Positive Selection for Colicin Diversity in Bacteria[1]

Margaret A. Riley
Department of Biology, Yale University

To examine the hypothesis that colicin proteins are subject to diversity-enhancing selection, we studied the rates of synonymous, nonsynonymous, and intergenic nucleotide substitution in three pairs of closely related colicin clusters. The results indicate that the immunity gene and the immunity-binding domain of the colicin gene, which interact to provide specific immunity from the lethal action of the colicin toxin, accumulate substitutions at synonymous and nonsynonymous sites several times more rapidly than does the remainder of the colicin cluster. We suggest that this increased level of divergence, centered at the immunity protein, may be the result of the combined action of recombination and positive selection acting to increase colicin diversity in natural populations of *Escherichia coli*.

Introduction

Colicins are toxic proteins produced by and active against *Escherichia coli* and related bacteria. More than 20 colicins have been described in *E. coli*, distinguished by the absence of cross-immunity between the producing strains (Fredericq 1957; Nomura 1967; Pugsley 1985). Colicins are encoded on a diverse group of colicin plasmids (Col plasmids), which usually encode at least two additional colicin-related proteins: an immunity protein, conferring specific immunity to the host and against that colicin, and a lysis protein, involved in lysing the host cell and in releasing colicin into the environment. This group of genes is called a "colicin cluster."

Colicins are abundant in natural populations of *E. coli*. Thirty percent of the 72 strains in the ECOR collection, which comprises *E. coli* isolates from man and domestic and zoo animals, possess one or more colicins (Riley and Gordon 1992). Although the ecological role of colicins is unclear, several studies suggest that they play a role in bacterial competition and invasion (Branche et al. 1963; Ikari et al. 1969; Kelstrup and Gibbons 1969; Smith 1974; Hardy 1975; Chao and Levin 1981).

Under conditions of stress, a small fraction of colicinogenic cells are induced to produce colicin and lysis proteins, an action mediated by the SOS response. The production of lysis protein results in the lysis of the host cell and in the release of colicin. The colicin protein attaches to specific cell surface receptors on nearby *E. coli* cells and kills the invaded cell by one of four mechanisms, which include DNA and rRNA degradation. If the invaded cells possess the same Col plasmid, they will be immune to the effects of the colicin, through a specific interaction between the immunity and colicin proteins (Pugsley 1984).

Numerous colicins and their associated immunity and lysis genes have been sequenced, and the biochemical details of the encoded proteins have been deduced

1. Key words: diversity-enhancing selection, synonymous substitution, nonsynonymous substitution, colicin evolution.

Address for correspondence and reprints: Margaret A. Riley, Department of Biology, Yale University, New Haven, Connecticut 06511.

Mol. Biol. Evol. 10(5):1048–1059. 1993.

0737-4038/93/1005-0009$02.00

(reviewed in Konisky 1982; Luria and Suit 1987). These DNA and protein sequences have been employed in investigations into the evolutionary origins and phylogenetic relationships of colicin and related proteins recently reviewed by Riley (accepted). Such studies reveal five families of related colicin proteins, with little or no sequence similarity observed between them.

The level of sequence similarity observed within a colicin family is variable (Riley, accepted). For example, colicins Ia and Ib belong to the same colicin family, and their DNA sequences differ by only 15% over a 3.7-kb region that includes the colicin and immunity genes as well as two open reading frames (ORFs). In contrast, colicin E1, considered a member of the same colicin family, can be aligned with colicins Ia and Ib only over the C-terminal 300 residues of the colicin protein and shows only 37% amino acid similarity in this restricted region. No further amino acid or DNA sequence similarity is observed between the E1 cluster and the Ia/Ib clusters. Further, if comparisons are restricted to closely related members of a colicin family, in which reliable DNA sequence alignments are possible, the level of similarity is observed to vary considerably across the colicin cluster. The immunity gene and the immunity-binding portion of the colicin gene accumulate substitutions many times more rapidly than does the remainder of the cluster (Riley, accepted).

In a recent review of these data, Riley (accepted) hypothesized that the rapidly evolving immunity gene and colicin-binding region may be under the influence of positive selection for increased immunity diversity. In order to address this hypothesis, I provide here an analysis of the levels and patterns of divergence between closely related colicin, immunity, and lysis genes. These data confirm the original observation by Riley (accepted) and provide further insight into the evolutionary mechanisms promoting colicin diversity in *E. coli*.

Material and Methods

Table 1 lists the colicins included in this study and indicates the killing function of the colicins, their modes of entry into the cell, and the size of plasmid on which they are found. DNA and protein sequence alignments were made by the CLUSTAL multiple-alignment program of Higgins et al. (1992). DNA sequence divergence was estimated after correction, for multiple substitution events, by the method of Jukes and Cantor (1969).

