JOURNAL OF BACTERIOLOGY, Mar. 1994, p. 1673–1682
0021-9193/94/$04.00+0

Vol. 176, No. 6

DNA Sequence Divergence among Derivatives of *Escherichia coli* K-12 Detected by Arbitrary Primer PCR (Random Amplified Polymorphic DNA) Fingerprinting

IGOR BRIKUN,[1]† KESTUTIS SUZIEDELIS,[1]‡ AND DOUGLAS E. BERG[1,2]*

Departments of Molecular Microbiology[1] and Genetics,[2] Washington University Medical School, St. Louis, Missouri 63110

Received 3 August 1993/Accepted 31 December 1993

Derivatives of *Escherichia coli* K-12 of known ancestry were characterized by random amplified polymorphic DNA (RAPD) fingerprinting to better understand genome evolution in this family of closely related strains. This sensitive method entails PCR amplification with arbitrary primers at low stringency and yields arrays of anonymous DNA fragments that are strain specific. Among 150 fragments scored, eight were polymorphic in that they were produced from some but not all strains. Seven polymorphic bands were chromosomal, and one was from the F-factor plasmid. Five of the six mapped polymorphic chromosomal bands came from just 7% of the genome, a 340-kb segment that includes the terminus of replication. Two of these were from the cryptic Rac prophage, and the inability to amplify them from some strains was attributable to deletion (excision) or to rearrangement of Rac. Two other terminus-region segments that resulted in polymorphic bands appeared to have sustained point mutations that affected the ability to amplify them. Control experiments showed that RAPD bands from the 340-kb terminus-region segment and also from two plasmids (P1 and F) were represented in approximate proportion to their size. Optimization experiments showed that the concentration of thermostable polymerase strongly affected the arrays of RAPD products obtained. Comparison of RAPD polymorphisms and positions of strains exhibiting them in the pedigree suggests that many sequence changes occurred in these historic *E. coli* strains during their storage. We propose that the clustering of such mutations near the terminus reflects errors during completion of chromosome replication, possibly during slow growth in the stab cultures that were often used to store *E. coli* strains in the early years of bacterial genetics.

Nearly half a century has passed since Gray and Tatum reported the first isolation of laboratory-generated nutritional mutations in *Escherichia coli* K-12 (7). Thousands of derivatives of the original strain have been generated since then, and the ancestries of many of them are known from Bachmann's painstaking tracing of pedigrees in publications and scattered laboratory notes (2, 3). Some early derivatives arose spontaneously, but most were induced, often with heavy doses of radiation or chemical mutagens. Diverse mutant phenotypes were selected; among the most frequent were requirements for particular amino acids or other metabolites, the inability to utilize particular carbon sources, and resistances to phages and other antimicrobial agents.

We have studied DNA sequence level divergence among historic *E. coli* K-12 strains, using the arbitrary primer PCR (or RAPD, for random amplified polymorphic DNA) method of DNA fingerprinting (27, 30) to better understand how these strains have evolved. In this method, single oligonucleotides of arbitrarily chosen sequence are used as primers in low-stringency PCRs. Arrays of DNA fragments are generated that reflect annealing of the primer to closely linked pairs of sites that fortuitously match or partially match the primer sequence. The arrays of fragments generated are strain specific and distinguish closely related strains with great sensitivity and efficiency relative to other strain typing methods (19). These features make RAPD fingerprinting extremely useful in epidemiology and in studies of microbial population genetics and evolution (1, 26, 28).

Here we report differences in arrays of RAPD products (DNA bands) obtained from some historic derivatives of *E. coli* K-12 and the characterization of representative polymorphic bands. Our results indicate that numerous sequence changes accumulated in strains isolated in the early years of microbial genetics and that these mutations tend to be clustered near the terminus of chromosome replication. We propose that such mutations result from errors while completing cycles of DNA replication, possibly associated with prolonged stalling of DNA replication complexes on incompletely replicated chromosomes.

MATERIALS AND METHODS

Bacterial growth and strains. Bacteria were grown in L broth (23) or minimal salts medium with glucose (25) unless specified otherwise at 37°C or, in the case of lysogens carrying a P1 CM *clr100* prophage, at 30°C. The strains of *E. coli* K-12 used in this work are listed in Table 1. The known relationships among the 26 historic strains, chosen to reflect early steps in the divergence of the K-12 family of *E. coli* strains (2, 3), are summarized in Fig. 1 and Table 1. These strains were obtained directly from Barbara Bachmann (*E. coli* Genetics Stock Center, Yale University) during 1992 and 1993. Upon receipt, single colony isolates of these strains were grown to stationary phase and stored at −70°C in LB plus 40% glycerol; they were also verified by testing for diagnostic nutritional requirements and/or fermentation traits.

* Corresponding author. Department of Molecular Microbiology, Box 8230, Washington University Medical School, St. Louis, MO 63110. Phone: (314) 362 2772. Fax: (314) 362 1232. Electronic mail address: Berg@Borcim.wustl.edu.

† Present address: Thermogene, Inc., Chicago, IL 60612.

‡ Present address: Department of Biochemistry and Biophysics, Vilnius University, Vilnius 2009, Lithuania.

TABLE 1. *E. coli* K-12 strains

Group	Designation	Genotype[a]	Year[b]	CGSC accession no.	Reference or source
Historic[c]	EMG2	F^+ *fnr* (λ^+)	1944	4401	
	679	As EMG2, but *thr-1*	1944	5588	
	679–680	As 679, but F^- *leuB6 rfbD1*	1945	5589	
	Y10	As 679-680, but *supE44 thi-1*	1947	5037	
	Y53	As Y10, but *lacY1* (unstable)	1947	4429	
	Y70	As Y53, but *lacY1* (stable)	1952	5356	
	C600	As Y70, but *tonA21* λ^-	1954	3004	
	CR34Thy$^-$	As C600, but *thyA6 deoC*	1960	6707	
	W1	As Y53, but *malA1* λ^r	1953	6674	
	W583	As W1, but *ara-13 tonA2 galT1 xyl-7*	1956	258	
	P678	As W583, but *malA*$^+$ *gal-6 galP63 malT1 mtlA*	1961	5265	
	PA340	As P678, but *hisG1 rpsL9 argH1 gdh-1* Δ(*gltB-gltF*) *500* λ^-	1969	5550	
	P678S^R	As P678, but *rpsL135* λ^-	1961	5239	
	W894	As W1, but *galK2*	1952	4956	
	AB1157	As W894, but *mal*$^+$ Δ(*gpt-proA*)*62 tsx-33 hisG4*(Oc) *rfbD1 mgl-51 rpsL31 kdgK51 xyl-5 mtl-1 argE3*(Oc) *ara-14* Rac$^-$ λ^-	1964	1157	
	JC7623	As AB1157, but *sbcC201 sbcB15 recB21 recC22*	1970	5188	
	58	As EMG2, but *bio-1 spoT1*	1944	5587	
	58-161	As 58, but *metB1*	1945	4425	
	W6	As 58-161, but *bio*$^+$ *relA1*	1950	4427	
	Hfr Broda 8	As W6, but Hfr PO118 λ^- λ^r	1967	5015	
	Hfr Hayes	As W6, but Hfr *azi-7 rpsL100* PO1	1953	5492	
	Hfr Cavalli	As W6, but Hfr PO2	1950	4410	
	W4032	As W6, but *proA3 lac-3 tsx-76*	1955	5399	
	WG1	FΔ (λ^+)	1953	5073	
	W1485	As WG1, but F^+ λ^- *fnr*	1953	5024	
	MG1655	As W1485, but F^-	1980	6300	
Other	DB1342	F^- *pro met thi* [P1 CM *clr100*]			21
	DB1631	W3110 [F1-4]			Laboratory collection
	DB7590 (Y10 Δ1608)	Y10Δ1608 *zdg-232*::Tn*10*			P1 transduction from PK1647
	DB7591 (C600Δ1608)	C600 Δ1608 *zdg-232*::Tn*10*			P1 transduction from PK1647
	DB7593 (C600 [F])	C600 [F1-4][d]			Transfer from DB1631
	DB7595 (C600 Δ1608 [F])	C600 Δ1608 *zdg-232*::Tn*10* [F1-4]			P1 transduction from PK1647, transfer from DB1631
	DB7596 (Y10 [P1])	Y10 [P1 CM *clr100*]			Lysogenization
	DB7598 (C600 [P1])	C600 [P1 CM *clr100*]			Lysogenization
	DB7599 (C600 [F, P1])	C600 [F1-4, P1 CM *clr100*]			Transfer from DB1631; lysogenization
	DH5α	*supE44* Δ*lac*U169 (ϕ80 lacZΔM15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*			23
	PK1647	F^- *trpR trpA9605*(Am) *his29*(Am) *ilv pro-2 arg-427 thyA deoB* or *deoC tsx Δ1608 zdg-232*::Tn*10*			9

