Nucleotide Polymorphism in Colicin E2 Gene Clusters: Evidence for Nonneutral Evolution

Ying Tan and Margaret A. Riley
Department of Biology, Yale University

To explore the molecular mechanisms behind the diversification of colicin gene clusters, we examined DNA sequence polymorphism for the colicin gene clusters of 14 colicin E2 (ColE2) plasmids obtained from natural isolates of *Escherichia coli.* Two types of ColE2 plasmids are revealed, with type II gene clusters generated by recombination between type I ColE2 and ColE7 gene clusters. The levels and patterns of DNA polymorphism are different between the two types. Type I polymorphism is distributed evenly along the gene cluster, while type II accumulates polymorphism at an elevated rate in the 5′ end of the colicin gene. These differences may be explained by recombinational origins of type II gene clusters. The pattern of divergence between the ColE2 gene cluster and its close relative ColE9 is not correlated with the pattern of polymorphism within ColE2, suggesting that this gene cluster is not evolving in a neutral fashion. A statistical test confirms significant departures from the predictions of neutrality. These data lend further support to the hypothesis that colicin gene clusters may evolve under the influence of nonneutral forces.

Introduction

Bacteriocins are compounds (usually proteins) that are produced by bacteria and inhibit or kill closely related species (Tagg et al. 1976). Their production occurs over the entire range of Gram-negative and Gram-positive bacteria, and may extend into the archaebacteria as well (Torreblanca, Meseguer, and Ventosa 1994; Dykes 1995). Due to the broad phylogenetic distribution of bacteriocin-producing species, high levels of diversity in molecular structures and mechanisms are predicted among the members of the bacteriocin family. Colicins, which are a class of extensively studied bacteriocins, have served as a model system with which to address mechanisms of bacteriocin diversity.

Colicins are large domain-structured proteins synthesized by *Escherichia coli.* Twenty-three colicin types have been identified by their corresponding specific immunities (Pugsley and Oudega 1987). Common features for all colicin types include: (1) all colicin proteins are organized into four functional domains, (2) the genes encoding colicins and associated proteins, i.e., immunity and lysis proteins, exist as tightly linked gene clusters on plasmid replicons, and (3) they are usually regulated via the SOS system of induction.

A large body of molecular data for colicin gene clusters has allowed DNA and amino acid sequence comparisons, which have added to our understanding of the evolutionary relationships of colicin, immunity and lysis proteins, and the molecular mechanisms involved in the origin and diversification of colicin gene clusters (Lau, Parsons, and Uchimura 1992; Riley 1993*a*, 1993*b*). Such mechanisms include horizontal transfer of colicin plasmids between strains of *E. coli* and subsequent plasmid cointegration, inter- and intragenic recombination, transposition, and rapid diversification of colicin and immunity proteins through the action of positive selection.

The role of positive selection in generating immunity diversification was the initial focus of this study. This mechanism was proposed to explain an unusual pattern of divergence between three pairs of closely related colicin gene clusters (ColE3/E6, ColE2/E9, and ColIa/Ib) (Riley 1993*a*, 1993*b*). DNA sequence comparisons reveal an apparent excess of substitutions in the immunity regions (i.e., the immunity gene and the immunity binding domain of the colicin gene) of these three pairs of colicin gene clusters. To account for such an unusual clustering of substitutions, Riley (1993*a*, 1993*b*) proposed that colicin gene clusters may diverge rapidly in the immunity region through a mutation-selection process. Repeated waves of this mutation-selection process result in high levels of substitution in the immunity binding region.

The diversifying-selection hypothesis was tested by comparison of DNA sequence polymorphism for the colicin gene clusters of six ColIa plasmids obtained from natural isolates of *E. coli* (Riley, Tan, and Wang 1994). Patterns of divergence between colicin gene clusters of ColIa and ColIb and patterns of polymorphism within ColIa revealed a sharp discrepancy, with an elevated level of divergence in the immunity region and an even distribution of DNA polymorphism across the entire colicin gene cluster. This result contrasts with neutral predictions, which suggest a positive correlation between patterns of divergence and polymorphism. However, due to the high level of divergence in the immunity regions of ColIa and ColIb, the available statistical tests could not be applied to the region predicted to experience the strongest positive selection.

In this study, the entire colicin gene cluster of 14 ColE2 plasmids obtained from natural isolates of *E. coli* was sequenced, and the levels and patterns of DNA polymorphism were examined. A close relative (ColE9) provided a reliable estimate of divergence for this region. Again, the patterns of divergence and polymorphism suggest a departure from neutrality. These data lend further support for the hypothesis that certain col-

Key words: colicin diversity, ColE2 plasmid, *Escherichia coli,* polymorphism, diversifying selection, molecular evolution.

Address for correspondence and reprints: Margaret A. Riley, Department of Biology, Yale University, New Haven, Connecticut 06520-8104. E-mail: riley@beagle.biology.yale.edu.