Table 1
Colicin, Lysis, and Immunity Sequences Included in the Present Study

	Mode of Action[a]	Colicin Receptor	Plasmid Size[b]	DNA Sequences[c]			Reference[d]
				Colicin	Lysis	Immunity	
E2 ...	3	BtuB	LMW	+	+	+	Cole et al. 1985
E3 ...	2	BtuB	LMW	+	+	+	Masaki and Ohta 1985
E6 ...	2	BtuB	LMW	+	+	+	Akutsu et al. 1989
E9 ...	3	BtuB	LMW	+	+	+	Curtis et al. 1989
Ia ...	1	Cir	HMW	+		+	Mankovich et al. 1986
Ib ...	1	Cir	HMW	+		+	Mankovich et al. 1986

[a] 1 = formation of channels in cytoplasmic membrane; 2 = 16s rRNA cleavage; and 3 = nonspecific DNA degradation.
[b] LMW = low-molecular-weight plasma; and HMW = high-molecular-weight plasmid.
[c] DNA sequences available from the literature.
[d] For DNA sequences.

Results and Discussion

Figure 1 illustrates the genetic organization of the three pairs of colicins (E3/E6, E2/E9, and Ia/Ib) included in this investigation. These colicins were chosen because each pair is closely related, allowing reliable alignments of long stretches of DNA, and they represent three levels of protein and DNA sequence divergence. Colicins E3 and E6 are 96% identical in DNA sequence over a 2.9-kb region that includes the colicin, lysis, and two immunity genes. Colicins E2 and E9 are 90% identical in sequence over a 2.2-kb region that includes the colicin, immunity, and lysis genes (Riley, accepted). Colicins Ia and Ib are identical at 85% of their base pairs over a 3.7-kb region that includes the colicin and immunity genes as well as two ORFs (Mankovich et al. 1986).

The average number of substitutions per site, corrected for multiple substitution events by the Jukes and Cantor (1969) method, is given for each pair of colicins, in table 2. Total, synonymous, and nonsynonymous sites are considered separately. For each pair of colicins, clear differences in the level of substitution are observed across the regions compared. In each case, the level of substitution per site peaks in the region encompassing the 3′ end of the colicin gene, which encodes the immunity-binding function, through the immunity gene. This segment of DNA is referred to as the "immunity region." Figure 2 illustrates this variable pattern of substitution across each of the colicin pairs. The same pattern of increased substitution in the immunity region is observed for total, synonymous, and nonsynonymous sites.

The significance of the observed differences in substitution levels across the regions compared was assessed with *G*-tests for homogeneity (table 3). This form of a *G*-test (Kreitman and Hudson 1991), which includes evolutionary variance as well as sampling variance and is thus a conservative test, is expected to be approximately χ^2 distributed when there is both no recombination within the regions and free recombination between the regions. The *G*-tests treat total, synonymous, and nonsynonymous sites separately and divide the sequences into functional blocks as delineated in table 1. All *G*-tests were significant at the 0.001 level, indicating significant departure from

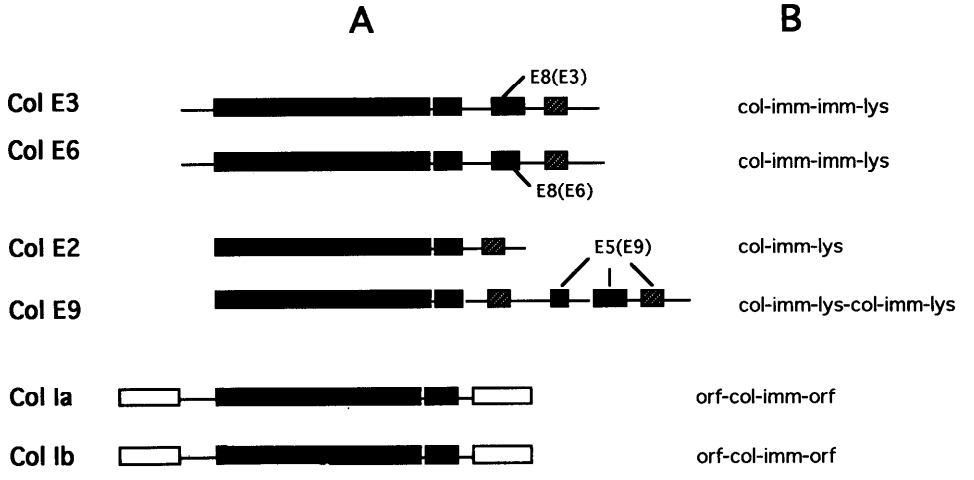


FIG. 1.—Colicin cluster organization. A, Colicin clusters drawn to scale. Rectangles indicate coding sequences, and connecting lines indicate spacer sequence. B, Gene order for each cluster. orf = open reading frame; col = colicin gene; imm = immunity gene; and lys = lysis gene. An additional immunity gene E8(E3) is located in the E3 and E6 clusters. An additional colicin, immunity, and lysis gene, E5(E9), is located in the E9 cluster.