[a] The following strains were λ lysogens: WG1, EMG2, 679, 679-680, P678, Y10, Y53, Y70, W1, W583, W894, 58, 58-161, W6, Hfr Hayes, Hfr Cavalli, and W4032; and the following were not lysogenic for λ: C600, CR34Thy$^-$, PA340, P678S^R, AB1157, JC7623, W1485, MG1655, and Hfr Broda 8 (3).

[b] Year of first publication (summarized in reference 3). For some strains, years of isolation are inferred as follows: W1 and W894 are inferred to have been isolated by 1953 because they are ancestral to W945, which was published in 1953 (3); W6 is inferred to have been isolated by 1950 because it is ancestral to Hfr Cavalli.

[c] All are from Barbara Bachmann, *E. coli* Genetics Stock Center (CGSC), and are reported in reference 3.

[d] F1-4 designates F-factor plasmid with Tn*3* insertion (5).

Derivatives of strains Y10 and C600 carrying a 340-kb deletion of a segment that contains the terminus of chromosome replication (Δ1608 [9]) (Fig. 2) and/or ~100-kb plasmid F or P1 (20, 29) were constructed. Δ1608 was introduced by P1 transduction from strain PK1647, in which Tn*10* (Tet^r) is linked to Δ1608. Tet^r recombinants were selected, and those containing the Δ1608 allele were identified by (i) their relatively small colony size on L agar and (ii) their simultaneous loss of *tyrR* and *manA* (Fig. 2), which were scored by growth on minimal medium containing a 20-μg/ml concentration of *m*-fluoro-DL-tyrosine, which inhibits *tyrR*$^+$ cells, and formation of white colonies on MacConkey mannose agar, respectively (9). Derivatives carrying an F factor were constructed by conjugation, using an F factor marked with Tn*3* (Amp^r) (designated F1-4) (5). Derivatives carrying the P1 plasmid prophage were generated by lysogenization with phage P1 CM *clr100* (Cam^r) (21).

Growth conditions and DNA preparation. *E. coli* cells were grown with aeration to stationary phase in 2 ml of L broth, and chromosomal DNA was extracted as follows. First, 0.5 ml of culture was centrifuged in a microcentrifuge, and the cells were resuspended in 200 μl of GTEL buffer (50 mM glucose, 50 mM Tris HCl [pH 8.0], 50 mM EDTA, 10 mg of lysozyme per ml) and incubated for 5 min at room temperature. Then 500 μl of