Mol. Biol. Evol. 14(6):666–673. 1997

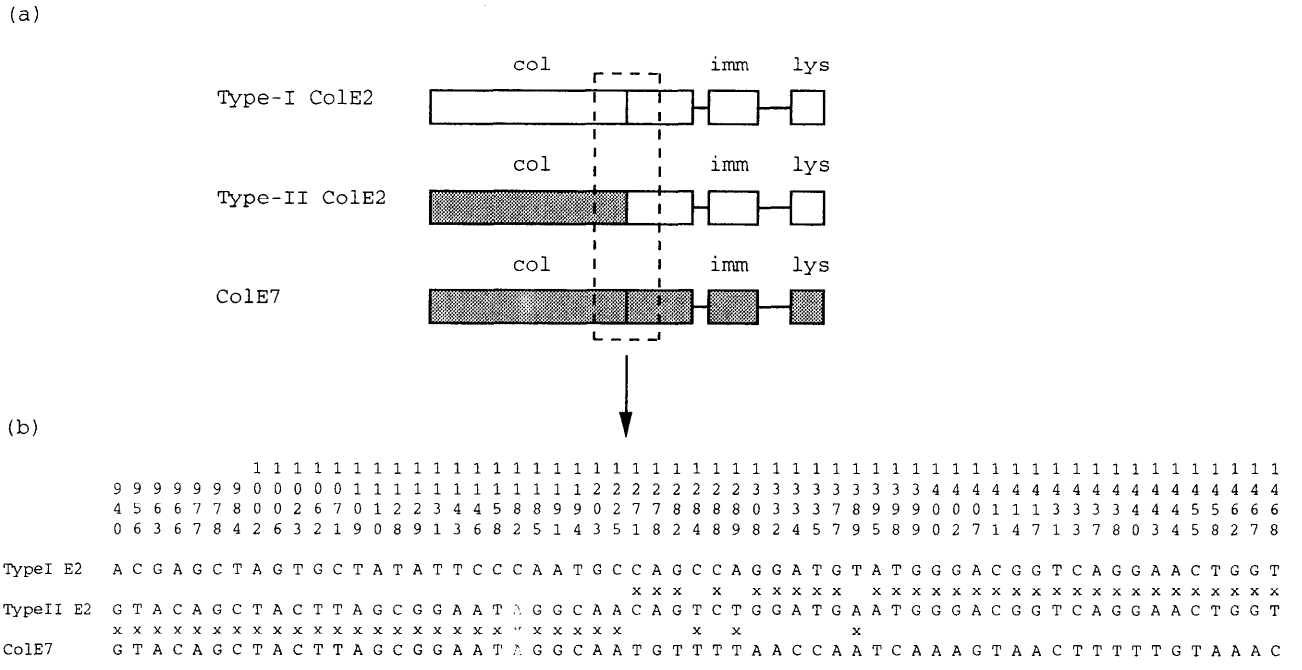

FIG. 1.—*a*, Genetic organization of type I and type II ColE2 and ColE7 gene clusters and their sequence similarity. Boxes indicate colicin, immunity, and lysis genes within a gene cluster. Connecting lines indicate intergenic sequences. The line in the colicin gene is the presumptive recombination breakpoint. *b*, Sequence comparison of 500 bp around the recombination boundary; only nucleotides different among the three sequences are shown.

icin gene clusters may diversify under the influence of some form of positive selection.

Materials and Methods

Plasmid Isolates

Five *E. coli* strains that carry ColE2 plasmids were isolated by Fredericq from unknown sources in France about 50 years ago and characterized by Watson et al. (1981): GEI288, GEI544, GEI602, K321, and CA42. The remaining nine ColE2-harboring *E. coli* strains were isolated from feral house mice in Australia over a period of 6 months in 1994 and identified by Gordon (unpublished data). These strains are DG21, DG36, DG79, DG163, DG176, DG266, DG279, DG353, and DG356.

Nucleotide Sequencing

Double-stranded DNA was prepared for each ColE2 plasmid, which is approximately 6.1 or 6.8 kb, using the Wizard Mini-prep Kit (Promega). This DNA was used as a template for PCR amplification of the ColE2 gene cluster employing primers designed based on the published colicin E2 gene cluster sequence (Cole, Saint-Joanis, and Pugsley 1985). The 5′ PCR primer was 5′-TTTATGAGCGGTGGCGAT-3′ (bp 328–345 from Cole, Saint-Joanis, and Pugsley 1985) and the 3′ primer was 5′-TCGGGTTACTGCGTTGCTAA-3′ (bp 2549–2530 from Cole, Saint-Joanis, and Pugsley 1985). A PCR fragment of 2.2 kb was subcloned using the TA Cloning Kit (Invitrogen Corporation), and double-stranded template was prepared using the Wizard Mini-prep Kit. Sequencing reactions were prepared for analysis on an ABI 373A automated sequencer using the Taq DyeDeoxy Terminator Cycle Sequencing Kit (ABI Applied Biosystems, Inc.). Oligonucleotide primers for sequencing were constructed at approximately 300-bp intervals based on available DNA sequence information (Cole, Saint-Joanis, and Pugsley 1985; Chak et al. 1991; Lau, Parsons, and Uchimura 1992). Both strands were sequenced.

Results

Two Types of ColE2 Plasmids

DNA sequences of the colicin, immunity, and lysis genes were determined for 14 natural isolates of the ColE2 plasmid of *E. coli*. The ColE2 (pColE2-P9) plasmid from the Pugsley colicin plasmid collection, whose ColE2 gene cluster had previously been sequenced (Cole, Saint-Joanis, and Pugsley 1985), was also included in this study.

