JOURNAL OF BACTERIOLOGY, Nov. 1990, p. 6175–6181
0021-9193/90/116175-07$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 172, No. 11

Phylogenetic Distribution of Branched RNA-Linked Multicopy Single-Stranded DNA among Natural Isolates of *Escherichia coli*

PETER J. HERZER,[1] SUMIKO INOUYE,[1] MASAYORI INOUYE,[1] AND THOMAS S. WHITTAM[2]*

Department of Biochemistry, Robert Wood Johnson Medical School–University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854,[1] and Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802[2]

Received 11 May 1990/Accepted 8 August 1990

Multicopy single-stranded DNA (msDNA), a branched DNA-RNA molecule, has been shown in *Escherichia coli* B and clinical strain Cl-1 to be synthesized by reverse transcriptase. We report that 13% of the strains of the ECOR collection, a sample of 72 *E. coli* isolates representing the breadth of genetic variation of the species, produce msDNA. Three of the four major subspecific groups include msDNA-producing strains. Screening of 25 isolates that are genetically related to msDNA-producing clinical strains uncovered 22 additional msDNA-producing strains. A phylogenetic tree based on allelic variation detected electrophoretically at 20 enzyme-encoding loci revealed two major clusters and several deep branches composed of strains that synthesize msDNA. Although *E. coli* K-12 does not harbor msDNA, other closely related strains of the K-12 family do. The results support the hypothesis that msDNA-synthesizing systems, including reverse transcriptase genes, were acquired recently and independently in different lineages of *E. coli*.

The recent discovery of bacterial reverse transcriptases (RTs) was engendered by the observation of a peculiar satellite DNA, now called multicopy single-stranded DNA (msDNA), in the soil bacterium *Myxococcus xanthus* (26). The satellite (msDNA-Mx162) is unusual because it consists of a 162-base single DNA strand linked to a 77-base RNA (msdRNA) sequence by a 2′,5′-phosphodiester linkage and occurs in hundreds of copies per cell (4). An msDNA of a similar structure (msDNA-Sa163) has been found in the closely related myxobacterium *Stigmatella aurantiaca* (5, 6). Recently it has been reported that *M. xanthus* contains another species of msDNA (msDNA-Mx65) in addition to msDNA-Mx162 (3). Most recently, msDNA molecules with structural features similar to those of myxobacterial msDNAs have been found in clinical isolate Cl-1 (9), strain B (10), and several other clinical isolates (21) of *Escherichia coli*.

In *M. xanthus* (8), *E. coli* clinical isolate Cl-1 (9), and *E. coli* B (10), the synthesis of msDNA depends on RT encoded by a sequence adjacent to the chromosomal region specifying msdRNA. Although the three bacterial RTs are substantially divergent in size and amino acid sequence, as inferred from the nucleotide sequences, they exhibit the common amino acid motifs of retroviral reverse transcriptases (22).

The objective of the study reported here was to determine the frequency of occurrence and genetic relationships among strains that harbor msDNA synthesized by RT in natural populations of *E. coli*. To accomplish this, we examined the 72 strains of the ECOR collection (13), a sample of natural isolates chosen to represent the major subspecific groups of the *E. coli* species as a whole. These groups have been identified on the basis of allelic variation at enzyme-encoding genes detected by multilocus enzyme electrophoresis (14, 23) and include isolates recovered from a variety of human and animal hosts in diverse geographic areas (13). In addition, we characterized the multilocus genotypes of previously reported msDNA-producing strains, including clinical strains (21) and *E. coli* B (10), to determine their overall genetic relatedness to other *E. coli* strains and screened close relatives of these strains for msDNA production.

MATERIALS AND METHODS

Bacterial strains. The 72 strains of the ECOR collection were provided by R. K. Selander. Details of the original sources for the ECOR strains are given by Ochman and Selander (13). Twelve msDNA-containing clinical strains were isolated from patients with extraintestinal infections at the Robert Wood Johnson Hospital (New Brunswick, N.J.). Twenty-five additional strains were obtained from I. K. Wachsmuth, Centers for Disease Control, Atlanta, Ga.; F. and I. Ørskov, Statens Seruminstitut, Copenhagen, Denmark; and R. A. Wilson, *E. coli* Reference Center, Department of Veterinary Sciences, Pennsylvania State University, University Park, Pa. Twenty-one of the strains for which somatic (O) and flagellar (H) serotypes were not available were serotyped with standard methods (25) by R. A. Wilson.

RT-labeling method. To detect msDNA, total RNA from bacterial cultures, prepared by the method described by Chomzynski and Sacchi (2), was treated with avian myeloblastosis virus RT in the presence of [α-^{32}P]dCTP (9, 21). The extension of the 3′ end of the msDNA, with the RNA molecule serving as template for RT, resulted in ^{32}P-labeled DNA fragments, which were electrophoresed in a 4% acrylamide–8 M urea gel and compared with molecular size markers. To determine the size of the single-stranded DNA, samples were treated with RNase A before electrophoresis (9).