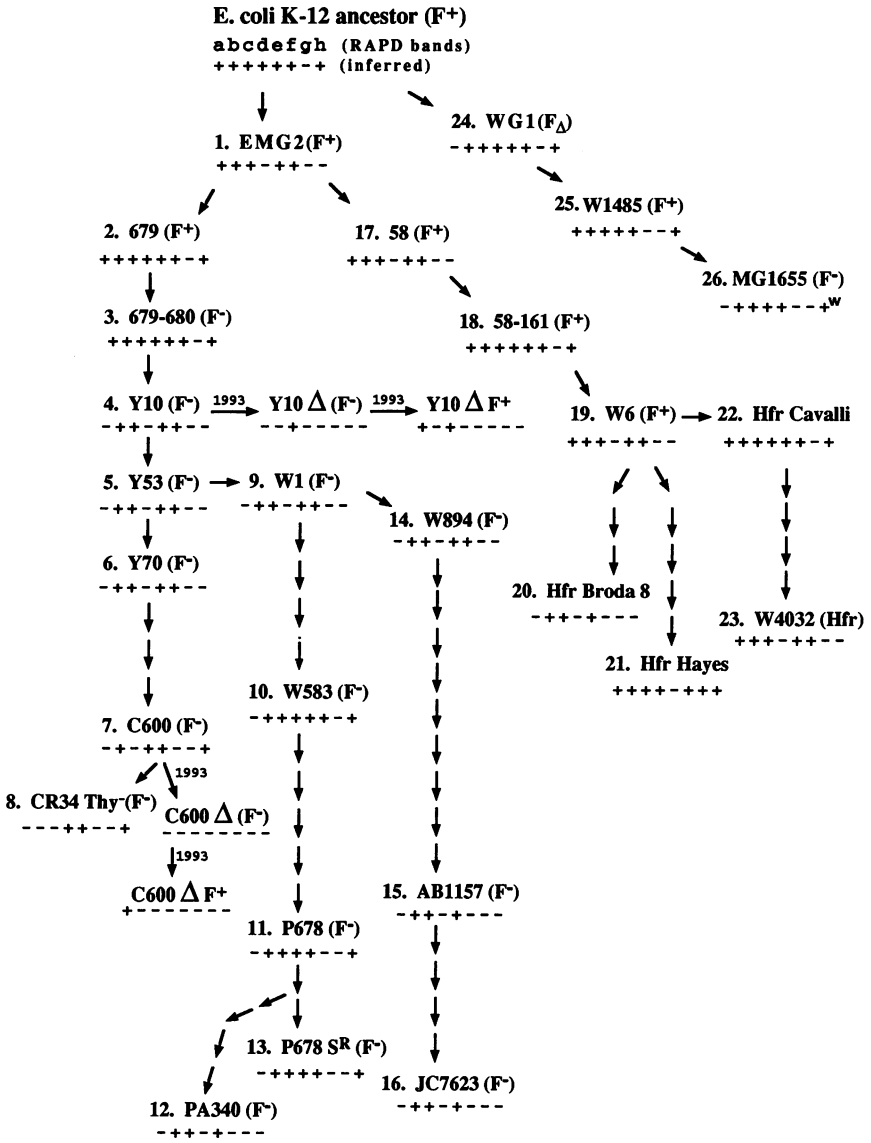

FIG. 1. Pedigrees of *E. coli* K-12 strains used in this study. Each arrow represents a single recorded cycle of spontaneous or induced mutation and deliberate selection. The historic strains are indicated by numbers 1 through 26; all of these strains were from the *E. coli* Genetics Stock Center (3), and most were isolated in the 1940s and 1950s (see also Table 1). The steps leading to four strains specifically constructed for this study are indicated with "1993," to indicate the year of construction. The series of + and − signs under each strain name indicates the ability or inability, respectively, to yield RAPD bands a through h which were polymorphic among these strains (see Fig. 4 and Table 2); $+^{w}$ designates weak amplification. The presence or absence of the F-factor plasmid in each historic strain is indicated in parentheses, because experiments summarized in the text and in Fig. 3 and 6 show that RAPD product a is derived from the F factor. The FΔ designation of WG1 indicates that its F factor contains a large (≥40-kb) deletion (24). Y10Δ and C600Δ designate derivatives of Y10 and C600 carrying the 340-kb Δ1608 deletion, which removes a region containing the replication terminus and closely linked *zdg232*::Tn*10* insertion (9). Most of the inferred relationships among strains presented here are summarized in reference 3. However, the inference that EMG2 is the ancestor of strains 679 and 58 was based on comparison of pedigree information in reference 7 with that in reference 6. WG1 is inferred to be the parent of W1485 on the basis of information given in reference 2. It should be noted that the results presented here indicate continued divergence of many pairs of parental and progeny or sibling strains after they were separated from one another (see Fig. 5). A strain identical to the ancestral wild-type *E. coli* K-12 strain used by Gray and Tatum (7) to isolate the first *E. coli* auxotrophs was not available to us; the array of polymorphic RAPD bands that it should have yielded is inferred from the arrays generated from its derivatives, as summarized here.

TESK lysis buffer (50 mM Tris HCl [pH 8.0], 50 mM EDTA, 1% sodium dodecyl sulfate [SDS], 50 μg of proteinase K per ml) was added, and the solution was incubated for 2 h at 55°C. RNase was added to a final concentration of 20 μg/ml, and the mixture was incubated for 10 min at 55°C. The solution was extracted with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1); 1/10 volume of 3 M sodium acetate was added. The DNA was precipitated with 2 volumes of ethanol, washed with 70% ethanol, and dissolved in 100 μl of distilled water. An aliquot was electrophoresed in a 1% agarose gel and stained with ethidium bromide to estimate the DNA yield and verify DNA integrity. Generally 30 to 50 μg of DNA was obtained from 0.5 ml of culture. Control experiments showed that RAPD profiles from samples prepared in this way were

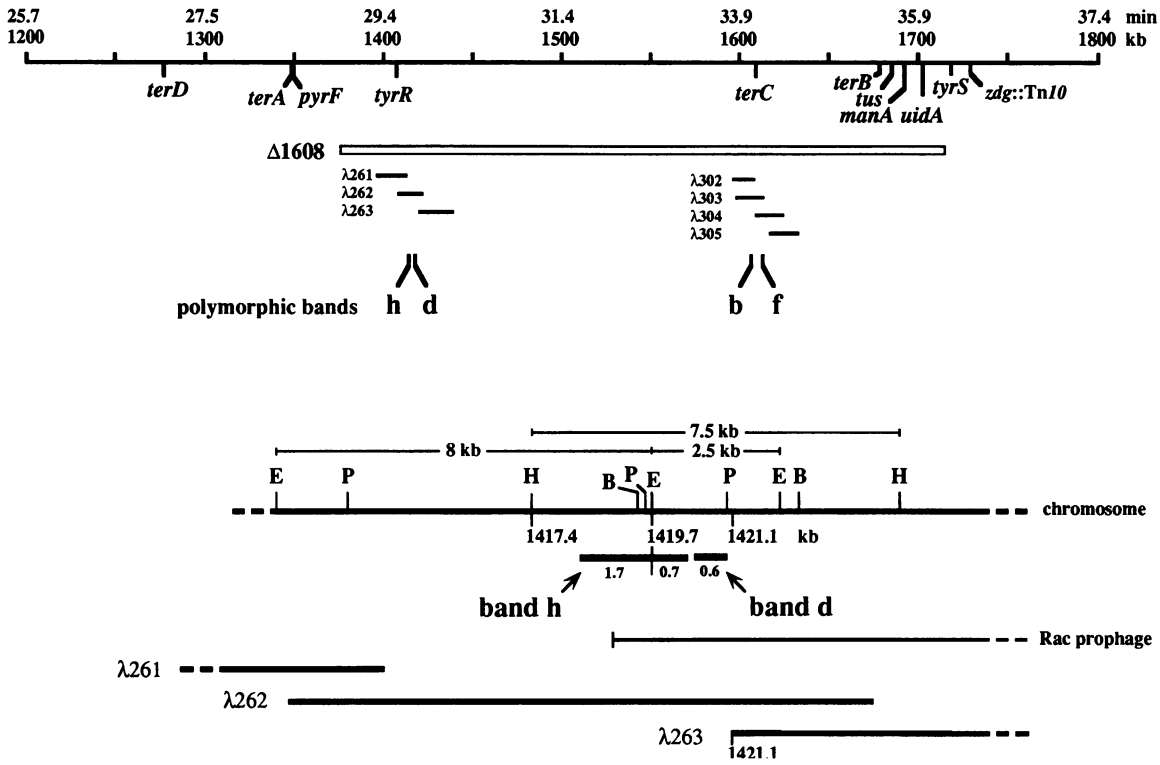

FIG. 2. Map of the region containing the terminus of chromosome replication (adapted from references 2, 4, 9, 13, 16, and 24 and this work). The positions of the four polymorphic RAPD bands (b, d, f, and h) that were fine structure mapped by hybridization to genomic Southern blots and to phage clones are indicated; a fifth polymorphic band (e) was inferred to come from this region because Δ1608 strains are E^-, but it was not cloned and thus was not mapped with precision.

more reproducible than those obtained by simply boiling colonies from plates or aliquots of liquid cultures (data not shown).

RAPD fingerprinting. PCR was carried out in 25 μl containing 30 ng of *E. coli* genomic DNA, 3 mM $MgCl_2$, 20 pmol of the primer (primer sequences are shown in Table 2), 1 U of AmpliTaq DNA polymerase (Perkin-Elmer Cetus), 250 μM each dCTP, dGTP, dATP, and dTTP (Boehringer Mannheim) in 10 mM Tris HCl (pH 8.3), 50 mM KCl, and 0.001% gelatin, all under a drop of mineral oil. A Perkin-Elmer TC480 thermal cycler was used for 45 cycles of amplification (94°C for 1 min, 36°C for 1 min, 72°C for 2 min) and then 72°C for 10 min in RAPDs with 15- as well as 10-nucleotide-long primers. After PCR, 8-μl aliquots were electrophoresed in 2% agarose gels containing 0.5 μg of ethidium bromide per ml and 1× Tris acetate running buffer and photographed under UV light. The 1-kb DNA ladder (Gibco BRL) was used as a size marker in all gels.

Reproducibility of RAPD fingerprints. The reproducibility of the sometimes small differences in band patterns that distinguish closely related strains was tested by using DNAs made from separate cultures of the same strain. Representative results are shown in Fig. 3A; equivalent assays with duplicate DNA preparations were used to test each polymorphism described below and in each case verified its reproducibility (data not shown).