Table 2
Average Number of Substitutions between Colicins

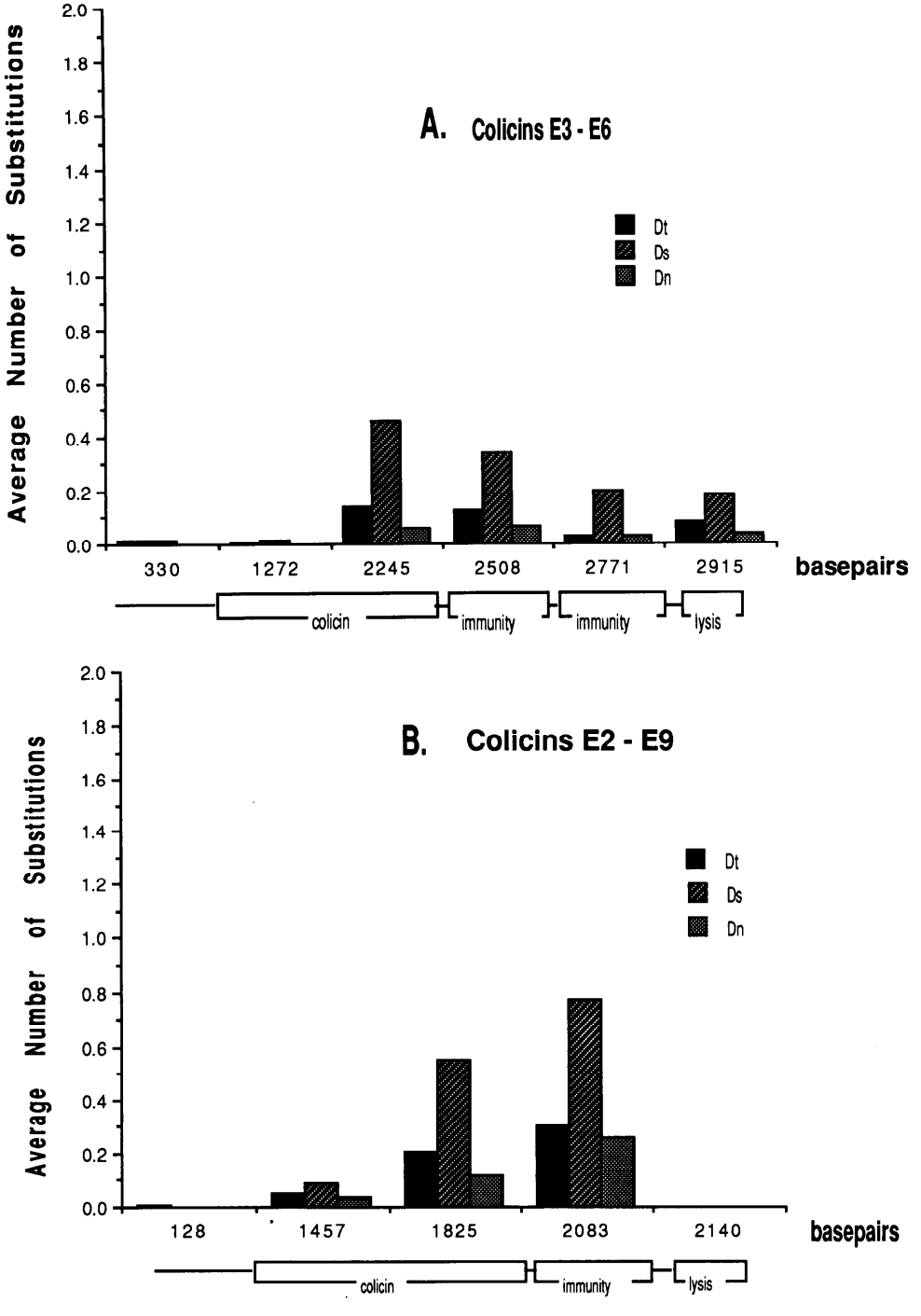
Region[a]	Average No. of Substitutions per Site[b]		
	Total	Synonymous	Nonsynonymous
A. Colicins E3 and E6:			
5′ Flanking	0.012 ± 0.00		
5′ Colicin	0.004 ± 0.00	0.015 ± 0.00	0.000 ± 0.00
3′ Colicin	0.143 ± 0.00	0.457 ± 0.01	0.058 ± 0.00
Intergenic	0.033 ± 0.00		
Immunity	0.126 ± 0.00	0.339 ± 0.02	0.066 ± 0.00
Intergenic	0.024 ± 0.00		
Immunity	0.030 ± 0.00	0.200 ± 0.00	0.030 ± 0.00
Intergenic	0.036 ± 0.00		
Lysis	0.079 ± 0.00	0.182 ± 0.01	0.038 ± 0.00
B. Colicins E2 and E9:			
5′ Flanking	0.008 ± 0.00		
5′ Colicin	0.051 ± 0.05	0.090 ± 0.20	0.037 ± 0.04
3′ Colicin	0.208 ± 0.00	0.556 ± 0.01	0.118 ± 0.00
Immunity	0.776 ± 0.03	0.258 ± 0.00	0.360 ± 0.00
Intergenic	0.048 ± 0.00		
Lysis	0.000 ± 0.00	0.000 ± 0.00	0.000 ± 0.00
C. Colicins Ia and Ib:			
5′ ORF	0.001 ± 0.00	0.00 ± 0.00	0.002 ± 0.00
Intergenic	0.023 ± 0.00		
5′ Colicin	0.013 ± 0.00	0.048 ± 0.00	0.005 ± 0.00
3′ Colicin	0.560 ± 0.00	1.771 ± 0.06	0.424 ± 0.00
Intergenic	1.134 ± 0.18		
Immunity	1.281 ± 0.02	1.144 ± 0.06	1.335 ± 0.03
Intergenic	0.213 ± 0.00		
3′ ORF	0.020 ± 0.00	0.039 ± 0.00	0.013 ± 0.00
3′ Flanking	0.048 ± 0.00		

