Molecular Mechanisms of Colicin Evolution[1]

Margaret A. Riley
Department of Biology, Yale University

This review explores features of the origin and evolution of colicins in *Escherichia coli*. First, the evolutionary relationships of 16 colicin and colicin-related proteins are inferred from amino acid and DNA sequence comparisons. These comparisons are employed to detail the evolutionary mechanisms involved in the origin and diversification of colicin clusters. Such mechanisms include movement of colicin plasmids between strains of *E. coli* and subsequent plasmid cointegration, transposition- and recombination-mediated transfer of colicin and related sequences, and rapid diversification of colicin and immunity proteins through the action of positive selection. The wealth of information contained in colicin sequence comparisons makes this an ideal system with which to explore molecular mechanisms of evolutionary change.

Introduction

Colicins are toxic proteins produced by and active against *Escherichia coli* and related bacteria. Nineteen colicins have been described in *E. coli*, distinguished by the absence of cross-immunity between the producing strains (Fredericq 1957; Nomura 1967; Pugsley 1984, 1985). Although colicins differ in their precise mode of killing, they share (1) a similar genetic structure, including the usual presence of three colicin-related genes—a colicin, a lysis, and an immunity gene, termed a "colicin cluster"; (2) the lethality of colicin release from the cell; (3) the specific protection afforded by the immunity protein; and (4) carriage on plasmids (Hardy 1975; Pugsley 1984; Luria and Suit 1987).

The biochemistry and molecular biology of colicins have been studied in great detail (reviewed in Konisky 1982; Luria and Suit 1987). Indeed, numerous colicin clusters have been sequenced, and the biochemical details of the encoded proteins have been deduced. In contrast, little is known about their ecological function. Colicin production may be involved in bacterial competition and virulence determination (Branche et al. 1963; Ikari et al. 1969; Kelstrup and Gibbons 1969; Smith 1974; Hardy 1975; Chao and Levin 1981; Pugsley 1984); however, the data supporting these claims are limited and often conflicting (Hardy 1975; Pugsley 1984).

Regardless of their exact role, colicins occur in appreciable frequencies in *E. coli* populations. Thirty percent of the gram-negative bacteria isolated from a river in France were characterized as colicinogenic (Pugsley 1984). Similar estimates have been obtained in surveys of 372 strains of *E. coli* isolated from man and from domestic, zoo, and wild animals (Riley and Gordon 1992; author's unpublished data).

The most obvious phenotype associated with the presence of colicins in bacterial populations is that, under stress conditions, a small fraction of colicinogenic bacteria

1. Key words: molecular evolution, colicins, *Escherichia coli*, diversifying selection.

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

Mol. Biol. Evol. 10(6):1380–1395. 1993.

0737-4038/93/1006-0018$02.00

produce colicin proteins. The release of colicins from the cell, mediated by the action of the lysis protein, is lethal for both the producing cell and any bacteria not immune or resistant to that colicin. Immunity is conferred by the presence of the immunity protein, which inactivates the corresponding colicin by interacting with a specific region of the protein (Nomura 1967; Pugsley 1984).

Colicins gain entry into a susceptible cell by absorption to a specific outer-membrane receptor. At least five cell-surface receptors are recognized by colicins, including those encoded by *btuB, fepA,* and *ompF* genes. Most of these receptors are also involved in nutrient uptake (Konisky 1982), and many are used by bacteriophage to gain entry into the cell. Having gained access to the cell interior, colicins kill their targets in one of four ways: (1) formation of ion-permeable channels in the cytoplasmic membrane, (2) nonspecific degradation of cellular DNA, (3) inhibition of protein synthesis through the cleavage of 16s ribosomal RNA, and (4) cell lysis resulting from inhibition of peptidoglycan synthesis (Konisky 1982; Pugsley 1984).

Colicin clusters are found exclusively on plasmid replicons. The sizes of these colicin plasmids (Col plasmids) range from 6 kb (e.g., Col E1 plasmid) to >75 kb (e.g., Col Ia plasmid). Little is known about the evolutionary relationship between colicin clusters and their plasmids. However, limited data suggest that several colicin clusters maintain long-lived associations with particular plasmid backgrounds while others are quite mobile and are found associated with multiple, unrelated plasmids (Waters and Crosa 1991; Riley and Gordon 1992).

This review explores the molecular mechanisms of colicin cluster evolution. The first section details the evolutionary relationships of colicin, lysis, and immunity proteins as inferred from amino acid and DNA sequence comparisons. The second section employs these comparisons and the resulting phylogenetic framework to infer the molecular mechanisms involved in the origin and diversification of colicin clusters.