The effect of concentration of AmpliTaq DNA polymerase on RAPD profiles was tested under conditions optimized for other critical parameters of Mg^{2+}, primer, and deoxynucleoside triphosphate concentration (1). Figure 3B shows that the amplification of certain bands was much more sensitive than

TABLE 2. Primers used for RAPD analysis[a]

Primer	Sequence	No. of bands	
		Total	Polymorphic[b]
1247	AAGAGCCCGT	16	0
1254	CCGCAGCCAA	13	0
1281	AACGCGCAAC	11	0
1283	GCGATCCCCA	12	1 (a)
1290	GTGGATGCGA	13	3 (b,c,d)
OPA-05	AGGGGTCTTG	12	0
OPA-10	GTGATCGCAG	8	3 (e, f,g)
OPB-08	GTCCACACGG	10	0
OPB-20	GGACCCTTAC	8	0
D-14801	AGCGAAAACGGTTGG	18	0
D-14802	GAAAACGGTTGGGTG	15	0
D-14803	AAACGGTTGGGTGAG	15	1 (h)

[a] The 12 primers used were chosen from among 30 primers tested as giving the largest number of reproducible bands with *E. coli* K-12 strains (1, 12). Primers 1247, 1254, 1281, 1283, and 1290 had been used by us in RAPD analyses of other organisms (1, 12). The OPA and OPB primers were from Operon Technologies Inc. (Alameda, Calif.). Primers D-14801, D-14802, and D-14803, although 15 nucleotides long, were used with the same cycling protocol used for 10-nucleotide primers, on the basis of reconstruction experiments showing that this protocol resulted in profiles that were similar to and as reproducible as those obtained with the variant protocol originally recommended (26) for longer primers.

[b] For polymorphisms, see Fig. 4. The inferences that particular bands are specific to F (primer 1283) or to the chromosome (primers 1290, OPA-10, and D-14803) are based on tests of C600 and its F^+ and Δ1608 derivatives that were generated for this study (see Fig. 3 and 6).

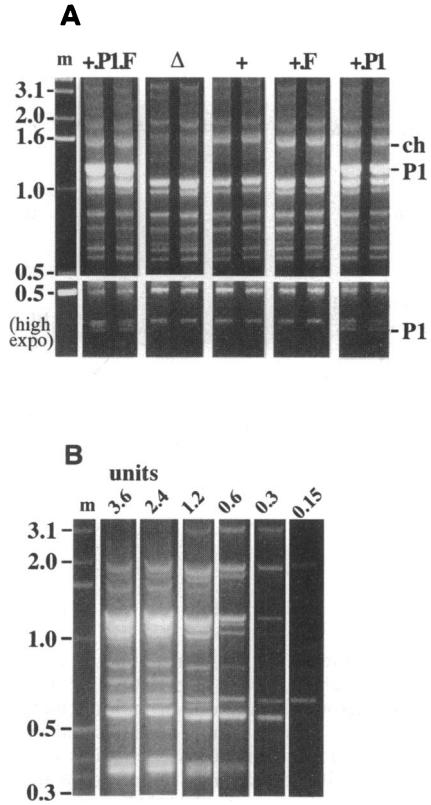

FIG. 3. Test of reproducibility of arrays of RAPD profiles. (A) Reproducibility of profiles from separate DNA preparations and identification of plasmid versus chromosomal (ch) $\Delta 1608^+$ segments as sources of polymorphic bands. m, size markers (1-kb ladder [Gibco BRL]; sizes in kilobases are shown at the left). The left and right lanes in each pair contain products of RAPD amplification using primer 1254 and DNAs extracted on different days from different cultures of the same strain. Designations: +, wild-type chromosome; Δ, Δ1608; F, presence of F-factor plasmid; P1, presence of P1 prophage; high expo, a segment of the gel printed at higher exposure in order to show the faint but reproducible polymorphic band at ~0.4 kb. The strains used were D7599 (+, P1, F), DB7591 (Δ), C600 (+), DB7593 (+F), and DB7598 (+P1). All lanes shown were from a single gel. This figure shows that the 1.5-kb band is specific to $\Delta 1608^+$ and the 0.4-kb band is specific to P1. (B) Effect of amount of polymerase on arrays of RAPD products. The DNA was from strain DB7599 (C600 carrying F and P1), the primer used was 1254, and the AmpliTaq polymerase concentration was 2.5 U/μl.

the amplification of others to changes in amounts of DNA polymerase. In our hands, 1.2 U (with this batch, 0.5 μl) of enzyme per 25-μl reaction seemed to be optimal, with disproportionately lower yields of certain bands when less enzyme was used. The yield of large (≥3-kb) bands also decreased when more enzyme was used, perhaps because of inhibition of amplification by a component of the storage buffer. Thus, differences in polymerase activity may underly the small differences in RAPD profiles that are seen (despite optimization of other more easily controlled parameters) in reactions carried out at different times or in different laboratories. We have sometimes used this differential sensitivity to DNA polymerase concentration to resolve two comigrating bands.

Nomenclature. RAPD products that were produced from some but not all strains will be referred to as polymorphic. The eight RAPD bands that were found to be polymorphic among historic K-12 strains will be referred to as bands a through h. In the text, the ability or inability to produce a given band is often indicated by using the band designation and superscript + or −. Thus, strain CR34Thy$^-$ is B$^-$, whereas each of the other 25 historic strains is B$^+$ (Fig. 1).

Hybridization analyses. For use as probes, polymorphic RAPD bands were first cloned into the pBluescript vector plasmid (Stratagene, La Jolla, Calif.) to ensure that only the RAPD product of interest was labeled. The cloning entailed cleavage of pBluescript at its *Eco*RV site, addition of single T residues to the blunt ends with *Taq* polymerase to facilitate PCR product cloning (17), elution of the desired RAPD bands from gels, ligation with vector DNA, and transformation of *E. coli* DH5α by standard methods (23). For use as probes, these cloned fragments were cleaved from plasmid clones at unique sites bracketing the *Eco*RV cloning site (*Eco*RI and *Hin*dIII for bands b, d, and f; *Sma*I and *Hin*dIII for band h). The cloned DNAs were then purified by electrophoresis and elution from gel slices and labeled by using the random primer kit and [^{32}P]dCTP as recommended by the supplier (Stratagene).

For Southern blots, DNA samples to be probed were electrophoresed in 1 or 2% agarose gels (restriction endonuclease-digested genomic DNA and RAPD products, respectively). The gels were treated with 2 volumes of 0.25 M HCl for 10 min (for partial depurination), soaked in a denaturation solution (1.5 M NaCl, 0.5 M NaOH) for 60 min, and then soaked in a neutralization solution (1.5 M NaCl, 1 M Tris HCl [pH 8.0]) for 60 min. The DNA was transferred to N-Hybond (Amersham) nylon filters by capillary blotting and fixed to the filters by UV irradiation (23). Filters containing RAPD products, genomic restriction fragments, or Kohara phage clones (13) (Takara Biochemicals, Berkeley, Calif.) were prehybridized at 42°C in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–5× Denhardt solution–0.5% SDS–100 mg of salmon sperm DNA per ml–50% formamide for 4 h and then hybridized with the radioactive probe under the same conditions for 12 h. They were washed in 2× SSC–0.5% SDS for 5 min at 20°C, in 2× SSC–0.1% SDS for 15 min at 20°C, and in 0.1× SSC–0.5% SDS for 2 h and then for 30 min at 68°C. To reuse filters, labeled probes were removed by washing the filters in 0.4 M NaOH for 1 h at 45°C, in 0.2 M Tris HCl (pH 8.0)–0.1× SSC–0.1% SDS for 30 min at 45°C, and in 0.1× SSC–0.1% SDS for 1 h at 52°C. The completeness of stripping of probes was verified by autoradiography.