Multilocus enzyme electrophoresis. Bacterial strains were characterized by enzyme variants separated in 11% starch gels by the electrophoretic methods and selective enzyme assays described by Selander et al. (17). A total of 20 different enzymes were analyzed, including the 19 listed in reference 24 and glutamate dehydrogenase. Mobility variants (electromorphs) of an enzyme were identified as distinctly migrating protein bands and were equated with alleles at the corresponding structural gene locus. The ECOR

* Corresponding author.

strains have been characterized previously for electrophoretically detected allelic variants for 38 enzymes (7, 17).

Phylogenetic analysis. The phylogenetic relationships among strains were inferred from the analysis of the genetic distance matrix by the neighbor-joining (NJ) method (16). The distance matrix consisted of the standard genetic distances (12) between all pairs of electrophoretic types in which the distance for the *i*th and *j*th electrophoretic types is given by $d_{ij} = -\ln(1 - D)$, where D is the proportion of loci with different alleles. The principle underlying the analysis is that any allelic difference in electrophoretic mobility results from at least one codon difference at the nucleotide level. Under the assumption that codon changes occur independently, standard genetic distance is an estimate of the mean number of net codon differences per genetic locus (12).

RESULTS

Occurrence of msDNA in ECOR strains. Screening of the 72 ECOR strains (13) for the production of msDNA with the RT-labeling method (9) uncovered 9 strains (13%) that produced quantities of msDNA. Seven of the nine strains contained amounts of msDNA comparable to those observed for Cl-1, although severalfold variation in msDNA production was observed (see below), suggesting that for these strains msDNA existed in hundreds of copies per cell. In contrast, two strains, ECOR-23 and ECOR-71, produced very low levels of msDNA.

A phylogenetic tree of the 72 ECOR strains was obtained by the NJ method (16) applied to a genetic distance matrix based on electrophoretically detected allelic variation at 38 enzyme-encoding loci (Fig. 1). The major topological features of the NJ tree are essentially the same as those found previously with an average-linkage cluster analysis (18). Three (groups A, B, and D) of the four major phylogenetic groups identified earlier (18) are apparent in Fig. 1. Group A is a distinct lineage comprising K-12 and K-12-like strains isolated for the most part from humans. Many of the B1 strains were originally isolated from nonprimate mammals, whereas the group B2 strains were obtained mostly from humans and other primates (18). The heterogeneous group of strains (formerly group C) that were found previously to lie outside of group B (18) have been reassigned to other lineages in the NJ tree.

An examination of the distribution of msDNA-producing strains in light of the phylogenetic relationships given in Fig. 1 yields several noteworthy observations. First, strains that harbor msDNA occur widely among the major subspecific groups of the ECOR collection: the only major group in which no msDNA-producing strains were found is the B2 group. Second, msDNA is produced by ECOR-70, -71, and -72, suggesting that the most recent common ancestor of these three strains also harbored msDNA. Third, in two cases, very closely related strains show sharp differences in the production of msDNA. Although ECOR-35 and ECOR-36 are virtually indistinguishable in the electrophoretic mobilities of 38 enzymes, indicating a very close genetic relationship, ECOR-35 contained msDNA, whereas ECOR-36 did not. In addition, ECOR-38 and ECOR-41, two strains representing a distinctive cluster in group D, produced msDNA, whereas the other members of the cluster (ECOR-39 and ECOR-40) did not. These observations suggest that the ability to synthesize msDNA has been recently gained or lost in the course of the evolutionary divergence of these strains.