Based on DNA sequences, two types of ColE2 plasmids can be distinguished. Type I plasmids, which include ColE2 plasmids isolated from *E. coli* host strains GEI288, GEI544, GEI602, and K321, share 95%–98% DNA sequence identity to the previously published ColE2 gene cluster sequence. Type II plasmids, which consist of the remaining ColE2 plasmids (isolated from host strains CA42, DG21, DG36, DG79, DG163, DG176, DG266, DG279, DG353, and DG356), possess a mosaic structure of the ColE2 gene cluster (fig. 1). The 5′ end of the type II colicin gene shares a reduced level sequence identity (80%) to the published ColE2 gene cluster sequence relative to that observed in type I plasmids. However, it shows 97% sequence identity to the 5′ end of the colicin gene of the ColE7 plasmid pColE7-K317 (Chak et al. 1991; Lau, Parsons, and Uchimura 1992). The 3′ region of the type II ColE2

```
                 col          imm
                      11111**22
                 44455677902455**23
                 13435236840512**40
                 48452717196621**96
P9               TCAGTATAAATAAT**CT
GEI602           .TG......C..G.**.C
GEI288           .TGAC..T.CCGGC**..
GEI544           .TG.CTG.GCCGGC**..
K321             CTGAC....CCGGC**T.
                    G  VK E N
                    S  GI A D
```

FIG. 2.—Polymorphic nucleotides and amino acids among the type I ColE2 gene clusters and encoded proteins. Standard single-letter amino acid abbreviations are used. Numbering of nucleotides is given above DNA sequences. Bold letters above numbers indicate each region. Stars are employed to distinguish between coding and noncoding sequences.

gene cluster, including the 3′ end of the colicin gene (encoding the immunity binding domain), the immunity gene, and the lysis gene, shows 97% sequence identity with the published ColE2 gene cluster sequence. This level of sequence identity is similar to that observed for type I ColE2 sequences.

Types I and II differ at 15% of their nucleotides. Such grouping of ColE2 sequences does not totally reflect geographical origin of the bacterial hosts. For example, host strain CA42, isolated in France, carries a ColE2 plasmid more closely related to those from DG strains, isolated in Australia. Due to the high level of divergence between types, the nucleotide polymorphisms were examined for type I and type II ColE2 plasmids separately as well as combined.

Total Nucleotide Polymorphism

There are 16 polymorphic sites in the 2,214-bp region examined among the five type I ColE2 plasmids (fig. 2 and table 1). ColE2 gene clusters from these plasmids differ, on average, at 7.4 nucleotides and have an average nucleotide diversity (or base pair heterozygosity) of 0.003 ± 0.002. Two subgroups of type I ColE2 gene clusters can be distinguished. The colicin gene clusters from strains GEI288, GEI544, and K321 form the first group, with an average of 4.7 nucleotide differences and an average nucleotide diversity of 0.002 ± 0.001. The remaining ColE2 gene clusters from host strains P9 and GEI602 fall into the second group, differing at 5 nucleotides with an average nucleotide diversity of 0.002 ± 0.000.

There are 52 polymorphic sites in the 2,202-bp region examined among the 10 type II ColE2 plasmids (fig. 3 and table 1), 6 of which are insertion/deletions. These plasmids differ, on average, at 13.9 nucleotides and have an average nucleotide diversity of 0.006 ± 0.004. ColE2 plasmids from strains DG163, DG176, and DG266 are clearly more closely related, sharing six identical insertion/deletions and several other unique polymorphic sites. These plasmids differ, on average, at 3.3 nucleotides and have an average nucleotide diversity of 0.002 ± 0.001. The six remaining type II ColE2 plasmids from DG strains (DG21, DG36, DG79, DG279, DG353, and DG356) form the second group and differ, on average, at 1.3 nucleotides, with an average nucleotide diversity of 0.001 ± 0.001. The last type II ColE2 plasmid, isolated from strain CA42, is distinct from all the other type II ColE2 plasmids carried by DG strains.

The total level of nucleotide diversity observed among the type I ColE2 sequences is two-fold lower than that of the type II ColE2 sequences. However, among the four subgroups of type I and type II ColE2 sequences, the levels of within-subgroup diversity are not significantly different ($G = 2.744$, df = 3, $P > 0.30$).