Colicin, Lysis, and Immunity Protein Relationships

The 16 colicin clusters included in this review are listed in table 1, which also indicates the killing function of the encoded colicins, their modes of entry into the cell, the size of plasmid on which the colicin clusters are found, and the availability of DNA sequences for the colicin, lysis, and immunity genes. The organization of each of the colicin clusters is depicted in figure 1. For ease of presentation, five groups of colicin clusters are distinguished in figure 1. The basis for this division is described below. Colicin B is included in groups 1 and 2 (see below). All DNA and protein sequence alignments (described below) were made with the CLUSTAL V multiple-sequence alignment program (Higgins et al. 1992). Protein-sequence-similarity assessment is based on the similarity matrix of Dayhoff et al. (1978). The phylogenetic trees described in the present study are unrooted and were done by using maximum parsimony, as implemented in PAUP (phylogenetic analysis using parsimony, version 3.0d; Swofford 1989).

Colicin Proteins

A feature common to all colicin proteins is their organization into four functional domains (fig. 2) (Ohno-Iwashita and Imahori 1980; Mock and Pugsley 1982; Pugsley 1984; Roos et al. 1989). Information required for movement across the cell membrane is found in the N-terminal 25% of the colicin protein. The central portion (~50% of the protein) is involved in receptor recognition and interaction, while the C-terminal

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(s)[d]
				Colicin	Lysis	Immunity	
A	1	OmpF/BtuB	LMW	+	+	+	Lloubes et al. 1984, Cavard et al. 1985
B	1	FepA	HMW	+		+	Schramm et al. 1987, 1988
D	2	FepA	LMW	+		+	Roos et al. 1989
DF13 ...	2	Iut	LMW	+	+	+	van den Elzen et al. 1983
E1	1	BtuB	LMW	+	+	+	Chan et al. 1985
E2	3	BtuB	LMW	+	+	+	Cole et al. 1985
E3	2	BtuB	LMW	+	+	+	Chak and James 1984, Toba et al. 1986
E5	2	BtuB	LMW	+	+	+	Curtis et al. 1989
E6	2	BtuB	LMW	+	+	+	Cole et al. 1985, Akutsu et al. 1989
E8	3	BtuB	LMW	+	+	+	Uchimura et al. 1987, Toba and Ohta 1988
E9	3	BtuB	LMW	+	+	+	James et al. 1987
Ia	1	Cir	HMW	+		+	Mankovich et al. 1986
Ib	1	Cir	HMW	+		+	Mankovich et al. 1986
M	4	FhuA	HMW	+		+	Kock et al. 1987, Oschlager et al. 1987
N	1	OmpF	LMW	+	+	+	Pugsley 1987, 1988
V	1	Cir	HMW	+	+		Gilson et al. 1990

[a] 1 = formation of channels in cytoplasmic membrane; 2 = 16s rRNA cleavage; 3 = nonspecific DNA degradation; and 4 = inhibition of murein biosynthesis.

[b] LMW = ~2–40-kb; and HMW = ~50–200-kb.

[c] DNA sequences available from the literature.

[d] For DNA sequences.

25% contains both the killing function and sequences that interact with the immunity protein (Konisky 1982).

Colicin proteins display varying levels of sequence similarity across these functional domains. For example, colicins Ia and Ib employ identical modes of entry into the cell and identical killing functions but have different immunity specificities (Mankovich et al. 1986). They differ at only 1% of the N-terminal 430 amino acid residues, whereas in the immunity binding region 64% of the C-terminal residues differ. As depicted in figure 1, with the exception of the immunity genes, the DNA sequence identity extends throughout the region sequenced. In figure 1 the shaded regions of the Ia and Ib clusters have 85% sequence identity. This high level of sequence identity argues that these two colicin clusters have recently shared a common ancestor.

Colicins A, B, E1, and N employ a killing function similar to that of Ia and Ib; however, genetic and biochemical studies indicate that they have different methods of entry into the cell and different immunity proteins. Comparisons of these colicin sequences reveal reduced levels of protein sequence similarity relative to that observed for colicins Ia and Ib, with similarity restricted to the C-terminal killing domain (Morlon