Figure preparation. Polaroid gel photographs and autoradiograms were processed for presentation by being scanned with a Microtek Scanmaker 600ZS, using Adobe Photoshop 2.0 software (Adobe Systems, Inc., Mountain View, Calif.), and imported into Canvas 3.0.2 (Deneba Software, Miami, Fla.) running on a Macintosh Quadra 700 computer.

RESULTS

Arrays of RAPD products from historic *E. coli* K-12 strains. RAPD tests were carried out on a set of 26 strains from divergent branches of the *E. coli* K-12 pedigree (Fig. 1), most of which had been isolated in the 1940s and 1950s (Table 1). All strains tested were clonal descendants of the original wild-type *E. coli* K-12, the products of mutation and selection, without DNA transfer from other strains. Twelve primers (Table 2) that had yielded informative profiles in preliminary tests were used, and the reactions were carried out by using phenol-extracted DNAs and concentrations of thermostable polymerase (AmpliTaq) and other components that had been optimized for quality and reproducibility of profiles (Fig. 3).

Eight of 150 RAPD bands scored with 12 primers (Table 2) were polymorphic, being present in the profiles from some

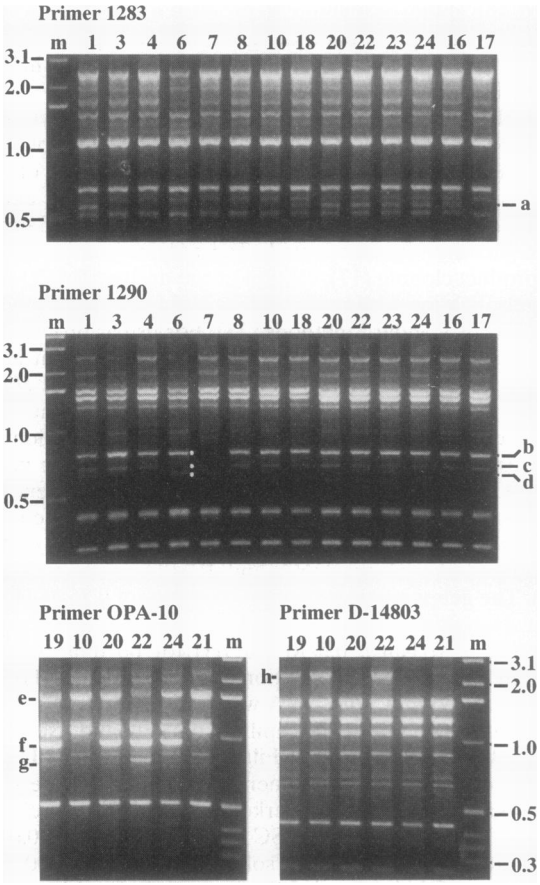

FIG. 4. Representative RAPD profiles from historic strains from the *E. coli* K-12 pedigree. Numbers above the lanes correspond to strain numbers used in Fig. 1. m, size markers (as in Fig. 3). The sequences of primers are listed in Table 2. All polymorphic bands detected in studying these strains (a through h) are indicated.

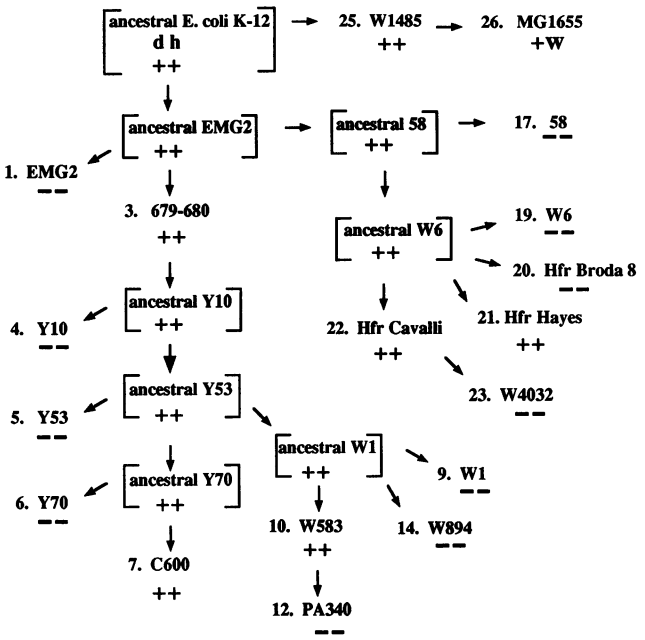

FIG. 5. Inferred patterns of change in historic *E. coli* K-12 strains. All known or inferred changes with respect to RAPD bands d and h are indicated. Each arrow indicates one or multiple mutational steps. Current versions of the historic strains are indicated by numbers (1 through 26) as in Fig. 1; versions of historic strains inferred to have been used to generate the next strain in the pedigree are indicated in brackets. The inference that ancestral D^+H^+ strains were D^+H^+ is based on the occurrence of D^+H^+ progeny.

strains but not others (Fig. 1 and 4). Thirty-seven changes in RAPD profile were seen among the 26 historic strains, whose lineages contained 59 separate steps of mutation and deliberate selection. The most divergent RAPD profiles, those from Hfr Broda 8 and Hfr Hayes, differed from each other by six bands, although these strains were separated by only eight known steps.

Several important features emerged from comparisons of the bands amplified from each strain with respect to the position of that strain in the pedigree. With four of eight polymorphic bands (b, c, e, and f), the loss of ability to produce the band was correlated with the pedigree: the ability to form bands b, c, and e was lost just once and not regained (strains CR34Thy$^-$, C600, and Hfr Hayes, respectively); the ability to form band f, although lost multiple times (C600, P678, AB1157, Hfr Broda 8, and W1485), was also not regained. The ability to form a fifth band (g) appeared only once (in Hfr Hayes); such a single appearance is also not unexpected.

The distribution of the ability to form three other bands (a, d, and h) did not correlate well with the pedigree: band a was amplified from W1485 but not from its parent, WG1, whereas the ability to produce this band was lost in three other branches of the pedigree (Y10 and descendants, Hfr Broda 8, and MG1655).

More complex patterns were observed with two other bands, d and h. They were not amplified from five strains whose immediate descendants did yield these bands: (i) EMG2 (D^-H^-) versus 679 (D^+H^+); (ii) 58 (D^-H^-) versus 58-161 (D^+H^+); (iii) Y70 (D^-H^-) versus C600 (D^+H^+); (iv) W1 (D^-H^-) versus W583 (D^+H^+); and (v) W6 (D^-H^-) versus Hfr Cavalli (D^+H^+). The ability to amplify band h was also diminished but not eliminated in MG1655, but its D^+ character was not affected (Southern blotting confirmed that this weakly amplified band contained h sequences [Fig. 1]).

