl. 4)**:**S394–S400.
21. **Fetherston, J. D., P. Schuetze, and R. D. Perry.** 1992. Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletions of 102 kb of chromosomal DNA, which is flanked by a repetitive element. Mol. Microbiol. **6:**2693–2704.
22. **Finne, J., M. Leinonen, and P. H. Mäkelä.** 1983. Antigenic similarities between brain components and bacteria causing meningitis. Lancet **ii:**355–357.
23. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function, and impact on microbial evolution. Mol. Microbiol. **23:**1089–1097.
24. **Hacker, J., H. Kestler, H. Hoschutzky, K. Jann, F. Lottspeich, and T. K. Korhonen.** 1993. Cloning and characterization of the S fimbrial adhesin II complex of an *Escherichia coli* O18:K1 meningitis isolate. Infect. Immun. **61:**544–550.
25. **Heath, J. D., J. D. Perkins, B. Sharma, and G. M. Weinstock.** 1992. *Not*I genomic cleavage map of *Escherichia coli* K-12 strain MG1655. J. Bacteriol. **174:**558–567.
26. **Hill, C. W., and B. W. Harnish.** 1981. Inversions between ribosomal RNA genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78:**7069–7072.
27. **Huang, S. H., C. Wass, Q. Fu, N. V. Prasadarao, M. Stins, and K. S. Kim.** 1995. *Escherichia coli* invasion of brain microvascular endothelial cells in vitro and in vivo: molecular cloning and characterization of invasion gene *ibe10*. Infect. Immun. **63:**4470–4475.
28. **Hull, R. A., R. E. Gill, P. Hsu, B. H. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33:**933–936.
29. **Knapp, S., J. Hacker, T. Jarchau, and W. Goebel.** 1986. Large, unstable inserts in the chromosome affect virulence properties of uropathogenic *Escherichia coli* O6 strain 536. J. Bacteriol. **168:**22–30.
30. **Lund, B., F. Lindberg, B. I. Marklund, and S. Normark.** 1987. The PapG protein is the alpha-D-galactopyranosyl-(1-4)-beta-D-galactopyranose-binding adhesin of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA **84:**5898–5902.
31. **Marklund, B.-I., J. M. Tennent, E. Garcia, A. Hamers, M. Baga, F. Lindberg, W. Gaastra, and S. Normark.** 1992. Horizontal gene transfer of the *Escherichia coli pap* and *prs* pili operons as a mechanism for the development of tissue-specific adhesive properties. Mol. Microbiol. **6:**2225–2242.
32. **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:**3943–3949.
33. **McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg, and J. B. Kaper.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. USA **92:**1664–1668.
34. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Mills, D. M., V. Bajaj, and C. A. Lee.** 1995. A 40kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. Mol. Microbiol. **15:**749–759.
36. **Monteilhet, C., A. Perrin, A. Thierry, L. Colleau, and B. Dujon.** 1990. Purification and characterization of the *in vitro* activity of I-*Sce*I, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acids Res. **18:**1407–1413.
37. **Morschhauser, J., V. Vetter, T. Korhonen, B. E. Uhlin, and J. Hacker.** 1993. Regulation and binding properties of S fimbriae cloned from *E. coli* strains causing urinary tract infection and meningitis. Zentbl. Bakteriol. **278:**165–176.
38. **Perkins, J. D., J. D. Heath, B. R. Sharma, and G. M. Weinstock.** 1993. *Xba*I and *Bln*I genomic cleavage maps of *Escherichia coli* K-12 strain MG1655 and comparative analysis of other strains. J. Mol. Biol. **232:**419–445.
39. **Riley, M., and S. Krawiec.** 1987. Genome organization, p. 967–981. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
40. **Rode, C. K., V. H. Obreque, and C. A. Bloch.** 1995. New tools for integrated genetic and physical analyses of the *Escherichia coli* chromosome. Gene **166:**1–9.
41. **Rode, C. K., L. Xhang, B. Foxman, and C. A. Bloch.** Unpublished data.
42. **Schmoll, T., H. Hoschutzky, J. Morschhauser, F. Lottspeich, K. Jann, and J. Hacker.** 1989. Analysis of genes coding for the sialic acid-binding adhesin and two other minor fimbrial subunits of the S-fimbrial adhesin determinant of *Escherichia coli*. Mol. Microbiol. **3:**1735–1744.
43. **Selander, R. K., D. A. Caugant, and T. S. Whittam.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1647. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
44. **Silver, R. P., C. W. Finn, W. F. Vann, W. Aaronson, R. Schneerson, P. J. Kretschmer, and C. F. Garon.** 1981. Molecular cloning of the K1 capsular polysaccharide genes of *E. coli*. Nature **289:**696–698.
45. **Sternberg, N. L., and R. Maurer.** 1991. Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella typhimurium*. Methods Enzymol. **204:**18–43.
46. **Sussman, M.** 1997. *Escherichia coli*: mechanisms of virulence. Cambridge University Press, New York, N.Y.
47. **Swenson, D., N. Bukanov, D. Berg, and R. Welch.** 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. Infect. Immun. **64:**3736–3743.
48. **Vimr, E.** 1991. Map position and genomic organization of the *kps* cluster for polysialic acid synthesis in *Escherichia coli* K1. J. Bacteriol. **173:**1335–1338.
49. **Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Orskov, I. Orskov, and R. A. Wilson.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect. Immun. **61:**1619–1629.

Editor: P. E. Orndorff