Since type I and type II gene clusters share high sequence similarity in the 3′ region of the gene cluster, combined nucleotide polymorphisms were also examined in this region. There are 23 polymorphic sites in the 831-bp region examined among the 15 ColE2 plasmids (table 1). These plasmids differ, on average, at 9.2

Table 1
Nucleotide Polymorphism and Diversity in ColE2 Gene Clusters

Region	Total Sites	Total Poly	Total % Poly	K_t	Syn Sites	Syn Poly	Syn % Poly	K_s	Nonsyn Sites	Nonsyn Poly	Nonsyn % Poly	K_n
Type I												
Col	1,746	14	0.80	0.004 ± 0.003	401	9	2.24	0.010 ± 0.007	1,344	5	0.37	0.002 ± 0.001
Imm	261	2	0.77	0.003 ± 0.003	51	2	3.92	0.015 ± 0.018	209	0	0	0
Inter	61	0	0	0								
Lys	144	0	0	0	37	0	0	0	106	0	0	0
Type II												
Col	1,728	43	2.49	0.008 ± 0.005	400	29	7.25	0.002 ± 0.014	1,329	14	1.05	0.003 ± 0.002
Imm	261	1	0.38	0.001 ± 0.002	52	0	0	0	208	1	0.48	0.001 ± 0.002
Inter	61	2	3.28	0.007 ± 0.011								
Lys	144	0	0	0	37	0	0	0	106	0	0	0
Combined												
3 Col ...	363	11	3.03	0.014 ± 0.006	71	6	8.45	0.047 ± 0.018	291	5	1.72	0.007 ± 0.004
Imm	261	7	2.68	0.009 ± 0.005	52	5	9.62	0.032 ± 0.021	208	2	0.96	0.003 ± 0.003
Inter	61	2	3.28	0.004 ± 0.009								
Lys	144	3	2.08	0.010 ± 0.005	37	2	5.41	0.025 ± 0.016	106	1	0.94	0.005 ± 0.004

NOTE.—Abbreviations used as follows: Poly, polymorphism; Syn, synonymous; Nonsyn, nonsynonymous; Col, colicin; Imm, immunity; Inter, intergenic; Lys, lysis.

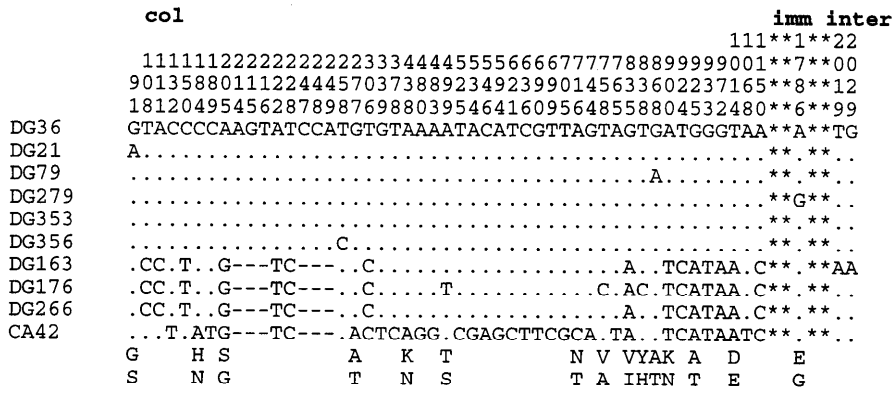

FIG. 3.—Polymorphic nucleotides and amino acids among the type II gene clusters and encoded proteins. Denotations are similar to those in figure 2.

nucleotides and have an average nucleotide diversity of 0.011 ± 0.005.

Synonymous and Nonsynonymous Polymorphisms

Table 1 summarizes the levels of noncoding, synonymous, and nonsynonymous polymorphism for each region sequenced in both types of ColE2 plasmids and provides an estimate of nucleotide diversity for synonymous (K_s) and nonsynonymous (K_n) sites for each gene. For type I ColE2 plasmids, estimates of synonymous and nonsynonymous polymorphism are not significantly different among the colicin, immunity, and lysis genes (synonymous: $G = 2.216$, df = 2, $P > 0.30$; nonsynonymous: $G = 2.110$, df = 2, $P > 0.20$). For type II ColE2 plasmids, there are significant differences in estimates of synonymous polymorphism among the three genes, with an increased level of synonymous polymorphism in the colicin gene relative to the immunity and lysis genes ($G = 12.053$, df = 2, $P < 0.01$). No significant differences in estimates of nonsynonymous polymorphism among the three genes of type II ColE2 plasmids are found ($G = 2.740$, df = 2, $P > 0.20$).

For type I ColE2 plasmids, synonymous and nonsynonymous polymorphic sites are distributed evenly among the colicin gene when the sequence is divided into six equal-sized blocks, each with 291 nucleotides (synonymous: $G = 5.676$, df = 5, $P > 0.30$; nonsynonymous: $G = 4.625$, df = 5, $P > 0.30$). In contrast, neither synonymous nor nonsynonymous polymorphic sites are distributed evenly along the colicin gene of type II ColE2 plasmids (synonymous: $G = 20.993$, df = 5, $P < 0.01$; nonsynonymous: $G = 13.054$, df = 5, $P < 0.05$). Synonymous and nonsynonymous polymorphic sites appear clustered in the 5′ end of type II colicin genes, i.e., the region that shares high levels of sequence similarity with colicin E7. Synonymous and nonsynonmous polymorphic sites are distributed evenly in the immunity and lysis genes of both types of plasmids if such polymorphic sites are found (type I immunity: synonymous $G = 1.561$, df = 2, $P > 0.30$; type II immunity: nonsynonymous $G = 2.207$, df = 2, $P > 0.30$).