Comparison with msDNA-producing clinical strains. Using

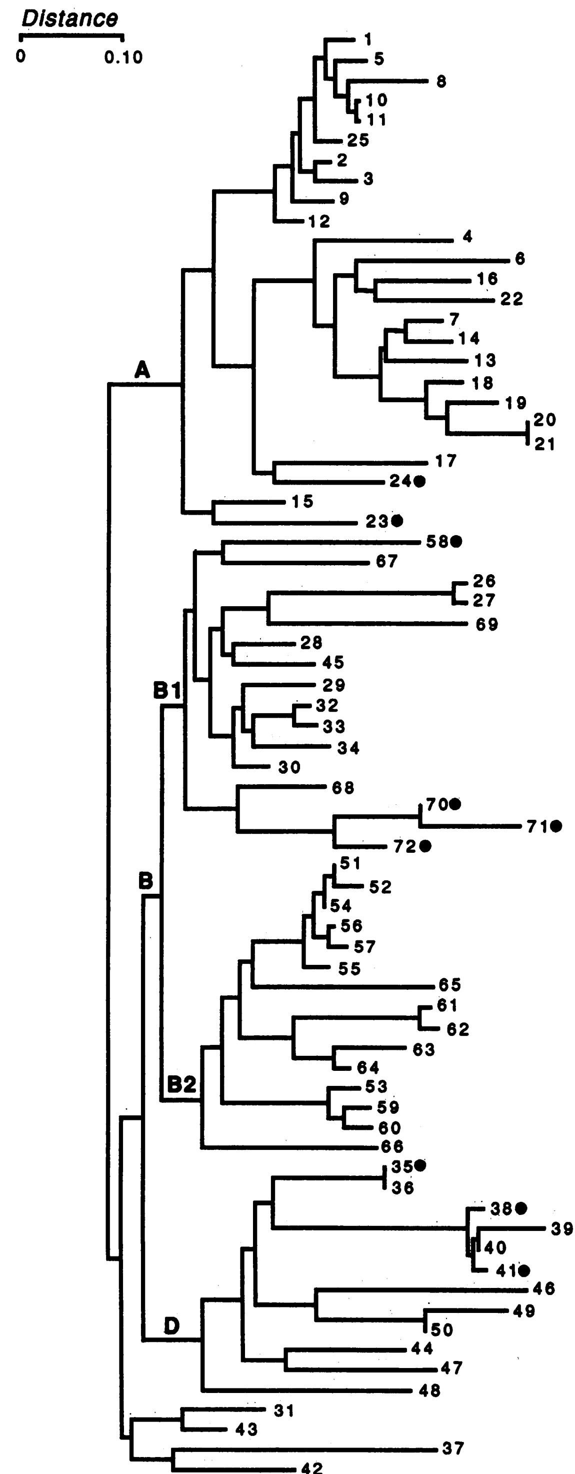


FIG. 1. Phylogenetic distribution of msDNA-containing *E. coli* strains of the ECOR collection (13). The tree was inferred by the NJ algorithm (16) applied to the genetic distances based on polymorphisms of 38 enzyme-encoding loci. Circles represent the nine ECOR strains in which msDNA was detected by the RT-labeling method (9, 10).

TABLE 1. Sources of 15 msDNA-producing *E. coli* strains of the ECOR collection and clinical strains (21)

Strain no.	Group	Serotype[a]	Host	Location	Position (bases)[b]
ECOR-23	A	O25:HN	Elephant[c]	United States (Washington)	110
ECOR-24	A	O15:NM	Human	Sweden	125
ECOR-35	D	O1:NM	Human	United States (Iowa)	100
Cl-1	D	O1:NM	Human	United States (New Jersey)	100
ECOR-38	D	O7:NM	Human	United States (Iowa)	130
ECOR-41	D	O7:NM	Human	Tonga	130
Cl-13	D	O7:NM	Human	United States (New Jersey)	130
Cl-30	D	O7:NM	Human	United States (New Jersey)	130
BC-43	D	ND	Human	United States (New Jersey)	130
ECOR-58	B1	ON:H8	Lion[c]	United States (Washington)	80
ECOR-70	B1	O78:NM	Gorilla[c]	United States (Washington)	130
ECOR-71	B1	ON:NM	Human	Sweden	130
ECOR-72	B1	O8:HN	Human	Sweden	130
BC-57	B1	O9:NM	Human	United States (New Jersey)	80
BC-61	B1	O9:NM	Human	United States (New Jersey)	130

[a] Abbreviations: N, nontypable (i.e., antigen could not be identified with standard antisera); NM, nonmotile cells with no H antigen; ND, not determined.
[b] Positions of msDNA were determined with single-stranded DNA as reference after treatment of labeled msDNA with RNase A. Although msDNA has an extensive secondary structure and migrates abnormally, the positions reflect the sizes of msDNA.
[c] From a zoo.

the RT-labeling method, Sun et al. (21) screened 113 clinical isolates of *E. coli* and found 7 strains (5%) that produced msDNA. Four of these strains (Cl-1, Cl-13, Cl-23, and Cl-30) were isolated from patients with urinary tract infections, and three (BC-43, BC-57, and BC-61) were recovered from patients with septicemia (21). We characterized these strains for electrophoretic variation of 20 enzymes and compared the electromorph profiles with those of the ECOR strains. Six of the clinical strains showed very close genetic relationships to specific genotypes of the ECOR collection (Table 1). Cl-13, Cl-30, and BC-43 were indistinguishable from ECOR-41 and ECOR-38, Cl-1 was identical to ECOR-35, and BC-57 and BC-61 exhibited single allelic differences from ECOR-70. Because the msDNA-producing clinical strains are identical or nearly identical in multilocus enzyme genotype to ECOR strains independently isolated from separate geographic localities and, in some cases, different host species (Table 1), these results show that these strains represent widespread, naturally occurring bacterial clones that have molecular systems for producing msDNA.