The lack of correlation of these patterns with the strain pedigree suggests numerous cases of mutation from D^+H^+ to D^-H^- in a strain after its derivative had been isolated (Fig. 5). This interpretation implies separate cycles of mutation leading to the current versions of strains such as Y10 and Y53 (each D^-H^-), since they were nominally ancestors of D^+H^+ strains W583 and C600. The D^-H^- strains PA340, W894, and Hfr Broda 8 also descend from D^+H^+ ancestors, but it is not clear whether the ability to amplify bands d and h was lost before, during, or after the isolation of each strain. On the basis of these considerations, we estimate that there were 11 separate losses of ability to amplify bands d and h (Fig. 5). Although such results could also suggest an alternative model in which cells repeatedly mutate from D^-H^- to D^+H^+, the finding that band d- and h-containing sequences are deleted from several D^-H^- strains (see below) rules out this alternative.

In summary, the versions of many historic strains available today may reflect continued divergence after their initial isolation. Many of the parent-progeny pairs that, on the basis of the pedigree, should differ by just a single mutation may actually be considerably less isogenic.

Tests of isogenic strains differing in plasmid content and chromosome size. Isogenic derivatives of strain C600 that

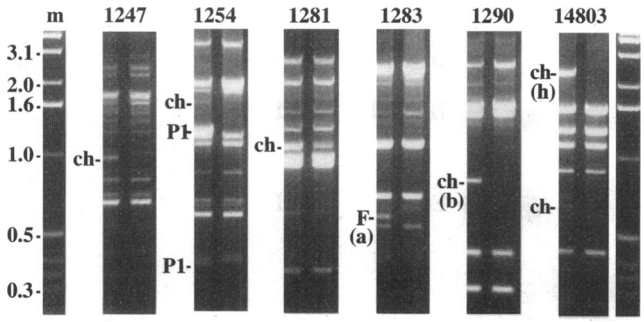

FIG. 6. Identification of polymorphic RAPD bands in strains that differ in plasmid content and chromosome size. With each pair of lanes, the left lane contains RAPD products from C600 carrying F and P1 (DB7599) and the right lane contains products from plasmid-free C600Δ1608 (DB7591). Designations: m, size markers (as in Fig. 3); 1247 through 14803, primers used for amplification (sequences in Table 2); ch, F, and P1, bands specific to the 340-kb segment removed by Δ1608, to the ~100-kb F plasmid, and to the ~100-kb P1 plasmid. The mapping of bands to Δ1608, F, or P1 was based on tests using isogenic strains that carry the Δ1608 or Δ1608$^+$ allele and were either plasmid free or carried just one of the plasmids (see, e.g., Fig. 3A). All lanes were from one gel; the gel photograph was cut and pasted for clarity of labeling key bands.

differed in the presence or absence of a ~340-kb deletion of the chromosome replication terminus, Δ1608 (Fig. 2), and also the F and P1 plasmids (each ~100 kb), were constructed and studied by RAPD fingerprinting. Nine of 78 prominent bands scored from strain C600 carrying plasmids P1 and F were missing from the profile of plasmid-free C600Δ1680 (Fig. 6). This outcome is in accord with Δ1608$^+$, F, and P1 collectively constituting about 11% of the *E. coli* genome (9, 13, 20, 29). Tests using other isogenic strains from this series and an equivalent series that we generated starting with strain Y10 (as in Fig. 3A) showed that one band came from the F plasmid and corresponded to band a from the historic strains; two came from P1, and six came from the Δ1608$^+$ segment. Two bands eliminated by the Δ1608 deletion in these experiments corresponded to bands b and h that were polymorphic among the historic strains. Band d, which was amplified faintly with primer 1290 in tests of the historic strains (Fig. 4), was not seen in these experiments, probably because a different batch of AmpliTaq DNA polymerase was used (see Fig. 3B).

Additional tests comparing Y10 and C600, and isogenic Δ1608 strains derived from them, indicated that five of the six chromosomal RAPD bands that were polymorphic among these historic strains came from the Δ1608$^+$ segment (d, e, and f as well as b and h) (data not shown). This finding suggested a remarkable tendency for mutations to accumulate in the terminus region of the chromosome.

Additional noteworthy features of the distribution of the ability to amplify band a became evident when it was localized to the F-factor plasmid (above). Band a was amplified from 9 of 11 historic strains tested that carried the F factor in either plasmid or integrated form and was not amplified from 14 of 15 strains tested that lacked the F factor. Surprisingly, band a was not amplified from one Hfr strain (Hfr Broda 8) and one F$^+$ strain (WG1), and it was amplified from one F$^-$ strain (679–680). The A$^-$ strain WG1 contains a deletion variant of the F factor (~40 kb missing) (24) that does not confer M13 sensitivity. Control experiments showed that the isolates of Hfr Broda 8 and 679-680 used in these experiments were sensitive and resistant, respectively, to the F-factor-specific phage M13 (as expected).

Strains that carried an F factor with a ~50-kb deletion, pOX38 (8), also did not yield band a. The failure of Hfr Broda 8 to yield band a can thus be ascribed to a deletion or other mutation in its F factor; the amplification of band a from the nominally F$^-$ strain 679-680 might reflect persistence of part of the F factor in this strain.

Genomic Southern hybridization analysis of RAPD polymorphisms. The DNAs of RAPD bands b, d, f, and h were cloned in the pBluescript plasmid vector, and the cloned DNAs were used as hybridization probes to better understand how the ability to amplify these bands was lost and to map the genomic loci from which they came. The seeming disappearance and then reappearance of the ability to amplify bands d and h was studied first, by probing Southern blots of genomic DNAs from representative strains with the cloned band d and band h DNAs (Fig. 7). Each of eight strains that yielded band d (designated D$^+$) contained a 2.5-kb *Eco*RI fragment with homology to band d, whereas three of four historic strains that were D$^-$ (all except strain Y10) lacked genomic sequences that hybridized to band d.

A consistent but more detailed picture emerged from hybridization to genomic *Hin*dIII digests. Two fragments with strong homology to band d were found: (i) a 7.5-kb fragment in seven of eight D$^+$ strains (all but MG1655; number 26 in Fig. 7), which was absent from D$^-$ strains; and (ii) a ≥14-kb fragment in one D$^-$ strain (Y10; number 4) and in three D$^+$ strains (679 [number 2]; W583 [number 10], and MG1655 [number 26]). (Strains 679 and W583 exhibited both the 7.5- and the ≥14-kb fragments.)

The same genomic DNAs were tested by using the band h probe. In *Hin*dIII digests, all H$^+$ strains exhibited a 7.5-kb fragment with homology to the h probe and also one or more fragments of ≥14 kb in size. H$^-$ strains lacked the characteristic 7.5-kb *Hin*dIII fragment, as did MG1655 (which allowed only weak amplification of band h) and C600Δ1608 (DB7591), which lacks the normal terminus of replication. Consistent results were obtained with *Eco*RI digests: most H$^+$ strains exhibited fragments of 2.5 and 8 kb with homology to band h, whereas most H$^-$ strains did not (although other fragments with weak homology were present). In summary, these experiments showed that the loss of ability to produce bands d and h was associated with deletion of corresponding chromosomal sequences in three of four cases. The fourth such case (Y10), and also the case of decreased amplification of band h (MG1655), seemed to be associated with rearrangements of the corresponding sequences, not deletions.