For the combined data, estimates of synonymous and nonsynonymous polymorphism are not significantly different when compared among the 3′ end of the colicin gene, the immunity gene, and the lysis gene (synonymous: $G = 0.5686$, df = 2, $P > 0.70$; nonsynonymous: $G = 0.6786$, df = 2, $P > 0.70$). Within each gene, synonymous and nonsynonymous polymorphic sites are also distributed evenly (colicin: synonymous $G = 5.199$, df = 2, $P > 0.05$; nonsynonymous $G = 0.4741$, df = 2, $P > 0.70$; immunity: synonymous $G = 5.695$, df = 2, $P > 0.05$; nonsynonymous $G = 1.632$, df = 2, $P > 0.30$; lysis: synonymous $G = 2.558$, df = 1, $P > 0.10$; nonsynonymous $G = 1.321$, df = 1, $P > 0.20$).

Statistical Tests of Neutrality

A statistical test of the neutral model, the HKA test (Hudson, Kreitman, and Aguade 1987), has been developed to detect a departure from the prediction of a positive correlation between patterns of nucleotide polymorphism within species and patterns of nucleotide divergence between species. Employing the polymorphism data of the ColE2 gene cluster and divergence data obtained from a comparison of the ColE2 and ColE9 gene clusters (Riley 1993*b*), two HKA tests were carried out. The first test employed the polymorphism data of type I ColE2 gene clusters, while the second made use of the combined polymorphism data of ColE2 gene clusters. In the test with combined data, the 5′ ends of type II colicins were not included because these sequences belong to the E7 type, which is very different from type I E2 sequences. Since the divergence data for the test were from type I E2 and E9 comparisons, only type I alleles for the 5′ end of the colicin gene could be used in the test. Corrections for unequal sample sizes were performed for the combined data according to Berry et al. (1991).

In both of our tests, the 5′ end of the colicin gene and the immunity region, which includes the 3′ end of the colicin gene (i.e., the immunity binding region) and the immunity gene, are used as the two compared regions. The boundaries of the two regions are defined based on functional domains of colicins. Numerous examples of recombination between functional domains that give rise to a mosaic structure of colicins (Roos, Harkness, and Braun 1989; Pilsl and Braun 1995) have shown more recombination between regions. Thus, we can assume that the two regions being compared are independent loci, as required for the application of the

Table 2
HKA Tests on Type I ColE2 Gene Clusters

	5′ End of Colicin Gene		Immunity Region	
	No. Sites Compared	No. Sites Variable	No. Sites Compared	No. Sites Variable
Within type I ColE2 (n = 5)	1,383	14	626	2
Between type I ColE2 and ColE9	1,386	52	626	150

NOTE.—The relevant parameters were calculated by solving simultaneous equations. The divergence estimates were calculated from the comparison of the ColE2 sequence from DG36 strain with the published ColE9 sequence. $\theta_1 = 1.726 \times 10^{-3}$, $\theta_2 = 8.454 \times 10^{-3}$, $T = 24.504$, $\chi^2 = 9.544$, $P < 0.005$.

HKA test. Both of our HKA tests reveal significant departures from neutral predictions (table 2 and 3).

A second test of neutral prediction, the MK test (McDonald and Kreitman 1991), compares the ratio of synonymous to nonsynonymous polymorphism to the ratio of synonymous to nonsynonymous divergence. If the ratios are not the same, one can reject the null hypothesis of neutrality. We have applied this test to the immunity region employing polymorphism data for the E2 immunity region and divergence data obtained from a comparison of the E2 and E9 immunity regions. This test also reveals significant departures from neutral predictions (table 4).

Discussion

DNA sequence data reveal two types of ColE2 gene clusters. A high level of DNA sequence difference is found when the colicin genes are compared between type I and type II ColE2 gene clusters. This genotypic difference may explain the previous phenotypic observation that strain CA42 produced a colicin E2 protein immunologically distinct from the characterized colicin E2-P9 (Lewis and Stocker 1965). The distinction between the two ColE2 sequence types is also consistent with a previous finding which reveals two structurally distinct ColE2 plasmid types based on the restriction mapping of six ColE2 plasmids (Watson, Vernet, and Visentin 1985). Watson, Vernet, and Visentin's study revealed that five ColE2 plasmids isolated from strains P9,

Table 3
HKA Test on Combined Data of ColE2 Gene Cluster

	5′ End of Colicin Gene		Immunity Region	
	No. Sites Compared	No. Sites Variable	No. Sites Compared	No. Sites Variable
Within ColE2	1,383	14 (n = 5)	626	18 (n = 15)
Between ColE2 and ColE9	1,386	52	626	150

NOTE.—The relevant parameters were calculated by solving simultaneous equations. The divergence estimates were calculated from the comparison of the ColE2 sequence from DG36 strain with the published ColE9 sequence. $\theta_1 = 2.331 \times 10^{-3}$, $\theta_2 = 1.242 \times 10^{-3}$, $T = 17.353$, $\chi^2 = 3.930$, $P < 0.05$.

Table 4
MK Test on the Immunity Region

	Nonsynonymous Substitutions	Synonymous Substitutions
Polymorphism within ColE2 (n = 15)	7	11
Divergence between ColE2 and ColE9	95	55

NOTE.—The divergence estimates were calculated from the comparison of the ColE2 sequence from DG36 strain with the published ColE9 sequence. $G = 3.919$, df = 1, $P < 0.05$.