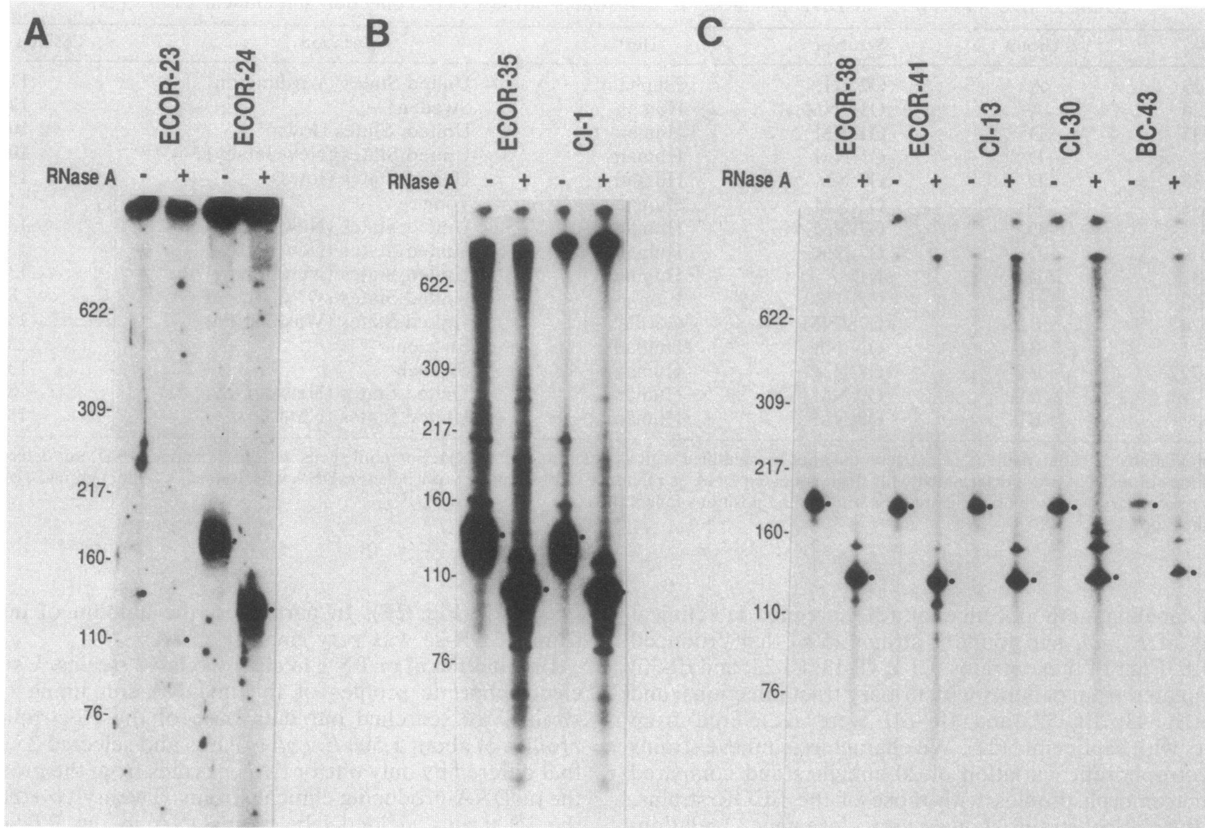
Variation in size and expression of msDNA. To compare the sizes and roughly estimate variation in amounts of msDNA synthesized by different strains, RT-labeled msDNAs prepared from total RNA fractions were treated with RNase A and electrophoretically separated as described above. After RNase A treatment, ECOR-35 produced a single band that migrated to the position equivalent to a 100-base single-stranded DNA, which was the same position and signal intensity as the band produced by Cl-1 (Fig. 2B and Table 1). ECOR-41 produced a band comparable in size (~130 bases) and intensity to those of Cl-13 and Cl-30. msDNA from strain BC-43, which was estimated previously to be 110 bases in length after RNase treatment (21), migrated to the same position as msDNAs from Cl-13 and Cl-30 under our electrophoretic conditions (Fig. 2C and Table 1). ECOR-70, -71, and -72 all produced single bands of ~130 bases in size after treatment with RNase A but showed considerable variation in band intensity. Because the same amounts of total RNA were used in these experiments, we interpret this variation as reflecting differences in the amount of msDNA synthesized by these related strains. ECOR-72 produced a signal severalfold weaker than that of ECOR-70, and ECOR-71 produced a further weaker signal than that of ECOR-72 (Fig. 2F). In particular, the amount of msDNA from ECOR-23 was very small (Fig. 2A).