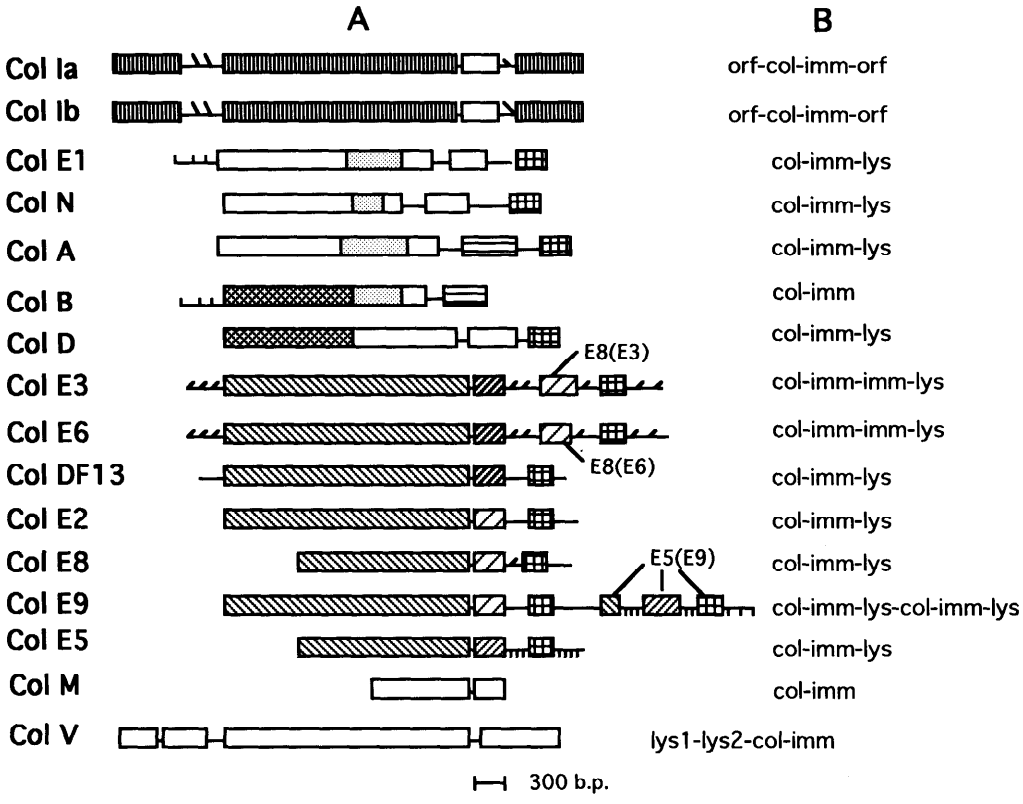

FIG. 1.—Colicin cluster organization. A, Colicin clusters drawn to scale. Regions of >80% DNA sequence identity between genes and spacer regions are indicated with matching hatching (angled and vertical lines). Rectangles indicate coding sequences, and connecting lines indicate spacer sequence. Bars on left border and corresponding numbers indicate colicin groups (as described in text). B, Gene order for each operon. orf = open reading frame; col = colicin gene; imm = immunity gene; and lys = lysis gene.

et al. 1983; Pattus et al. 1983; Lloubes et al. 1984; Pugsley 1987, 1988; Schramm et al. 1987; Parker et al. 1989). Levels of similarity in this domain range from 50% among colicins A, B, and N to 32% between colicin E1 and Ia or Ib. No similarity (i.e., >30%) is detected in the N-terminal two-thirds of these colicins. Further, with the exception of colicin B, no amino acid or DNA sequence similarity is detected between the group 1 colicin proteins and those of any other group (fig. 1).

A comparison of hydrophobicity plots for this group of colicins reveals that, even with the variable levels of amino acid sequence similarity observed, the proteins consist of a similar distribution of hydrophobic and hydrophilic domains in the C-terminus (Pugsley 1987; Schramm et al. 1987). The similarity in inferred protein structure in the killing domain suggests that the proteins have either (*a*) shared a more recent common ancestor in this region, relative to the more divergent N-terminal domains, (*b*) shared a distant common ancestor and experienced similar functional constraints associated with the killing function while the remaining functions have been free to diverge, or (*c*) converged on a similar array of amino acids in their killing domain.

A phylogeny for the group 1 colicins is given in figure 3A. The tree in figure 3A is based on the C-terminal 100 amino acids. Three clusters can be distinguished in the tree: colicin A; colicins E1, Ia, and Ib; and colicins B and N.

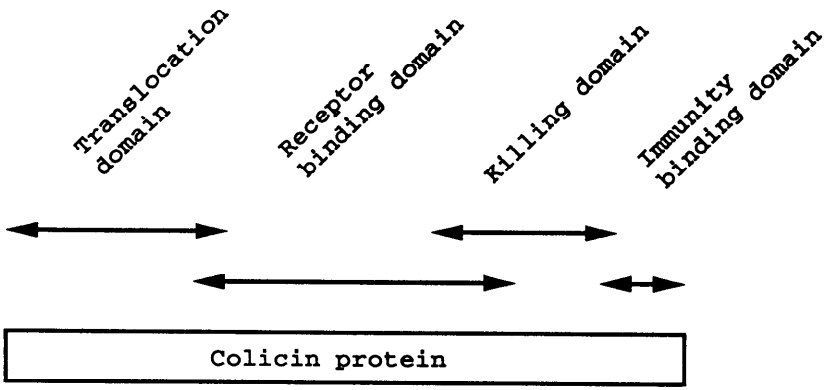

FIG. 2.—Colicin protein functional domains. Arrows above the box indicate the limits of each domain within the colicin protein. The names of the domains are indicated above the arrows.