GEI288, GEI544, GEI602, and K321 shared most restriction sites and, thus, were closely related, while plasmid ColE2-CA42 remained as a unique type, sharing many restriction sites with plasmid ColE7-K317. The presence of certain restriction sites for the ColE2 plasmids from nine DG strains also reveals that their ColE2 plasmid type falls into the same class as CA42, as does their ColE2 gene cluster type.

The type II ColE2 plasmid appears to be derived from a recombination event between type I ColE2 and ColE7 plasmids, with the 5′ end of the colicin gene and the remainder of the plasmid "outside" of the colicin gene cluster originating from ColE7 (fig. 4). This conclusion is supported by several lines of evidence. First, our sequence data show that there is a high level of sequence similarity between the 5′ ends of the type II ColE2 and ColE7 gene clusters. Further, both previous studies and our data on restriction sites reveal that type II ColE2 plasmids share the same size (i.e., 6.1 kb) and most of the restriction sites flanking the colicin gene cluster region with ColE7 plasmid (Watson, Vernet, and

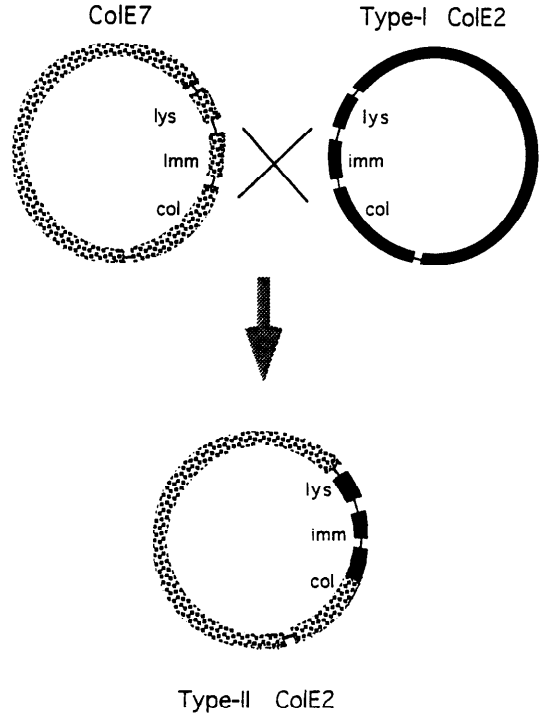

FIG. 4.—The recombinational origination of the type II ColE2 plasmid.

Visentin 1985; this study). In addition, ColE2 and ColE7 plasmids are compatible with each other and thus can coexist in a cell (Lau, Parsons, and Uchimura 1992). Indeed, a ColE7 plasmid was found coexisting with a cryptic ColE2 plasmid in strain K317 (Males and Stocker 1980; Watson et al. 1981).

The recombinational origin of the type II ColE2 gene cluster may explain the differences in the levels and patterns of nucleotide polymorphism between type I and type II ColE2 gene clusters. Our data reveal that the level of polymorphism in the type II ColE2 gene cluster is slightly higher than that of type I. Within the subgroups of type I and type II ColE2 sequences, the levels of diversity are about the same. Moreover, polymorphic sites are distributed evenly among the three genes of the colicin gene cluster and also within each gene for type I ColE2 gene clusters but not for type II ColE2 gene clusters. This heterogeneity in substitution levels among type II E2 plasmids may be explained by multiple recombinational origins for type II ColE2 plasmids, i.e., different subgroups of type II ColE2 plasmids may have different ancestors generated by different recombinations between type I ColE2 and ColE7 plasmids. The higher level of polymorphism for type II ColE2 may reflect polymorphism among the different ColE7 origins. In fact, this view is reinforced by closer examination of the distribution of type II ColE2 polymorphism. A higher level of polymorphism is observed in the colicin gene relative to the remainder of the colicin gene cluster. Further, within the colicin gene, the polymorphic sites are clustered at the 5′ end. Thus, most of the polymorphisms for type II ColE2 are within the region that appears to have recombined in from ColE7.

These data support the subsequent conclusion that there have been multiple origins for type II ColE2 plasmids. An alternative explanation could involve a single origin for type II ColE2 plasmids and a higher evolutionary rate in the 5′ end of the colicin gene relative to the 3′ end. However, previous studies on colicin E plasmids suggest that the 5′ end of the colicin gene is not less constrained than the remainder of the gene cluster (Riley 1993*a*). Thus, the single-origin hypothesis is less likely to be the explanation for the observed polymorphism patterns.

The ColE2 gene cluster is closely related to that of ColE9. Their immunity regions can be unambiguously aligned. Similar to the situation observed for ColIa, the synonymous and nonsynonymous polymorphisms are evenly distributed along the whole gene cluster for type I ColE2. When both ColE2 types are combined in the 3′ end region of the gene cluster, where the sequence originating from ColE7 is excluded, the synonymous and nonsynonymous polymorphisms are also evenly distributed along that region. This pattern of polymorphism is quite different from the pattern of divergence observed between ColE2 and its closely related ColE9 gene cluster, where the divergence is clustered in the immunity region. These results are not consistent with the neutral theory, which predicts that levels of polymorphism should be positively correlated with levels of divergence, i.e., regions that diverge more rapidly should accumulate polymorphism more rapidly (Kimura 1983).