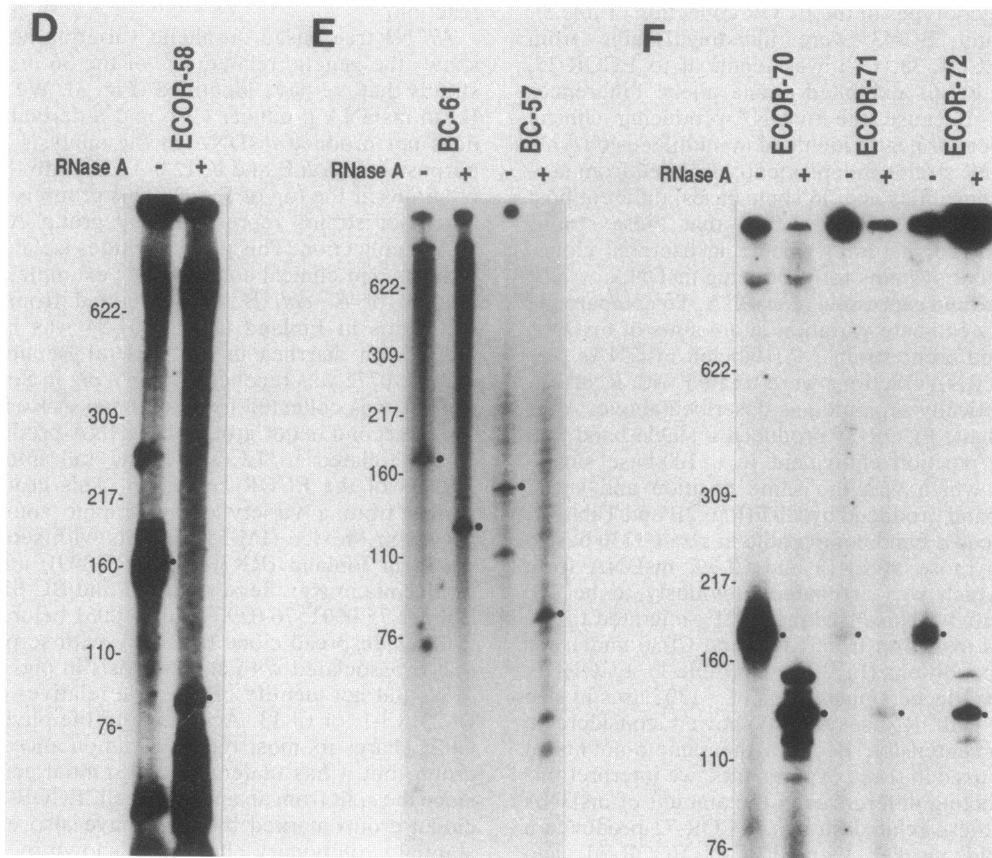
Distribution of msDNA in closely related strains. Using the electrophoretic profiles of the msDNA-containing clinical strains, we searched our data base of the electrophoretic profiles of about 1,500 *E. coli* isolates and selected 25 strains that differed by only one or two enzymes from the profiles of the msDNA-producing clinical strains. Twenty-two (88%) of the 25 strains yielded labeled msDNA in the RT-labeling reaction.

An NJ tree based on allelic variation at 20 enzyme loci shows the genetic relatedness of the 36 msDNA-producing strains that we have identified (Fig. 3). We included *E. coli* B, an msDNA producer (10), and a derivative of K-12 that does not produce msDNA in the analysis for comparative purposes. *E. coli* B and K-12 both fall into the distinct group of strains at the top of Fig. 3. This group is part of the K-12 family of strains represented by group A isolates in the ECOR collection. This group includes isolates from a variety of hosts and clinical sources. For example, RK294, a close relative of *E. coli* B, was obtained from a patient with meningitis in Finland (19), C404-84 was isolated from an infant with diarrhea in the Federal Republic of Germany (15), 82-0772 was recovered from a pig in South Dakota, and 1762-58 was collected from a human in Kentucky.

The second major group of msDNA-producing strains are closely related to ECOR-72 and fall into the group B1 isolates of the ECOR collection. This group also includes strains from a variety of geographic sources, such as a person in Mexico (JM-55), infants with septicemia or meningitis in Finland (RK306 and RK309), and patients with septicemia in New Jersey (BC-57 and BC-61). Interestingly, isolates 75-0001, 76-0296, and 77-0061 belong to a geographically widespread clone (typically with serotype O157:H43) that is associated with colibacillosis in pigs (24).

We did not identify other close relatives of clinical strain Cl-23, Cl-1, or Cl-13. According to the phylogeny in Fig. 3, Cl-23 shares its most recent common ancestor with the B1 group, but it has undergone substantial genetic divergence since the split from an ancestral cell. ECOR-58, Cl-1, and the clonal group marked by Cl-13 have also experienced substantial evolutionary change, as shown by the long branch lengths ($d > 0.20$) in the tree.


A
ECOR-23
ECOR-24
RNase A - + - +
622-
309-
217-
160-
110-
76-
B
ECOR-35
Cl-1
RNase A - + - +
622-
309-
217-
160-
110-
76-
C
ECOR-38
ECOR-41
Cl-13
Cl-30
BC-43
RNase A - + - + - + - + - +
622-
309-
217-
160-
110-
76-


D
ECOR-58
RNase A - +
622-
309-
217-
160-
110-
76-
E
BC-61
BC-57
RNase A - + - +
622-
309-
217-
160-
110-
76-
F
ECOR-70
ECOR-71
ECOR-72
RNase A - + - + - +
622-
309-
217-
160-
110-
76-

DISCUSSION

Detecting msDNA and RT. In both *E. coli* B and Cl-1, the synthesis of msDNA has been demonstrated to depend on the presence of an open reading frame encoding RT (9, 10). Comparisons of the sequences specifying msDNA and encoding RT in these strains have revealed extensive sequence divergence, and hybridization studies have shown that specific msDNA and RT probes fail to cross-react with chromosomal DNA from other strains containing msDNA (21). As a consequence, we used the RT-labeling method to detect msDNA because of its increased sensitivity compared with the direct staining of msDNA with ethidium bromide (26) and its ability to detect the presence of highly divergent msDNA molecules. However, we suspect that the RT labeling underestimates the actual frequency of RT genes in the *E. coli* population because some strains may lack sequences or contain defective sequences for producing the msDNA-RNA transcript.