[a] Regions correspond to functional domains delineated in fig. 1.
[b] Corrected by the Jukes and Cantor (1969) method.

the null hypothesis of equal substitution rates across the gene clusters. The *G*-tests revealed significant deviations in substitution rates for total, synonymous, and nonsynonymous sites.

How do we account for the varying levels of substitution across the colicin clusters? Variable rates of nonsynonymous substitution in the different genes may simply reflect the different levels of functional constraint experienced by the encoded colicin, immunity, and lysis proteins. According to the neutral mutation hypothesis (Kimura 1983, pp. 149–214), proteins that experience more stringent functional constraints will accumulate neutral nonsynonymous substitutions more slowly than will proteins—or domains within a protein—experiencing fewer constraints. However, relaxing functional constraints at the protein level—and, therefore, at nonsynonymous sites—should not effect the rate of substitution at synonymous and intergenic sites (Kimura 1983, pp. 149–214). Thus, any explanation for the increased level of substitution in the immunity region must account for the increased substitution rates at apparently functionless sites (i.e., synonymous and intergenic positions).

Figure 3 provides a schematic alignment of 2,664 bp between colicins E3 and E6, including the entire colicin, immunity, and lysis gene region. This figure illustrates



the clustered nature of the substitutions observed between these two colicins. Outside the immunity region the sequences are nearly identical, and the few differences encountered are apparently randomly distributed across the sequences. Within the immunity region the differences are clustered into pockets of substituted sites separated by long stretches of identical base pairs. This same pattern is observed in an alignment

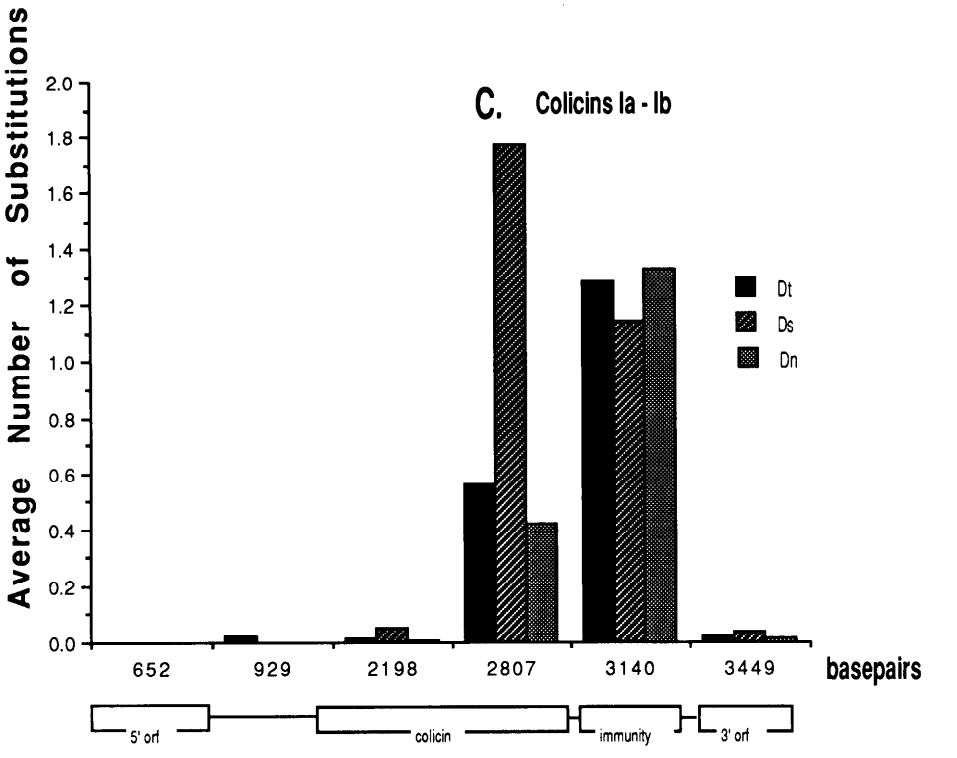


FIG. 2.—Distribution of nucleotide substitutions along the colicin cluster region for three pairs of colicins. The average number of substitutions, corrected for multiple substitution events by the Jukes and Cantor (1969) method, is plotted for the functional domains depicted along the *Y*-axis. Total, nonsynonymous, and synonymous sites are considered separately.