Moreover, HKA tests for the 5′ end of the colicin gene versus the immunity region detect statistically significant departures from neutral predictions, indicating that one of the compared regions evolves in a nonneutral fashion. An MK test applied to the immunity region also allowed rejection of the null hypothesis of neutrality. Thus, the HKA and MK tests in concert argue that at least one portion of the colicin gene cluster (the MK test would argue for the immunity region) has experienced nonneutral evolution.

It should be noted that it may not be appropriate to apply the HKA and MK tests to these data. First, there is no way to assess whether *E. coli* is at equilibrium with respect to the forces of mutation and drift. Given the clear geographic subdivision in the plasmid populations sampled, it is likely that the plasmid pool of E2 in *E. coli* does not represent a freely recombining population of plasmids. This would violate an assumption of most tests of neutrality.

However, it should be noted that even though there is obvious geographic structuring to the plasmid population, the subdivision is not complete. One E2 plasmid isolated 50 years ago in France is strikingly similar to the Australian E2 isolates. Further, it is clear that plasmids are moving in the *E. coli* population. We find evidence for recombination between E2 and E7 plasmids even though we currently do not find E2 and E7 in the same *E. coli* isolates.

At least two hypotheses could explain the observed disparity between patterns of polymorphism and divergence in E2 and E9 plasmids: recombination and selection. Multiple recombination events could explain the divergence patterns of ColE2/E9 and ColE3/E6 pairs. This hypothesis requires multiple recombination events comprising short regions of exchange centered on the immunity region (Riley 1993*a*). Although recombination has clearly played a role in the diversification of colicin types, it is unlikely that recombination alone can account for the observed patterns of divergence between colicins E2 and E9. First, it is unlikely that there has been enough time for multiple recombination events across the immunity region to have occurred given the near identity of the E2 and E9 flanking sequences, unless conjugation and recombination rates are many orders of magnitude higher in natural populations than anticipated (Gordon 1992). Second, given that the frequency of a specific colicin type is usually less than 1% in natural *E. coli* populations (Achtman et al. 1983; Riley and Gordon 1992), the availability of an appropriate recombination template for such repeated events is problematic. Finally, the specific interaction required between the immunity binding domains of the colicin and immunity proteins would prevent recombination events within the immunity region that break up the required protein interactions, again reducing the pool of potential functional templates. All of these features argue that it is unlikely that multiple recombination events across the immunity region could generate the observed patterns of divergence.

One alternative explanation for the observed lack of correspondence between the polymorphism and divergence of colicin E2 and E9 plasmids is the action of positive selection. As detailed in the introduction and previously (Riley 1993*a*, 1993*b*), selection for novel colicin types could result in clustered substitutions accumulating in the immunity regions of colicin gene clusters. This pattern of diversification would not necessarily be reflected in inflated levels of polymorphism in the same region. It is proposed that mutation, recombination, and drift would be the dominant evolutionary forces acting on the variation segregating within plasmid populations. However, according to the diversifying-selection hypothesis, recombinants between colicin types (or "species") in the immunity region would be eliminated due to the lethal effect of recombination disrupting immunity function. Thus, the forces acting at the population level and at the plasmid "species" level would be distinctly different and would result in different patterns of polymorphism and divergence.

The finding of a mosaic structure for type II ColE2 gene clusters provides further support for the action of positive selection. To explain why synonymous and intergenic sites within the immunity region experience elevated rates of substitution related to the flanking region, it was proposed that between events of diversification in immunity function, neutral substitutions accumulate randomly along the plasmid replicon. Recombination between closely related Col plasmids would release these neutral substitutions from their linkage with the selected sites and, therefore, homogenize the evolved and ancestral Col plasmids. However, due to the specific interaction required between the immunity protein and the immunity binding domain of the colicin protein, any recombination events within the immunity region that disrupt immunity function would be selected against (Riley 1993*a*). Thus, recombination would tend to occur only outside the immunity region. This prediction is precisely reflected in the mosaic structure of type II ColE2 gene clusters, which appears to have been created by recombination event(s) between ColE2 and ColE7 plasmids. It is found that in type II ColE2 gene clusters, both the immunity binding region of the colicin gene (3′ end) and the immunity gene originated from the ColE2 gene cluster, while the 5′ end of the colicin gene originated from the ColE7 gene cluster. The existence of an intact E2 immunity region within the mosaic structure of type II ColE2 gene clusters suggests that the specific interaction between the E2 immunity protein and the immunity binding domain of a colicin selects against recombination events within the immunity region in order to maintain the immunity function, as predicted by Riley (1993*a*).

As revealed by previous studies, mechanisms like horizontal transfer of colicin plasmids, inter- and intragenic recombination, transposition, and rapid diversification of immunity function through the action of positive selection are all responsible for the diversification of colicin gene clusters (Lau, Parsons, and Uchimura 1992; Riley 1993*a*, 1993*b*). Among these mechanisms, two are further supported by this study. The sequence polymorphism data of the ColE2 gene cluster suggest a role for positive selection in generating diversification of the immunity region of colicin gene clusters. The finding of a new type of ColE2 sequence with a mosaic structure and its recombinational origin indicate the important contribution of recombination to the diversification of other regions of the colicin gene cluster.