A second group of colicins comprises colicins B and D. These two colicins share the same receptor specificity and are both dependent on *tonB* for uptake into the cell (Schramm et al. 1987; Roos et al. 1989). However, they differ both in their killing mechanisms and in their immunity specificities (table 1). The first and second domains of these colicins differ at only 10% of the N-terminal 400 residues (Roos et al. 1989), whereas the C-terminal domains cannot be aligned. At the DNA level, this similarity is reflected in 94% identity over the 5′ 1,000 nucleotides compared, with no similarity observed either 3′ to this region or with any other colicin proteins, save the previously mentioned similarity observed between colicin B and the group 1 colicins (fig. 1). This great DNA sequence identity argues that colicins B and D have shared a more recent common ancestor in the 5′ portion of their colicin genes than in the remainder of the clusters, presumably through some form of recombination or gene conversion. The presence of such long stretches of DNA sequence identity argues against either convergence onto similar function in the N-terminal domains or ancient shared common ancestry and subsequent divergence of the remainder of the cluster.

Colicins E2, E3, E5, E6, E8, E9, and DF13 make up a third group of colicins. The E-series members of this group share a similar mode of entry into the cell. Colicin DF13 differs in receptor recognition; however, it has the same killing function as do E3, E5, and E6: cleavage of ribosomal RNA (Curtis et al. 1989). Colicins E2, E8, and E9 kill by degrading DNA (Pugsley 1985; Akutsu et al. 1989; Curtis et al. 1989) (table 1). With one exception (DF13), those group 3 colicins for which there are complete amino acid sequences available (E2, E3, E6, and E9) are 97% identical in sequence over the first 400 amino acid residues. The C-terminal 100 amino acids are more divergent, averaging 49% similarity. DF13 is the most divergent member of this group, with an average of 50% amino acid similarity over the length of the protein, compared with the remaining E-series colicins (Cole et al. 1985; Masaki and Ohta 1985). The similarity detected in the group 3 colicin proteins is mirrored at the DNA level and extends into the 5′ and 3′ flanking genes and spacer regions of these operons (fig. 1).

A second incomplete colicin sequence was detected on the Col E9 plasmid (Curtis et al. 1989). This partial sequence shows similarity to a partial sequence available for colicin E5 and has been labeled "colicin E5(E9)" to distinguish it from the native E9 and the "true" E5 colicins (fig. 1).

The inferred phylogenetic relationship for the group 3 colicins is presented in

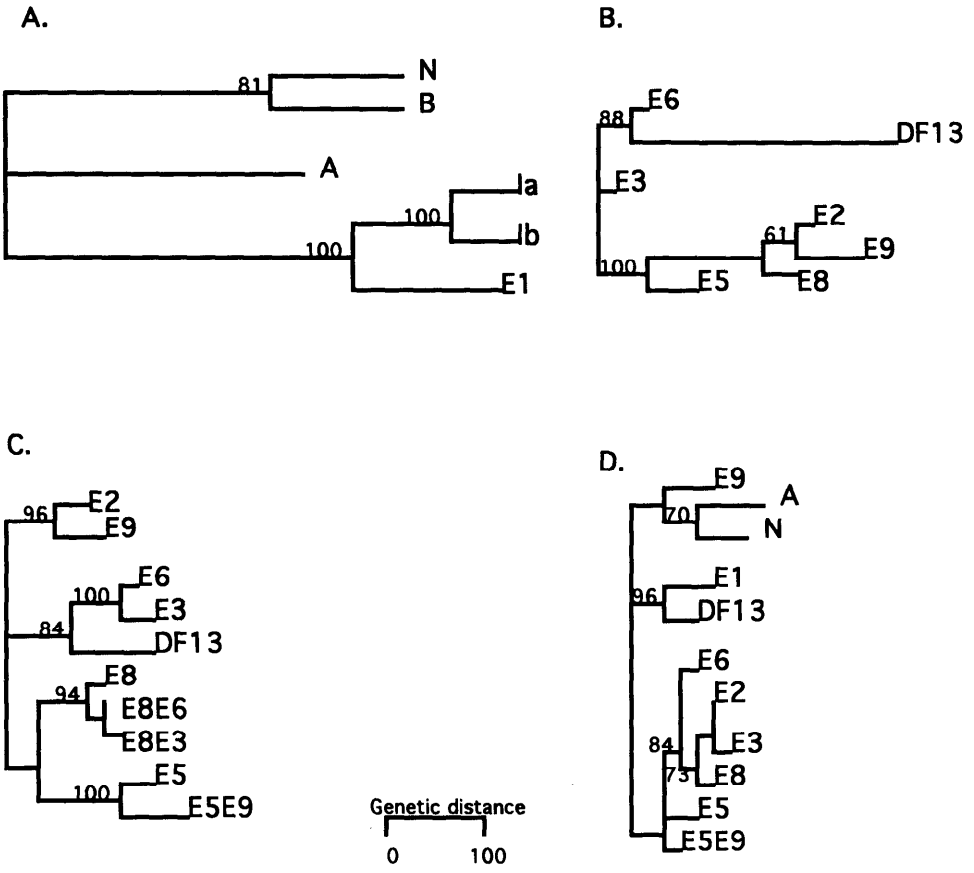

FIG. 3.—Maximum-parsimony trees relating colicin, lysis, and immunity proteins. Distance (in substitutions/100 sites) is indicated along the branches. Numbers with arrows indicate bootstrapping values based on 1,000 replications. A, Group 1 colicin proteins. B, Group 2 colicin proteins. C, Group 2 immunity proteins. D, Group 1 and 2 lysis proteins.