Among several strains of the ECOR collection, we found reproducible variation in the intensities of bands of RT-labeled msDNA (Fig. 2F). Southern blot analysis of *Pst*I and *Blg*II digests of ECOR-70, -71, and -72 chromosomal DNAs with RT-labeled msDNA from ECOR-70 as probe suggests that all three strains contain a single gene for msDNA located at the same position on the chromosome (data not shown). We conclude, therefore, that the variation in band intensity shown in Fig. 2F indicates that these strains produce different amounts of msDNA of the same structure. The reason underlying this variation in the amount of msDNA is not known.

Genetic diversity of msDNA-producing strains. The results of our survey of the ECOR collection revise upward the estimated frequency of *E. coli* strains that carry msDNA. About 13% (9 of 72) of the ECOR isolates contain msDNA, compared with about 6% (7 of 113) estimated previously for clinical isolates (21). From the electrophoretic analysis of enzyme polymorphism, we inferred the phylogenetic relationships among the chromosomal genomes of ECOR strains and found that three of the four major phylogenetic groups included strains that produce msDNA. This observation is consistent with the hypothesis that msDNA-synthesizing systems, including the RT genes, were acquired relatively recently in the evolutionary radiation of *E. coli* lineages (21). The observation that only one (ECOR-36) of two strains with identical multilocus enzyme genotypes produced msDNA is clear evidence of the rapid gain or loss of an msDNA-synthesizing system.

The genetic relationships inferred from the analysis of multilocus enzyme polymorphisms of msDNA-producing strains are supported by earlier observations (21). The genetic similarity in chromosomal genotypes of strains Cl-13 and Cl-30, as indicated by their identity in multilocus allelic profile, was suggested by the observation that these strains have the same chromosomal restriction fragment patterns with two restriction enzymes in Southern hybridizations with two different chromosomal gene probes (21). Cl-13 and

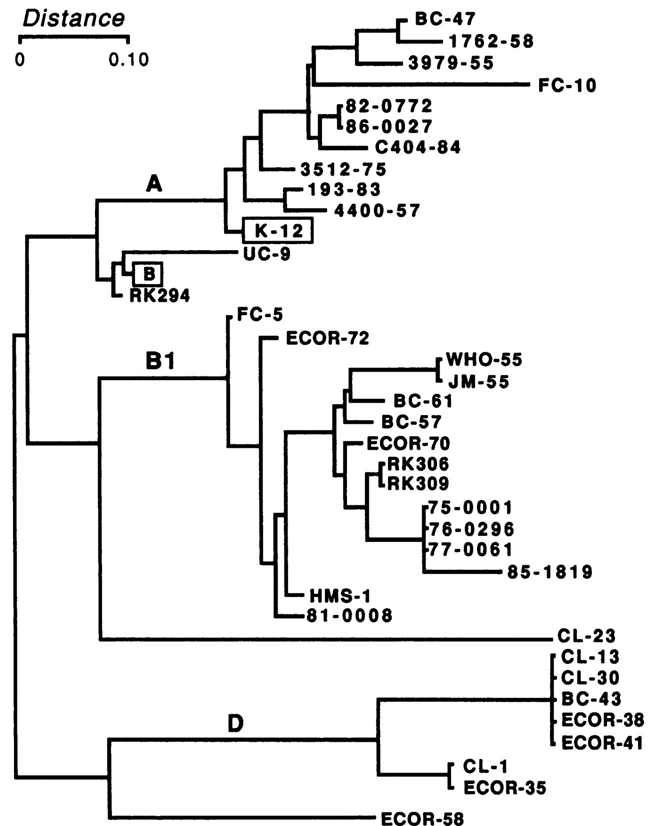


FIG. 3. Phylogenetic tree of the evolutionary divergence of 36 msDNA-producing *E. coli* strains. The tree was constructed by the NJ algorithm applied to a genetic distance matrix based on polymorphisms for 20 enzyme-encoding loci. *E. coli* K-12 does not produce msDNA and is included for comparative purposes. The phylogenetic groups identified in the ECOR collection (see Fig. 1) are designated by letters. Note that ECOR-58 is not included in the B1 group based on the 20 enzyme loci. The serotypes and geographic sources of strains are as follows: BC-47 (ON:NM), United States (New Jersey); 1762-58 (O55:NM), United States (Kentucky); 3979-55 (O55:H6), United States; FC-10 (O16:NM), United States (New Jersey); 82-0772 (O157:H8), United States (South Dakota); 86-0027 (O157:H8), United States (New Jersey); C404-84 (O126:H2), Federal Republic of Germany; 3512-75 (O26:H32), United States; 193-83 (O26:H32), United States; 4400-57 (O55:H6), Denmark; UC-9 (OX2:H35), United States; RK294 (O25:H1), Finland; FC-5 (O8:NM), United States (New Jersey); WHO-55 (O55:NM), unknown; JM-55 (O55:NM), Mexico; RK306 (O8:H19), Finland; RK309 (O8:H19), Finland; 75-0001 (O157:H43), United States; 76-0296 (O157:H43) and 77-0061 (O157:H43), United States (Pennsylvania); 85-1819 (O157:NM), United States (Virginia); HMS-1, United States (Massachusetts); and 81-0008 (O55:H9), United States. See Table 1, footnote *a*, for the serotype abbreviations.