Acknowledgments

Special thanks to Etsuko Moriyama for her help in polymorphism data analysis and to David Gordon and Peter Lau for kindly providing the ColE2 strains. Thanks to members of the Riley group and to Julian Adams and an anonymous reviewer for their extensive comments on this manuscript. This work was supported by an NIH First Award (GM47471) and an NSF Young Investigator Award (DEB-9458247) to M.A.R. Additional support was provided by the Biospherics Institute of Yale University through a grant from the General Reinsurance Corporation.

LITERATURE CITED

ACHTMAN, M., A. MERCER, B. KUSECEK, A. POHL, M. HEUZENROEDER, W. AARONSON, A. SUTTON, and R. P. SILVER. 1983. Six widespread bacterial clones among *Escherichia coli* K1 isolates. Infect. Immun. **39**:315–335.

BERRY, A. J., W. A. JAMES, and M. KREITMAN. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics **129**:1111–1117.

CHAK, K., W. KUO, F. LIU, and R. JAMES. 1991. Cloning and characterization of the ColE7 plasmid. J. Gen. Microbiol. **137**:91–100.

COLE, S. T., B. SAINT-JOANIS, and A. P. PUGSLEY. 1985. Molecular characterization of the colicin E2 operon and identification of its products. Mol. Gen. Genet. **198**:465–472.

DYKES, G. A. 1995. Bacteriocins: ecological and evolutionary significance. Trends Ecol. Evol. **10**:186–189.

GORDON, D. M. 1992. The rate of plasmid transfer among *Escherichia coli* strains. J. Gen. Microbiol. **138**:17–21.

HUDSON, R. R., M. KREITMAN, and M. AGUADE. 1987. A test of neutral molecular evolution based upon nucleotide data. Genetics **116**:153–159.

KIMURA, K. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

LAU, P. C. K., M. PARSONS, and T. UCHIMURA. 1992. Molecular evolution of E colicin plasmids with emphasis on the endonuclease types. Pp. 353–378 *in* R. JAMES, C. LAZDUNSKI, and F. PATTUS, eds. Bacteriocins, microcins and lantibiotics. Springer-Verlag, Berlin and Heidelberg.

LEWIS, M. J., and B. A. D. STOCKER. 1965. Properties of some group E colicine factors. Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. 1 Orig. **196**:173–183.

MALES, B. M., and B. A. D. STOCKER. 1980. *Escherichia coli* K317, formerly used to define colicin group E2, produces colicin E7, is immune to colicin E2, and carries a bacteriophage-restricting conjugative plasmids. J. Bacteriol. **144**: 524–531.

MCDONALD, J. H., and M. KREITMAN. 1991. Adaptive protein evolution at the *adh* locus in *Drosophila*. Nature **351**:652–654.

PILSL, H., and V. BRAUN. 1995. Novel colicin 10: assignment of four domains to TonB- and TolC-dependent uptake via the Tsx receptor and to pore formation. Mol. Microbiol. **16**: 57–67.

PUGSLEY, A. P., and B. OUDEGA. 1987. Methods for studying colicins and their plasmids. Pp. 105–161 *in* K. G. HARDY, ed. Plasmids, a practical approach. IRL Press, Oxford.

RILEY, M. A. 1993*a*. Positive selection for colicin diversity in bacteria. Mol. Biol. Evol. **10**:1048–1059.

———. 1993*b*. Molecular mechanisms of colicin evolution. Mol. Biol. Evol. **10**:1380–1395.

RILEY, M. A., and D. M. GORDON. 1992. A survey of col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of col-plasmid lineages. J. Gen. Microbiol. **138**:1345–1352.

RILEY, M. A., Y. TAN, and J. WANG. 1994. Nucleotide polymorphism in colicin E1 and Ia plasmids from natural isolates of *E. coli*. Proc. Natl. Acad. Sci. USA **91**:11276–11280.

ROOS, U., R. E. HARKNESS, and V. BRAUN. 1989. Assembly of colicin genes from a few DNA fragments: nucleotide sequence of colicin D. Mol. Microbiol. **3**:891–902.

TAGG, J. R., A. S. DAJANI, and L. W. WANNAMAKER. 1976. Bacteriocins of a gram positive bacteria. Bacteriol. Rev. **40**: 722–756.

TORREBLANCA, M., I. MESEGUER, and A. VENTOSA. 1994. Production of halocin is a practically universal feature of archeal halophilic rods. Appl. Microbiol. **19**:201–205.

WATSON, R., W. ROWSOME, J. TSAO, and L. P. VISENTIN. 1981. Identification and characterization of Col plasmids from classical colicin E-producing strains. J. Bacteriol. **147**:569–577.

WATSON, R., T. VERNET, and L. P. VISENTIN. 1985. Relationships of the Col plasmids E2, E3, E4, E5, E6 and E7: restriction mapping and colicin gene fusion. Plasmid **13**:205–210.

JULIAN P. ADAMS, reviewing editor

Accepted March 6, 1997