of colicins E2 and E9 (Curtis et al. 1989); however, the number of substitutions within each cluster in the immunity region is increased. In the Ia and Ib alignment (Mankovich et al. 1986), the immunity and colicin-binding region of the aligned sequences has accumulated even more substitution, to the point that synonymous sites have reached saturation and nonsynonymous substitutions are close to saturation.

In the case of Ia/Ib a single recombination event between either colicin and some highly divergent colicin cluster serving as the template could produce the observed marked transition between similar and divergent sequences across the cluster. The effect of recombination on the different classes of sites within the immunity region would be the same; that is, all would show elevated levels of substitution relative to the regions that have not experienced recombination. This same explanation will not account for the pattern of substitution observed in the E3/E6 or E2/E9 alignment. Multiple recombination events, comprising short regions of exchange, superimposed across the immunity region of E3/E6 and E2/E9 could result both in increased levels of divergence in this region, relative to the 5′ and 3′ regions, and in the clusters of divergent sequence.

In all scenarios involving recombination, the sequences must recombine within the colicin gene, in frame, to introduce a novel immunity-binding function. Further, in the case of E3/E6 and E2/E9, multiple recombination events are required that must occur over a very short time frame, to account for the near identity of the

Table 3
***G*-tests of Homogeneity, in Substitution Levels, across Colicin Sequences**

	Average No. of Substitutions per Site		
	Total	Synonymous	Nonsynonymous
A. E3 and E6:			
G	187.26	117.53	38.23
Degrees of freedom	7	4	4
P	$\ll 0.001$	$\ll 0.001$	$\ll 0.001$
B. E2 and E9:			
G	206.60	68.56	81.64
Degrees of freedom	5	3	3
P	$\ll 0.001$	$\ll 0.001$	$\ll 0.001$
C. Ia and Ib:			
G	1,134.27	266.45	785.49
Degrees of freedom	8	4	4
P	$\ll 0.001$	$\ll 0.001$	$\ll 0.001$

flanking sequences. If the recombination rate in bacteria is on the order of—or, more likely, less than—the mutation rate (Whittam and Ake 1992) during the time required for multiple successful recombination events to have occurred within the immunity region, then the synonymous and intergenic regions outside this region should have accumulated neutral substitutions. However, in the three comparisons presented here, the sequences flanking the immunity region have significantly reduced levels of synonymous and intergenic substitution. Given that (*a*) Col plasmids are found in $\leq 30\%$ of natural isolates of *Escherichia coli* (Riley and Gordon 1992), (*b*) estimates of plasmid transfer in *E. coli* are on the order of 10^{-12}–10^{-18} transfers/cell/generation in natural isolates (Gordon 1992), and (*c*) estimates of the recombination rate in *E. coli* are on the order of the mutation rate, 10^{-9} (Whittam and Ake 1992), recombination seems an unlikely explanation for the patterns of divergence observed.

DNA sequences that stimulate genetic recombination have been identified in a number of organisms (Voelkel et al. 1987). It is possible that some as yet uncharacterized sequences are acting to increase the rate of recombination in the immunity-gene region. However, these sequences would have to occur in unrelated colicin clusters that exhibit no protein or DNA sequence similarity. Even if the rate of recombination is increased, one still has to account for the availability of the appropriate template. In light of both the specific function of the immunity protein and the requirement for binding between it and the binding region of the colicin protein, there must exist limitations as to what can serve as donor sequences in this sort of "diversification by recombination" mechanism.

In addition, it is unlikely that some very restricted mutator phenomenon can account for the pattern observed. Mutator genes come in two basic forms (Cox 1976): (1) those that act genome wide, e.g., by increasing the rate of transitions over transversions, and (2) those that are sequence and/or locus specific, which generally involve transposable elements. The genome-wide mutators clearly cannot account for the pattern observed. Further, although effort has been made to identify transposons associated with colicin and related sequences, the presence of only a single case has been reported. A degenerate transposon-like structure has been observed in Col E9 (Curtis

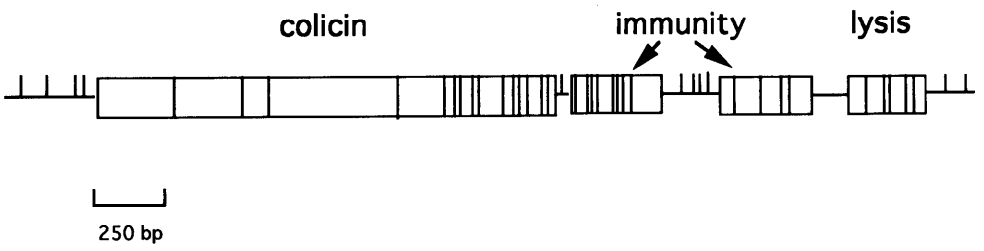


FIG. 3.—Schematic DNA sequence alignment of colicin E3 and E6. The aligned region includes the colicin, immunity, immunity (E8), and lysis genes. Rectangles indicate coding sequences, and connecting lines indicate intergenic regions. Vertical lines indicate the position of substitutions between E3 and E6 colicin clusters.

et al. 1989). There are no additional indications of transposons in the colicin cluster region.