FIG. 2. Size variation of msDNA in nine *E. coli* strains of the ECOR collection and six related clinical msDNA-producing isolates. Total RNA fraction extracted from a 2.5-ml overnight culture was applied to each lane in panel A. The RNA fractions were treated without (−) and with (+) RNase A. The following amounts, relative to those from the 2.5-ml cultures, were applied to other lanes in panels B to F: ECOR-23 (−), 1/9; ECOR-23 (+), 1/6; Cl-1 (−), 1/12; Cl-1 (+), 1/4; ECOR-38 (−), 1/30; ECOR-38 (+), 1/15; ECOR-41 (−), 1/30; ECOR-41 (+), 1/15; Cl-13 (−), 1/15; Cl-13 (+), 1/8; Cl-30 (−), 1/6; Cl-30 (+), 1/3; BC-43 (−), 1/2; BC-43 (+), 1; ECOR-58 (−) and (+), 1/30; BC-61 (−) and (+), 1/15; BC-57 (−), 1/15; BC-57 (+), 1/5; ECOR-70 (−) and (+), 1/30; ECOR-71 (−) and (+), 1/30; and ECOR-72 (−) and (+), 1/30. The positions of msDNA are indicated by dots at the right side of each lane.

Cl-30 also were shown to be distinct from Cl-1, *E. coli* K-12, and *E. coli* B in their chromosomal restriction sites (21) and multilocus enzyme profiles (Fig. 3).

A key finding of our study is that in three cases, msDNA-producing clinical strains were identical or nearly identical in multilocus enzyme genotype to msDNA-producing strains of the ECOR collection (Table 1). Because the strains were collected at different times from unassociated hosts in separate geographic localities, this observation indicates that these bacteria represent widespread naturally occurring clones that owe their similarity to descent from a recent common ancestral cell. In all three cases, strains of the same clone produced msDNA of similar size and amounts as found after treatment with RNase. Thus, it is likely that isolates of the same clone, as marked by distinct electrophoretic types, have inherited the same msDNA-synthesizing system from a progenitor cell of each clone.

The diversity of chromosomal backgrounds in *E. coli* in which msDNA is synthesized is reflected in the deep phylogenetic branches shown in Fig. 3. By screening for msDNA in closely related strains of known msDNA producers, we uncovered two major clusters of strains with multiple independent isolates of diverse origin. One of these clusters (labeled A at the top of Fig. 3) includes laboratory strain K-12 and 10 other isolates that harbor msDNA. Laboratory strain B and isolates RK294 and UC-9 form a distinct lineage that shares the most recent common ancestor with the K-12 cluster. The extent to which the isolates of these clusters share characteristics in their msDNA-synthesizing systems, including RT genes, is unknown, but preliminary screening of msDNA sizes indicates some variation in the sizes of single-stranded DNAs after RNase treatment (data not shown). A second distinct cluster, including ECOR-70 and -72, includes 15 isolates, most of which produce DNAs of about 130 bases in sizes after RNase treatment (data not shown). The three remaining lineages (marked by Cl-23, Cl-1, and ECOR-58) are only distantly related to the two major clusters. Thus, on the basis of the phylogenetic analysis, we recognize six genetically distinct subspecific lineages within the *E. coli* population in which msDNA synthesis occurs.

Comparisons of the msDNAs and RT-encoding genes of the clinical strains and *E. coli* B show that their msDNA-synthesizing systems are highly divergent. Sun et al. (21), using specific probes for msDNA and RT obtained from Cl-1, found no hybridization with either probe to chromosomal DNA from *E. coli* B, Cl-13, Cl-23, and Cl-30, a result indicating little homology in the msDNA-synthesizing systems of these strains. Furthermore, the msdRNA and the predicted proteins encoded by the RT genes of *E. coli* B and Cl-1 differ substantially in size (22) and bear little similarity in primary sequence (21).

Two hypotheses that account for the existence of highly divergent msDNA-synthesizing systems in selected lineages of *E. coli* are as follows. First, production of msDNA is an ancient attribute of *E. coli* populations, and the genes mediating msDNA synthesis have undergone rapid divergence and been repeatedly lost in independent lines. Second, the genes specifying distinct msDNA-synthesizing systems have diverged in other bacterial species and were subsequently transferred independently and relatively recently into different lineages that constitute extant *E. coli* populations. Although the recent loss of msDNA production is consistent with the observation in the ECOR strains of genetically similar isolates with and without msDNA, support for the latter hypothesis comes from the analysis of the codon usage in the RT genes, which differs from the usage in most *E. coli* protein-encoding genes (9). Indeed, the codon adaptation indices (20), a measure of the degree of synonymous codon usage bias, for the RTs of Cl-1 and *E. coli* B are 0.172 and 0.170, respectively, values lower than those reported for 165 *E. coli* genes (20). Such a low degree of codon adaptation strongly argues that the RT genes involved in msDNA synthesis are foreign to *E. coli* and were relatively recently acquired via horizontal transfer.