The filters used above were stripped of labeled DNA and reprobed with cloned band b and f DNAs. Much simpler patterns were obtained. In the case of *Eco*RI digests, the band b probe hybridized to one fragment of >23 kb from all strains used in the tests shown in Fig. 6 and also from CR34Thy$^-$, the only B$^-$ strain that was found (data not shown). In the case of the *Hin*dIII digests, the band b probe also hybridized to just one fragment of about 5.3 kb from all strains except CR34Thy$^-$, which exhibited two fragments of 1.5 and 3.8 kb with band b homology. The CR34Thy$^-$ pattern probably reflects a mutation that created an internal *Hin*dIII site and is not related to the inability to amplify band b. It did not hybridize to any sequences in the Δ1608 derivative of C600.

The cloned band f DNA was used to probe the same filters that had been used for hybridization with band d and h probes in the analysis described above. The band f probe hybridized to just one *Eco*RI fragment and one *Hin*dIII fragment, each ≥23 kb, from all historic strains tested, regardless of whether they yielded band f (679, Y10, W1, W583, 58, 58-161, W6, Hfr Hayes, and Hfr Cavalli) or did not (C600, CR34Thy$^-$, W1485,

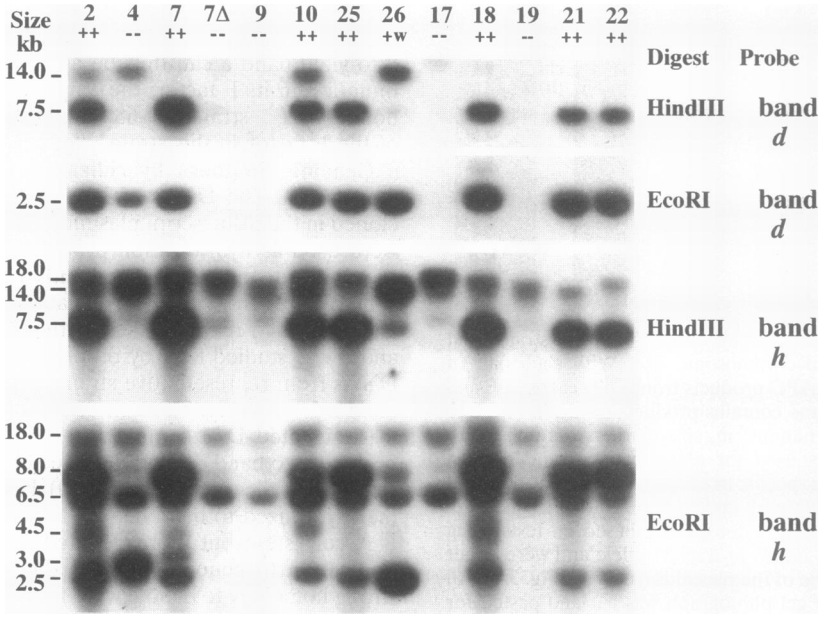

FIG. 7. Southern blot hybridization with band d and h probes. 7Δ indicates the Δ1608 deletion derivative of C600 (DB7591). The numbers above all other lanes refer to strains listed in Fig. 1; the series of + and − signs above each lane indicates the ability and inability, respectively, to yield RAPD bands d and h, respectively (see Fig. 1).

and MG1655) (data not shown). It did not hybridize to any sequences in the Δ1608 derivative of C600.

These results suggest that point mutations, rather than rearrangements, were responsible for the loss of ability to amplify band f in five different lineages.

Hybridization to arrays of RAPD products. Southern blots of arrays of RAPD fragments were probed with cloned band d and h DNAs. These tests showed that (i) these two bands were not homologous to one another, even though the ability or inability to amplify one band was usually associated with the ability or inability to amplify the other; (ii) each probe hybridized strongly to the band from which it was cloned and also to several smaller fragments in arrays from D^+ strains; and (iii) arrays from D^- strains did not exhibit hybridization with the band d probe, whereas arrays from H^- strains (including C600Δ1608) contained smaller bands that hybridized weakly to the h probe (data not shown). This finding indicates that additional band h related sequences are present elsewhere in the chromosome.

Map locations of DNA segments yielding polymorphic RAPD products. The chromosomal segments corresponding to bands b, d, f, and h were located more precisely by probing a filter containing the Kohara library of ordered λ clones that cover most of the *E. coli* genome (13). The band d and h probes each hybridized strongly to clone λ262 (4A6) in the vicinity of min 30.3 in the chromosome but not to clones λ261 (3G3) and λ263 (2C3), which overlap λ262 (data summarized in Fig. 2). Inspection of the restriction map of this region (13, 22) (Fig. 2) revealed a segment unique to clone λ262 that is within a 7.5-kb *Hin*dIII fragment and that partially overlaps adjacent 8.1- and 2.5-kb *Eco*RI fragments. Thus, band d comes from a sequence between kb 1420.4 and 1421.1 (~0.7 kb to the right of the *Eco*RI site at kb 1419.7 and at the left end of clone λ263, respectively). Band h comes from a sequence just to the left of band d, at a position overlapping the *Eco*RI site at kb 1419.7. This region contains one end of the defective Rac prophage, which is partially related to phage λ and is believed to have been excised from some *E. coli* K-12 strains (11, 15, 16). The band h probe also hybridized weakly to pairs of overlapping clones from two other loci: λ428 (2D5) and λ429 (7F8) at min 56.8 and λ572 (2A6) and λ573 (4D2) at min 82. Band d, in contrast, did not hybridize to other clones in the Kohara library (data not shown). Thus, the sequences in band d are unique, whereas some of those in band h are repeated.

The cloned band b DNA hybridized strongly to clones λ302 (2A9) and λ303 (5F9) (estimated position, kb 1603.7 to 1605.1; min 34.4), and band f hybridized strongly to clones λ304 (1F9) and λ305 (2H2) (kb 1623.8 to 1625.9; min 34.8) (data not shown). Thus, bands b and f come from loci about 180 and 200 kb from the sequences in or near the end of the Rac prophage that yield bands d and h (Fig. 2).

DISCUSSION

Strain-specific, reproducible differences were found in DNA fingerprints generated from historic strains of *E. coli* K-12 by the arbitrary primer PCR (or RAPD) method. Eight of 150 DNA bands (RAPD products) scored in tests of 26 strains from several branches of the pedigree were polymorphic. Analyses of these bands suggest (i) that additional mutations, distinct from those deliberately selected in generating new strains, have accumulated and (ii) that such mutations tend to be near the terminus of chromosome replication.

The polymorphic bands d and h provided the most striking demonstration of such nondeliberate changes. Paradoxically, the ability to amplify these bands (designated D^+H^+) disappeared independently from 11 different historic strains in the *E. coli* K-12 pedigree yet seemed to reappear in six different sublineages derived from nominally D^-H^- strains (Fig. 1). These findings implicated genomic changes in parental strains after progeny strains had been isolated, that is, during storage or subculturing (Fig. 5). Since bands d and h came from within or near the Rac prophage, and the D^-H^- character was attributable to deletion (Rac excision) in many of the strains, it

could be argued that they represent a special case. However, (i) other rearrangements were implicated in changes affecting band d in one case (Y10) and band h in two cases (Y10 and MG1655); (ii) an equivalent incomplete correlation with the pedigree was seen with band a (F factor specific); and (iii) three other polymorphic RAPD fragments came from the terminus region (but outside the Rac prophage), whereas none of the mutations deliberately selected by researchers map to this region. Thus, we propose that there are many sequence differences among historic strains of *E. coli* K-12, distinct from those changes deliberately selected in obtaining new phenotypes.