figure 3B. This tree distinguishes three subgroups: (1) colicins E3, E6, and E5; (2) colicin DF13; and (3) colicins E2, E8, and E9. The same branching patterns are obtained if colicin DF13 is removed from the analysis.

The two remaining colicins, M and V, bear no detectable amino acid sequence similarity (i.e., >30%) to each other or to any other sequenced colicin (Dreher et al. 1985; Kock et al. 1987; Gilson et al. 1990; Waters and Crosa 1991). Colicin M is unusual in being the only known colicin that kills by inhibiting murein biosynthesis (Dreher et al. 1985). Colicin V forms channels in the cytoplasmic membrane, as do the group 1 colicins; however, no amino acid or DNA sequence similarity is detected between the killing domain of colicin V and that of the group 1 colicins (Waters and Crosa 1991).

Several of the colicins share additional short stretches of similarity. For those colicins dependent on *tonB,* including colicins B, D, Ia, Ib, and M, a short sequence of four amino acids, called the "tonB box," has been determined in the amino terminus (Kock et al. 1987; Roos et al. 1989). This sequence is required for colicin uptake by the cell receptor (Kock et al. 1987; Roos et al. 1989). Colicin proteins A and E1 have short regions of similarity that have been implicated in *btuB* receptor recognition

(Morlon et al. 1983; Cole et al. 1985). However, other colicins that recognize the *btuB* receptor do not share this region of sequence similarity. Thus, colicins that recognize the same receptor may do so through interactions with different regions of the colicin and receptor proteins (Pugsley 1984; James et al. 1987).

In summary, examination of colicin proteins reveals a similarity in the organization of four functional domains within the proteins. Further, protein sequence comparisons reveal five distinct groups of colicins (group 1—A, E1, Ia, Ib, N, and B; group 2—B and D; group 3—E2, E3, E5, E6, E8, E9, and DF13; group 4—M; and group 5—V). Members of a group have either (*a*) shared a recent common ancestor (e.g., all of the E series colicins of group 3), (*b*) shared a distant common ancestor with portions of the colicin free to diverge while other portions have been constrained (e.g., DF13 compared with the remaining group 3 colicins), or (*c*) evolved a similar array of amino acids in particular domains (e.g., the killing domains of the group 1 colicins). Further, a mosaic pattern of colicin protein sequence similarity is observed with blocks of between-colicin similarity that correspond to functional domains of the proteins (e.g., the domains of colicins B and D).

Immunity Proteins

The immunity proteins confer specific immunity to the corresponding colicin. Mechanisms for the activity of several immunity proteins have been deduced (Hardy 1975; Konisky 1982; Lloubes et al. 1984; Pugsley 1984, 1988; Akutsu et al. 1989). These mechanisms involve a high affinity binding of the immunity protein to the immunity-binding region located in the C-terminus of the colicin protein. Three of the E-series Col plasmids have more than one immunity gene. The Col E3 plasmid and the Col E6 plasmid each possess a second immunity gene 5′ to their lysis genes, which encode proteins similar in sequence to the E8 immunity protein; these immunity genes are labeled, respectively, "immunity E8(E3)" and "immunity E8(E6)" (Chak and James 1984) (fig. 1). The Col E9 plasmid has a second immunity gene, 3′ to its partial E5(E9) colicin gene and similar in sequence to the E5 immunity gene; this second immunity gene is labeled "immunity E5(E9)" (James et al. 1987) (fig. 1).

The immunity proteins are the most variable of the colicin-related sequences. Most pairwise comparisons of immunity proteins reveal no detectable amino acid sequence similarity. The exceptions are immunity proteins A and B (Schramm et al. 1988) and the group 3 immunity proteins. At 39% of its residues immunity protein A is identical with immunity B, and it is similar at an additional 32%. At the DNA level, the identity between A and B is restricted to the immunity gene and to the killing domain of their colicin proteins (see above). The group 3 immunity proteins range from being nearly identical, e.g., the E5 and E5(E9) immunity proteins, which differ by a single amino acid replacement, to similar, e.g., E2, E8, and E9 immunity sequences, which are similar at 54%–97% of their residues.