Although the ecological role of colicins is unknown, it has been proposed that they provide strains of *E. coli* with a competitive advantage. Several studies support this hypothesis (Ikari et al. 1969; Smith 1974; Chao and Levin 1981; Bradley 1991); for example, the presence of a D-like colicin in symptomatic strains of enterohemorrhagic *E. coli* is believed to be responsible for the ability of these strains to invade the intestine and to displace the resident *E. coli* strains (Bradley 1991). Strains of *E. coli* carrying the Col E3 plasmid were shown to displace colicin-free strains when introduced at certain frequencies and grown on plates rather than in liquid media (Chao and Levin 1981). Each of these studies describes one or more conditions under which colicin production can provide a competitive advantage.

However, it is likely that there is a cost associated with colicin production. Surveys of colicinogenic bacteria in natural populations of *E. coli* suggest that populations do not become fixed for particular colicinogenic strains; rather, they suggest that these strains exist at low or intermediate frequencies (Riley and Gordon 1992). It is thus likely that there is a trade-off between the benefits and disadvantages of colicinogenicity, a trade-off that may differ for different populations of *E. coli* and for different environments.

If colicins do serve in bacterial competition, then a strain possessing a colicin to which other strains are not immune would be at a competitive advantage in certain situations. The production of such a novel immunity function could be accomplished in several ways. Studies on colicins E3, E6, and DF13 have permitted the localization of those residues likely involved in immunity binding (Lau and Condie 1989). These studies narrow the candidate amino acids to eight in the binding domain of the colicin protein and to nine in the immunity protein. This suggests that, at least for this group of colicins, relatively few substitutions in both the immunity protein and the binding region of the colicin protein are required to generate novel immunity functions.

The co-occurrence of mutations at one or more of these binding sites in a single colicin cluster could result in a strain that possesses a novel colicin phenotype. However, for the novel colicin phenotype to be viable, compensatory mutations will likely have to occur, in both the immunity gene and the binding region of the colicin gene, that retain binding function between the two encoded proteins. Further, only those strains that also retain immunity to the ancestral colicin are likely to survive; since the ancestral colicin will already be present in the population and without immunity to its ancestor, strains carrying the novel immunity function are likely to be killed by the ancestral

colicin. Several instances of partial immunity have been described between members of the closely related family of E-series colicins (Akutsu et al. 1989).

If the novel colicin phenotype involves mutations in the immunity gene that retain some affinity for the ancestral colicin, in addition to affinity for its own novel colicin protein, this strain will have an immediate and large advantage in environments where colicinogenicity provides a competitive advantage. Because of the interaction required between the immunity protein and the immunity-binding region of the colicin protein, the only way likely to produce these novel colicins is for selection for variants in one protein to occur simultaneously with selection for mutations in the other protein. I will call these "clusters" of selected sites, as it is essential that mutations occur at both sites if the evolved colicin is to be viable. Selection would thus act on clusters of variant sites rather than on individual sites.

In this scenario, positive selection will drive novel Col plasmids rapidly into the population, either until some new colicin phenotype evolves from it, in yet another round of diversification in immunity function, or until a strain carrying an unrelated colicin invades and possibly displaces the resident colicinogenic strain(s). The result of repeated rounds of this form of diversification of immunity function would be the accumulation of nonsynonymous substitutions in both the immunity gene and the binding region of the colicin gene, when closely related colicin clusters are compared.

In fact, this form of diversification in protein sequence is somewhat similar to that envisioned, by Hughes and Nei (1988, 1989), for the immunoglobin Vh loci in mammals. Those authors suggest that the high rate of nonsynonymous substitution observed at this locus is likely due to overdominant selection. New mutant alleles at the Vh loci enjoy a selective advantage until pathogens evolve a mechanism to avoid the new antibodies produced by the novel Vh alleles. Thus, there is a constant turnover of alleles that results in an increased level of nonsynonymous substitution at the Vh locus.