It is remarkable that five of the six polymorphic chromosomal bands that were mapped came from just 7% of the chromosome. This 340-kb segment, defined by the Δ1608 deletion, contains the normal terminus of chromosome replication (9). Comparison of restriction maps of several historic strains also revealed a tendency of sequence changes to cluster in this region (18). Homologous recombination close to the terminus is also extremely frequent (14) and perhaps it is related to the apparent high local mutation frequency.

What could account for the remarkable clustering of mutations near the terminus of DNA replication? We note that most strains studied here were isolated more than 30 years ago. It was common practice in those days to store *E. coli* strains as slowly growing stab cultures at room temperature (6). The selective pressures operating under these conditions must differ markedly from those usually studied, and thus the accumulation of inadvertent mutations might be ascribed to selection. The relative lack of genes in the terminus region that are known to contribute significantly to the phenotype as usually studied (4, 9) may argue against a selection model, but this possibility has not been tested critically, e.g., by measuring bacterial fitness in stab cultures.

In an alternative model, the actual probability of a mutational event might be highest near the terminus, perhaps as a result of changes in a replication complex stalled on incompletely replicated chromosomes. If this effect operated during growth, it might reflect a local inhibition of replication by the binding of Tus (terminator) protein to its cognate recognition sites (*ter* in Fig. 2) (10). Alternatively, perhaps some replication complexes remain stalled for months or years in cells persisting in long-lived stab cultures. Any changes in key fidelity components of the replication complex might lead to mutation, either during this persistent state or upon resumption of growth. The observed clustering of mutations near the replication terminus could be explained if replication forks tended to accumulate near this region. Alternatively, perhaps only a subset of incompletely replicated chromosomes, specifically those needing the least DNA synthesis to complete their cycles of DNA replication, survive long periods of storage.

ACKNOWLEDGMENTS

We thank B. Bachmann for providing many strains from the *E. coli* Genetics Stock Center strain collection and for suggestions and pedigree information, P. Kuempel for strain PK1647, and N. Akopyanz, C. M. Berg, N. Bukanov, D. Kersulyte, P. Kuempel, F. La Banca, J. M. Louarn, N. Murray, S. Short, and J. P. Woods for unpublished data, stimulating discussion, and/or critical reading of the manuscript.

This work was supported by Public Health Service grant HG00563.

REFERENCES

1. **Akopyanz, N., N. O. Bukanov, T. U. Westblom, S. Kresovich, and D. E. Berg.** 1992. DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. Nucleic Acids Res. **20**:5137–5142.
2. **Bachmann, B. J.** 1972. Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol. Rev. **36**:525–557.
3. **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. American Society for Microbiology, Washington, D.C.
4. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
5. **Bachmann, B. J.** Personal communication.
6. **Clowes, R. C., and W. Hayes.** 1968. Experiments in microbial genetics. Blackwell Scientific Publications, Oxford.
7. **Gray, C. H., and E. L. Tatum.** 1944. X-ray induced growth factor requirements in bacteria. Proc. Natl. Acad. Sci. USA **30**:404–410.
8. **Guyer, M. S., R. R. Reed, J. A. Steitz, and K. B. Low.** 1981. Identification of a sex-factor-affinity site in *E. coli* as γδ. Cold Spring Harbor Symp. Quant. Biol. **45**:135–140.
9. **Henson, J. M., and P. L. Kuempel.** 1985. Deletion of the terminus region (340 kilobase pairs of DNA) from the chromosome of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **82**:3766–3770.
10. **Hill, T. M., M. L. Tecklenburg, A. J. Pelletier, and P. L. Kuempel.** 1989. *Tus*, the trans-acting gene required for termination of DNA replication in *Escherichia coli*, encodes a DNA-binding protein. Proc. Natl. Acad. Sci. USA **86**:1593–1597.
11. **Kaiser, K., and N. E. Murray.** 1979. Physical characterisation of the "Rac prophage" in *E. coli* K12. Mol. Gen. Genet. **175**:159–174.
12. **Kersulyte, D., J. P. Woods, E. J. Keath, W. E. Goldman, and D. E. Berg.** 1992. Diversity among clinical isolates of *Histoplasma capsulatum* detected by polymerase chain reaction with arbitrary primers. J. Bacteriol. **174**:7075–7079.
13. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
14. **Louarn, J.-M., J. Louarn, V. Francois, and J. Patte.** 1991. Analysis and possible role of the hyperrecombination in the termination region of the *Escherichia coli* chromosome. J. Bacteriol. **173**:5097–5104.
15. **Low, B.** 1973. Restoration by the *rac* locus of recombinant forming ability in *recB*$^-$ and *recC*$^-$ merozygotes of *Escherichia coli* K-12. Mol. Gen. Genet. **122**:119–130.
16. **Mahajan, S. K., C. C. Chu, D. K. Willis, A. Templin, and A. J. Clark.** 1990. Physical analysis of spontaneous and mutagen-induced mutants of *Escherichia coli* K-12 expressing DNA exonuclease VIII activity. Genetics **125**:261–273.
17. **Marchuk, D., M. Drumm, A. Saulino, and F. S. Collins.** 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. **19**:1154.
18. **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.
19. **Persing, D. H., T. F. Smith, F. C. Tenover, and T. J. White (ed.).** 1993. Diagnostic molecular microbiology. American Society for Microbiology, Washington, D.C.
20. **Pierce, J. C., and N. L. Sternberg.** 1992. Using bacteriophage P1 system to clone high molecular weight genomic DNA. Methods Enzymol. **216**:549–574.
21. **Rosner, J. L.** 1972. Formation, induction and curing of bacteriophage P1 lysogens. Virology **48**:679–689.
22. **Rudd, K. E.** Personal communication.
23. **Sambrook, J., E. F. Fristch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Short, S.** Personal communication.
25. **Vogel, H. J., and D. M. Bonner.** 1956. Acetylornithinase of *Escherichia coli*: partial purification and some properties. J. Biol. Chem. **218**:97–106.
26. **Wang, G., T. Whittam, C. M. Berg, and D. E. Berg.** 1993. RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains. Nucleic Acids Res. **21**:5930–5933.

27. **Welsh, J., and M. McClelland.** 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. **18:**7213–7218.
28. **Welsh, J., and M. McClelland.** 1993. Characterization of pathogenic microorganisms by genomic fingerprinting using arbitrary primed PCR, p. 595–602. *In* D. H. Persing, T. F. Smith, F. C. Tenover, and T. J. White (ed.), 1993. Diagnostic molecular microbiology. American Society for Microbiology, Washington, D.C.
29. **Willetts, N., and R. Skurray.** 1987. Structure and function of the F factor and mechanisms of conjugation, p. 1110–1133. *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. American Society for Microbiology, Washington, D.C.
30. **Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18:**6531–6535.