Figure 3C provides a phylogeny for the group 3 immunity proteins, including the extra immunity proteins encoded in the E3, E6, and E9 colicin operons. The proteins are labeled according to the colicin operons with which they are associated. Visual comparison of group 3 colicin and immunity trees (fig. 3B and C) reveals a similar phylogeny for the two proteins.

In summary, this analysis reveals that the immunity proteins, although apparently quite similar in function, are more divergent in amino acid sequences than are the colicin and lysis proteins (see below). Even those colicin operons showing extensive protein and DNA sequence similarity, e.g., colicins Ia and Ib and the E-series colicins

of group 3, display little or reduced immunity-protein similarity relative to the remainder of the operons. As the phylogenies inferred for the group 3 immunity and colicin proteins agree in broad outline, it is likely that the different levels of divergence observed in these two proteins are an indication that they experienced quite different forms of natural selection, rather than being an indication that the colicin proteins share a more recent common ancestor than do the immunity proteins (see below).

Lysis Proteins

Lysis proteins facilitate the release of colicins through the cytoplasmic membrane (Pugsley and Schwartz 1984; Thum et al. 1988). Although the precise mechanism of cell lysis is unknown, all lysis proteins examined to date possess an N-terminal signal peptide and a cysteine residue located after the signal peptide cleavage site (Wu and Tokunaga 1986; Pugsley 1988). Immediately following the signal peptide is a conserved block of amino acids corresponding to the lipoprotein box found in precursors of bacterial lipoproteins (Vlasuk et al. 1983; Wu and Tokunaga 1986; Pugsley 1988). Several colicins apparently have no lysis-like protein, e.g., colicins Ia, Ib, B, and M (Thum et al. 1988). In contrast, colicin E9 has an additional lysis gene, lysis E5(E9), positioned 3′ to the E5-like immunity sequence (fig. 1). Colicin V possesses two proteins that function in cell lysis that are not homologous to the lysis proteins encoded by the remaining colicin operons (Waters and Crosa 1991).

With the exception of colicin V, lysis proteins average 45% amino acid sequence identity over all lysis protein comparisons (Cavard et al. 1985; Chan et al. 1985; Cole et al. 1985; Toba et al. 1986, 1988; Pugsley 1988; Thum et al. 1988). This comparison includes lysis proteins from different colicin cluster groups.

A maximum-parsimony tree relating the lysis proteins is given in figure 3D. The lysis genes are identified by the name of the colicin cluster with which they are associated. Colicin, immunity, and lysis sequences are not available for all of the colicin operons. If we allow for this limitation, with the exception of lysis DF13 the lysis proteins from group 1 and 3 colicin clusters are easily distinguished, as are their colicin and immunity sequences. Lysis DF13, a member of the group 3 colicin clusters, unexpectedly shares 76% of its residues with lysis E1 and, on average, only 44% with the group 3 lysis proteins. In addition, lysis proteins E2 and E3 are identical and thus fall together on the lysis tree, while the corresponding E2 and E3 colicin and immunity proteins are well separated on their respective trees. These discrepancies between the lysis, colicin, and immunity trees are suggestive of recombination events moving lysis genes between colicin clusters.

Two features of lysis-protein evolution are revealed in this analysis. First, lysis proteins are more conserved over all groups of colicin clusters when compared with the colicin and immunity proteins. Although certain pairs of colicin proteins may show higher levels of sequence similarity than does the average pair of lysis proteins, all of the lysis-protein sequences, save those of colicin V, can be aligned. In contrast, in only one case can colicin proteins from different groups be aligned. This suggests either that the lysis proteins have shared a more recent common ancestor than have the immunity and colicin proteins or that the lysis proteins have experienced very similar selective constraints resulting in the maintenance of a common core of amino acids. If we allow for the caveat that the colicin, immunity, and lysis trees were not constructed with the same set of sequences, the phylogenies indicate several instances in which recombination has apparently occurred between colicin clusters, resulting

in the movement of lysis genes between clusters. There are two instances (lysis E1/DF13 and lysis E2/E3) where significant disparity between the trees occurs.

Molecular Mechanisms of Colicin Cluster Evolution

Three conspicuous features of colicin cluster evolution are revealed by the sequence comparisons described above. First, although the phylogenies for the colicin, lysis, and immunity proteins broadly agree, several discrepancies between the lysis and colicin trees suggest that recombination or some other mechanism of rearrangement of the colicin-related genes occurs. The observation of additional colicin, lysis, and immunity sequences flanking the E-series colicin clusters supports this suggestion. Second, the patterns of domain-restricted sequence similarity and differences among colicin proteins implicate a more restricted form of transfer, i.e., intragenic recombination, generating novel colicin functions. Finally, in light of the varying levels of sequence conservation observed for the three proteins, natural selection has acted quite differently on these proteins.