There are two significant differences between the form of selection envisioned for the Vh loci and that which may be acting to produce colicin and immunity diversity. First, in the former, selection is likely acting on a single site during a selective sweep; thus, only those nonsynonymous sites that are mutated diverge during repeated rounds of this sort of selection (Hughes and Nei 1989). In contrast, in the case of colicin diversification, clusters of sites within both the immunity gene and the binding region of the colicin gene are selected as a unit. This difference is critical to the explanation of why synonymous and intergenic sites within the immunity region experience elevated rates of substitution (see below). Second, the overdominant-selection hypothesis cannot be applied directly to haploid bacterial populations. A form of frequency-dependent selection, which will maintain colicins in intermediate frequencies once they have successfully invaded a bacterial population, may be a more appropriate selective model. In this case, novel colicins are initially selected for because of their competitive advantage over their ancestral forms. Having reached appreciable frequencies in the bacterial population, these colicins are maintained by frequency-dependent selection in those environments where colicin presence provides a competitive advantage to the colicinogenic bacterial cells.

When a cluster of selected sites is selectively driven into the population, the entire plasmid replicon will hitchhike with the cluster; that is, the plasmid will increase in frequency, as a unit. Further, surveys of Col plasmid distributions suggest that this plasmid will likely then be maintained in the population at some intermediate frequency until either a second novel colicin arises or an unrelated colicin invades the population.

During this time, neutral substitutions that accumulate on the plasmid replicon will be randomly distributed across the plasmid DNA. However, if recombination occurs between the selected Col plasmids and either the ancestral Col plasmid or any other closely related Col plasmids, this will release the neutral polymorphisms from their linkage with the selected sites. This process will tend to decrease the level of divergence between the ancestral and novel Col plasmids. However, the clusters of selected sites in the immunity and colicin-binding regions will still serve as traps for neutral polymorphism. Neutral mutations that occur within the trap are forced to maintain linkage with the cluster under selection, as the clusters must remain intact if the colicin is to be viable. The further removed the neutral polymorphisms are from the selected cluster, the greater the chance that recombination will disrupt the hitchhiking effect.

Unlike the case discussed above for "diversifying recombination," which requires homologous yet divergent templates, the present case requires closely related templates. This requirement is met if the Col plasmids are being maintained in high frequencies in the local population because of frequency-dependent selection. Not only will multiple copies of the selected Col plasmid serve as templates, but any closely related Col plasmids—e.g., the ancestral Col plasmid—will serve as well. Thus the population of suitable templates may be quite large in the local population.

This form of molecular hitchhiking bears some similarity with that reported to occur at the *Adh* locus in *Drosophila melanogaster* (Kreitman 1983). In this latter case, it is the presence of a balanced polymorphism that is apparently holding particular pairs of alleles in the population, with the concomitant result that neutral polymorphisms located near the selected site are also being maintained. While recombination whittles the level of neutral polymorphism down, the further removed it is from the selected site, those neutral polymorphisms that occur near the selected site are held in the population longer than can be explained simply by genetic drift (Kreitman 1988).

In contrast to the recombinational diversification hypothesis discussed above, diversifying selection can explain the slightly different patterns of divergence observed in the three pairs of colicins examined here. Recall from figure 2 that (*a*) E3/E6 has an extended tail of divergent sites centered around the immunity region, (*b*) E2/E9 has a slightly reduced tail of divergent sites centered around the immunity region, and (*c*) Ia/Ib shows an abrupt transition between the divergent immunity region and the flanking sequences. These different patterns may be explained by the increasing effectiveness of recombination in whittling away neutral polymorphism that falls outside the immunity-region trap. In this scenario E3 and E6 have been diverging for a short time, such that mutation and selection pressures exceed recombination pressures. Ia and Ib are more divergent overall, perhaps indicating that these two colicins have existed as separately selected units long enough for recombination to have become a more important force in whittling away the neutral substitutions outside the immunity-region trap.

As was noted by Tanaka and Nei (1989), very few cases of diversifying selection have been detected at the molecular level. However, in all instances the genes implicated are involved in mechanisms of defense (Tanaka and Nei 1989). The major histocompatibility complex, the Vh loci, and serine proteases of mammals are involved in defense against pathogenic microorganisms, while restriction-modification genes and colicins in bacteria are likely involved in either defense against invasion of phage or competition with other closely related strains or species of bacteria (Sharp et al. 1992; Riley, accepted).

LITERATURE CITED

AKUTSU, A., H. MASAKI, and T. OHTA. 1989. Molecular structure and immunity specificity of colicin E6, an evolutionary intermediate between E-group colicins and cloacin DF13. J. Bacteriol. **171**:6430–9436.

BRADLEY, D. E. 1991. Colicinogeny of 055 EPEC diarrhoeagenic *Escherichia coli.* FEMS Microbiol. Lett. **90**:49–52.

BRANCHE, W. C., V. M. YOUNG, H. G. ROBINET, and R. D. MASSEY. 1963. Effect of colicin production on *Escherichia coli* in the normal human intestine. Proc. Soc. Exp. Biol. Med. **114**:198–201.

CHAO, L., and B. LEVIN. 1981. Structured habitats and the evolution of anticompetitors in bacteria. Proc. Natl. Acad. Sci. USA **78**:6324–6328.