Evolutionary significance of msDNA. Although the synthesis of msDNA by RT is not essential for bacterial growth and reproduction, as evidenced by the fact that most *E. coli* strains do not produce msDNA, reverse transcription might have a significant evolutionary role in the generation of genetic variation. One possibility, suggested by Cairns and colleagues (1), is that cells could reverse transcribe DNA from a variable set of mRNAs and incorporate mutations in a directed fashion, a versatile process for adaptation to strong selection pressures (but see reference 11 for new data that do not support the directed mutation hypothesis). Reverse transcription might also provide a source of DNA for gene duplications or DNA repair (22). Even if it only rarely occurs, reverse transcription, like homologous recombination, could play a significant role in the origin of new bacterial genotypes.

ACKNOWLEDGMENTS

We thank Marie Wolfe for technical assistance, Anita Marcase, Michael Davis, and Richard Wilson for serotyping strains, and Robert Selander for comments on the manuscript.

The research was supported by Public Health Service grants GM 26843 (S.I. and M.I.) and AI 24566 (T.S.W.) from the National Institutes of Health.

LITERATURE CITED

1. **Cairns, J., J. Overbaugh, and S. Miller.** 1988. The origin of mutants. Nature (London) **335**:142–145.
2. **Chomzynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.
3. **Dhundale, A., M. Inouye, and S. Inouye.** 1988. A new species of multicopy single-stranded DNA from *Myxococcus xanthus* with conserved structural features. J. Biol. Chem. **263**:9055–9058.
4. **Dhundale, A., B. Lampson, T. Furuichi, M. Inouye, and S. Inouye.** 1987. Structure of msDNA from Myxococcus xanthus: evidence for a long, self-annealing RNA precursor for the covalently linked, branched RNA. Cell **51**:1105–1112.
5. **Furuichi, T., A. Dhundale, M. Inouye, and S. Inouye.** 1987. Branched RNA covalently linked to the 5′ end of a single-stranded DNA in Stigmatella aurantiaca: structure of msDNA. Cell **48**:47–53.
6. **Furuichi, T., S. Inouye, and M. Inouye.** 1987. Biosynthesis and structure of stable branched RNA covalently linked to the 5′ end of multicopy single-stranded DNA of Stigmatella aurantiaca. Cell **48**:55–62.
7. **Goullet, PH., and B. Picard.** 1989. Comparative electrophoretic polymorphism of esterases and other enzymes in *Escherichia coli*. J. Gen. Microbiol. **135**:135–143.
8. **Inouye, S., M.-Y. Hsu, S. Eagle, and M. Inouye.** 1989. Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in Myxococcus xanthus. Cell **56**:709–717.
9. **Lampson, B. C., J. Sun, M.-Y. Hsu, J. Vallejo-Ramirez, S. Inouye, and M. Inouye.** 1989. Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. Science **243**:1033–1038.
10. **Lim, D., and W. K. Maas.** 1989. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in E. coli B. Cell **56**:891–904.
11. **Mittler, J. E., and R. E. Lenski.** 1990. New data on excisions of

Mu from *E. coli* MCS2 cast doubt on directed mutation hypothesis. Nature (London) **344**:173–175.
12. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York.
13. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157**:690–693.
14. **Ochman, H., T. S. Whittam, D. A. Caugant, and R. K. Selander.** 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J. Gen. Microbiol. **129**:2715–2726.
15. **Ørksov, F., T. S. Whittam, A. Cravioto, and I. Ørskov.** 1990. Clonal relationships among classic enteropathogenic *Escherichia coli* (EPEC) belonging to different O groups. J. Infect. Dis. **162**:76–81.
16. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
17. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51**:873–884.
18. **Selander, R. K., D. A. Caugant, and T. S. Whittam.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. *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.
19. **Selander, R. K., T. K. Korhonen, V. Väisänen-Rhen, P. H. Williams, P. Pattison, and D. A. Caugant.** 1986. Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. Infect. Immun. **52**:213–222.
20. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of the directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**:1281–1295.
21. **Sun, J., P. J. Herzer, M. P. Weinstein, B. C. Lampson, M. Inouye, and S. Inouye.** 1989. Extensive diversity of branched-RNA-linked multicopy single-stranded DNAs in clinical strains of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **86**:7208–7212.
22. **Varmus, H. E.** 1989. Reverse transcription in bacteria. Cell **56**:721–724.
23. **Whittam, T. S., H. Ochman, and R. K. Selander.** 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80**:1751–1755.
24. **Whittam, T. S., and R. A. Wilson.** 1988. Genetic relationships among pathogenic *Escherichia coli* of serogroup O157. Infect. Immun. **56**:2467–2473.
25. **Whittam, T. S., and R. A. Wilson.** 1988. Genetic relationships among pathogenic strains of avian *Escherichia coli*. Infect. Immun. **56**:2458–2466.
26. **Yee, T., T. Furuichi, S. Inouye, and M. Inouye.** 1984. Multicopy single-stranded DNA isolated from a gram-negative bacterium, Myxococcus xanthus. Cell **38**:203–209.