The second half of this review explores the molecular mechanisms underlying these and related features of colicin evolution. These mechanisms are shown to include those acting (1) at the level of the plasmid, e.g., plasmid transfer between hosts, and subsequent plasmid cointegration; (2) at the level of the cluster, e.g., transposition- and recombination-mediated movement of genes and clusters; (3) at the level of the gene, e.g., intragenic recombination between colicin domains; and (4) at the level of the protein, e.g., positive selection for colicin diversification.

Col Plasmid Transfer

Many Col plasmids encode the information required for movement between strains of *Escherichia coli* (Pugsley 1984). With few exceptions, the high-molecular-weight (HMW) Col plasmids (e.g., Col Ia, Col V, and Col M plasmids) are self-transmissible through conjugation mechanisms. The low-molecular-weight (LMW) Col plasmids (e.g., Col E1, Col E2, and Col A plasmids) often encode mobilization factors that enable them to transfer during the conjugation process governed by HMW plasmids. Although the presence of these transfer systems implies that Col plasmids can move between host strains, there are few documented cases of Col plasmid movement outside laboratory settings. One study (Riley and Gordon 1992) suggests that certain classes of Col plasmids may exist in a long-term stable association with particular hosts. Thus, the frequency of successful Col plasmid movement may be quite low.

However, there are examples in which movement of Col plasmids between *E. coli* hosts is clearly implicated. Achtman et al. (1983) report several instances of multiple Col plasmids in a single *E. coli* host. Since these plasmids have also been found singly or with different plasmids in different hosts, these cases likely represent plasmid movement. A second example is provided by the Col B and M plasmids, which have been isolated both individually in separate hosts and as a cointegrate Col B:M plasmid (Thum et al. 1988). The process of cointegration requires that the plasmids coexist within the same host, and thus cointegration must have been preceded by the transfer of one or both plasmids.

The frequency with which transfer of Col plasmids occurs under laboratory settings, although variable, is quite high, on the order of 1×10^{-6} transfers/cell/generation (author's unpublished data). However, the rate of conjugation or mobilization under more natural conditions is likely to be many orders of magnitude lower (Gordon

1992). Thus mechanisms exist enabling Col plasmids to move between strains of *E. coli*, although the frequency of successful movement remains to be established.

Col Plasmid Cointegration

The co-occurrence of Col plasmids in a common host provides the opportunity for subsequent rearrangement at the plasmid level. One form of rearrangement involves the cointegration of unrelated plasmids. The process of cointegration is well documented in the evolution of antibiotic resistance–encoding plasmids (Broda 1979, p. 16). A similar process has been invoked to explain the presence of colicins B and M on the same plasmid (Thum et al. 1988). Although this is the only reported example of Col plasmid cointegration, previous studies were not designed to distinguish between cointegration and a more localized form of recombination. Therefore, the importance of this mechanism in the evolution of Col plasmids cannot be assessed. However, the fact that Col plasmid cointegrates may be selectively maintained is suggested by data that indicate that the cointegrate Col B:M plasmid is the dominant form of these Col plasmids found in natural populations of *E. coli* isolated from chickens (M. A. Riley and C. Buscher, unpublished data). Thus, cointegration may play a major role in generating novel combinations of colicin clusters on which selection can act.

Transposition-mediated Movement of Colicin Cluster Sequences

Two mechanisms exist enabling a more localized movement of colicin sequences between Col plasmids: transposition and recombination. Although efforts have been made to identify the presence of transposable elements flanking colicin clusters or individual colicin-related sequences, only a single case has been reported. A degenerate transposon-like structure has been observed in the Col E9 plasmid (Lau and Condie 1989). This element is located within the coding sequence of the partial E5(E9) colicin gene present on the Col E9 plasmid. The authors suggest that the Col E9 plasmid is the result of a cointegration event between a Col E9-like plasmid and a Col E5-like plasmid. The transposable element subsequently inserted into the E5 colicin gene, resulting in inactivation of the protein. Thus the result of the transposition event was not to create a new combination of colicin-related genes but to modify a preexisting combination. Despite efforts to find additional evidence of transposable elements, no further cases have been reported.

Recombination Within and Between Colicin Clusters

In contrast to transposition, recombination has clearly played a role both in the generation of new combinations of domains within colicin genes and in the generation of new combinations of genes within colicin clusters (Pugsley 1987; Roos et al. 1989). The most dramatic example of recombination involved in the generation of novel colicin proteins is provided by colicin B. The N-terminal 313 residues of colicin B are almost identical to those in colicin D (Roos et al. 1989). This similarity is reflected at the DNA level which, with 94% of the 5′ nucleotides identical, suggests that the two colicins shared a common ancestor on the order of 32 Mya, if we employ both an estimate of $0.1\bar{6}$ hits/synonymous site for these sequences and an estimate on the order of 1×10^{-9} substitutions/synonymous site/Myr for the rate of synonymous substitution in *E. coli* (Ochman and Wilson 1987). As a point of reference, *E. coli* and *Salmonella* are estimated to have diverged on the order of 130–160 Mya (Ochman and Wilson 1987). Although this estimate of colicin divergence is clearly quite rough,

it is useful when comparing the relative ages of different colicins and colicin-related sequences.