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

COX, E. C. 1976. Bacterial mutator genes and the control of spontaneous mutation. Pp. 135–156 *in* H. L. ROMAN, ed. Annual Review of Genetics, Palo Alto.

CURTIS, M. D., R. JAMES, and A. CODDINGTON. 1989. An evolutionary relationship between the colE5-099 and the colE9-J plasmids revealed by nucleotide sequencing. J. Gen. Microbiol. **135**:2783–2788.

FREDERICQ, P. 1957. Genetics of two different mechanisms of resistance to colicins: resistance by loss of specific receptors and immunity by transfer of colicinogenic factors. Pp. 323–335 *in* G. E. W. Wolstenholme and M. O'Connor, eds. Ciba symposium on drug resistance in microorganisms. Churchill, London.

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

HARDY, K. G. 1975. Colicinogeny and related phenomena. Bacteriol. Rev. **39**:464–515.

HIGGINS, D. G., A. J. BLEASBY, and R. FUCHS. 1992. CLUSTAL V: improved software for multiple sequence alignment. CABIOS **8**:189–191.

HUGHES, A. L., and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167–170.

———. 1989. Nucleotide substitution at class II MHC loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**:958–962.

IKARI, N. S., D. M. KENTA, and V. M. YOUNG. 1969. Interaction of the germfree mouse intestine of colicinogenic and colicin-sensitive microorganisms. Proc. Soc. Exp. Med. **130**: 1280–1284.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. Munro, ed. Mammalian protein metabolism. Academic Press, New York.

KELSTRUP, J., and R. J. GIBBONS. 1969. Inactivation of bacteriocins in the intestinal and oral cavity. J. Bacteriol. **99**:888–890.

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

KONISKY, J. 1982. Colicins and other bacteriocins with established modes of action. Annu. Rev. Microbiol. **36**:125–144.

KREITMAN, M. 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster.* Nature **304**:412–417.

———. 1988. Molecular population genetics. Oxf. Surv. Evol. Biol. **4**:38–60.

KREITMAN, M., and R. R. HUDSON. 1991. Inferring the evolutionary histories of the Adh and Adh-dup loci in Drosophila melanogaster from patterns of polymorphism and divergence. Genetics **127**:565–582.

LAU, P. C., and J. A. CONDIE. 1989. Nucleotide sequences from the colicins E5, E6 and E9 operons: presence of a degenerate transposon-like structure in the ColE9-J plasmid. Mol. Gen. Genet. **217**:269–277.

LURIA, S. E., and R. J. SUIT. 1987. Colicins and Col plasmids. Pp. 1615–1624 *in* F. C. e. a.

NEIDHARDT, ed. *Escherichia coli* and *Salmonella typhimurium:* cellular and molecular biology. American Society of Microbiologists, Washington, D.C.

MANKOVICH, J. A., C.-H. HSU, and J. KONISKY. 1986. DNA and amino acid sequence analysis of structural and immunity genes of colicins Ia and Ib. J. Bacteriol. **168**:228–236.

MASAKI, H., and T. OHTA. 1985. Colicin E3 and its immunity genes. J. Mol. Biol. **182**:217–227.

NOMURA, M. 1967. Colicins and related bacteriocins. Annu. Rev. Microbiol. **21**:257–284.

PUGSLEY, A. 1984. The ins and outs of colicins. Microbiol. Sci. **1**:168–175, 203–205.

———. 1985. *Escherichia coli* K12 strains for use in the identification and characterization of colicins. J. Gen. Microbiol. **131**:369–376.

RILEY, M. A. Molecular mechanisms of colicin evolution. Mol. Biol. Evol. **10** (accepted).

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.

SHARP, P. M., J. E. KELLEHER, A. S. DANIEL, G. M. COWAN, and N. E. MURRAY. 1992. Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteri. Proc. Natl. Acad. Sci. USA **89**:9836–9840.

SMITH, H. W. 1974. A search for transmissible pathogenic characteristics in invasive strains of *Escherichia coli:* the discovery of a plasmid-controlled lethal character closely associated, or identical, with colicin V. J. Gen. Microbiol. **83**:95–111.

TANAKA, T., and M. NEI. 1989. Positive Darwinian selection observed at the variable-region genes of immunoglobins. Mol. Biol. Evol. **6**:447–459.

VOELKEL, K., R. L. KEIL, and G. S. ROEDER. 1987. Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. Cell **48**:1071–1079.

WHITTAM, T. S., and S. E. AKE. 1992. Genetic polymorphisms and recombination in natural populations of Escherichia coli. Pp. 223–245 *in* N. TAKAHATA and A. G. CLARK, eds. Molecular paleopopulation biology. Japan Scientific Society Press, Tokyo.

MARTIN KREITMAN, reviewing editor

Received December 24, 1992; revision received April 30, 1993

Accepted April 30, 1993