As sequences 5′ to the coding region in colicin D have not been determined, the 5′ boundary of this inferred recombination event between colicins B and D cannot be established. However, the distinction between similar and differing regions at the 3′ boundary is quite pronounced. The C-terminal sequences of these clusters bear no detectable similarity at either the protein level or the DNA level (fig. 1). This abrupt transition between similar and dissimilar sequences argues for the horizontal transfer of sequences between the Col B and D plasmids.

The pattern of recombination in colicin B is even more complex. Schramm et al. (1987) suggest that intergenic recombination will explain the origin of a 294-nucleotide sequence that is upstream of the structural genes of colicins B and E1 and that is 85% similar (fig. 1). The 5′ boundary of this stretch has not been established. The 3′ boundary occurs abruptly at the 5′ end of the colicin genes. Colicins E1 and B show no detectable similarity at the protein and DNA levels over the N-terminal two-thirds of their sequences. The recombination event involving colicins B and E1 is estimated to have occurred on the order of 32 Mya, i.e., quite close in time to the event involving colicins B and D.

Indications of further recombination events for colicin B have been described. Schramm et al. (1988) note that 52% of the nucleotides in the A and B immunity sequences are identical. This similarity extends through a short intergenic region and into the 3′ end of the colicin genes. On average, 65% of the nucleotides are identical throughout this region; however, numerous indels are required to obtain this level of identity. No additional regions of similarity are detected 5′ or 3′ to this block of sequence. The unique presence of such high levels of identity has been invoked as support for the idea of a common origin of the carboxy-terminal domain for these two colicins (Pugsley 1987; Schramm et al. 1988). Colicins A and B are estimated to have shared a common ancestor on the order of 95 Mya, placing this recombination event prior to those involving colicins D and E1.

Thus, colicin B is composed of three segments: the E1-like upstream region, the D-like N-terminal and central region, and the A-like carboxy-terminal region (Pugsley 1988; Schramm et al. 1988; Roos et al. 1989). Recombination is the most likely explanation for this patchwork of similarities and differences.

Recombination has also been invoked to explain the generation of new combinations of colicin, lysis, and immunity genes. The E-series colicins reflect the fluid nature of colicin cluster composition (Chak and James 1984; James et al. 1987; Chak et al. 1991). For example, colicin clusters E3 and E6 each have an additional immunity gene (Chak and James 1984, 1986), which shows 93% and 97% sequence identity, respectively, with the E8 immunity gene. This identity ends just 5′ to the immunity genes in both E3 and E6 but extends in the 3′ direction through the E3 and E6 lysis genes.

Several additional examples of likely recombination events exist. The lysis genes of colicins E1 and DF13 have 71% amino acid sequence identity nested within a several-kilobase-long region that cannot otherwise be aligned. Although lysis genes are, in general, highly conserved, the average pairwise comparison of lysis proteins reveals 45% amino acid sequence similarity. Thus, the lysis proteins for colicins E1 and DF13 clearly reflect a more recent common ancestry, on the order of 30 Mya.

A further example is provided by comparing the E5(E9) lysis gene on the Col E9 plasmid with the "true" colicin E5 lysis gene (fig. 1). In this example, the similarity

between the E9 and E5 clusters extends upstream from the E5(E9) lysis gene into a partial E5(E9) colicin gene and downstream through the remainder of the 888 bp sequenced. Eighteen nucleotide differences in total are detected over this entire region. As mentioned previously, Lau and Condie (1989) suggest that this similarity resulted from a fusion of E9-like and E5-like ancestral plasmids. However, a more localized form of recombination cannot be rejected. Further, they suggest that the lysis gene, which is so highly conserved between most colicin clusters, may serve as a hot spot for much of the observed recombination between colicin clusters. Thus recombination clearly plays a role in generating colicin-cluster and colicin-protein diversity. This finding suggests that, even if Col plasmid transfer is infrequent in the wild, there is sufficient movement for recombination between colicin clusters to play a dominant role in the origin of novel combinations of colicin, lysis, and immunity genes.

Diversifying and Purifying Selection

Colicin, immunity, and lysis genes display quite different levels of nucleotide and amino acid sequence divergence. In this section the importance of diversifying and purifying selection in producing these wide-ranging levels of diversity is discussed.

Colicin clusters Ia and Ib provide the most dramatic example of these two contrasting forms of selection. A comparison of the rates of divergence between the Ia and Ib clusters is provided in figure 4A. The average number of substitutions per site is indicated separately for coding and noncoding regions of the cluster and, within a gene, for synonymous and nonsynonymous sites. On the basis of the level of substitution at synonymous sites in the 5′ half of the colicin gene, it is estimated that Ia and Ib shared a common ancestor on the order of 8.7 Mya. However, during that relatively

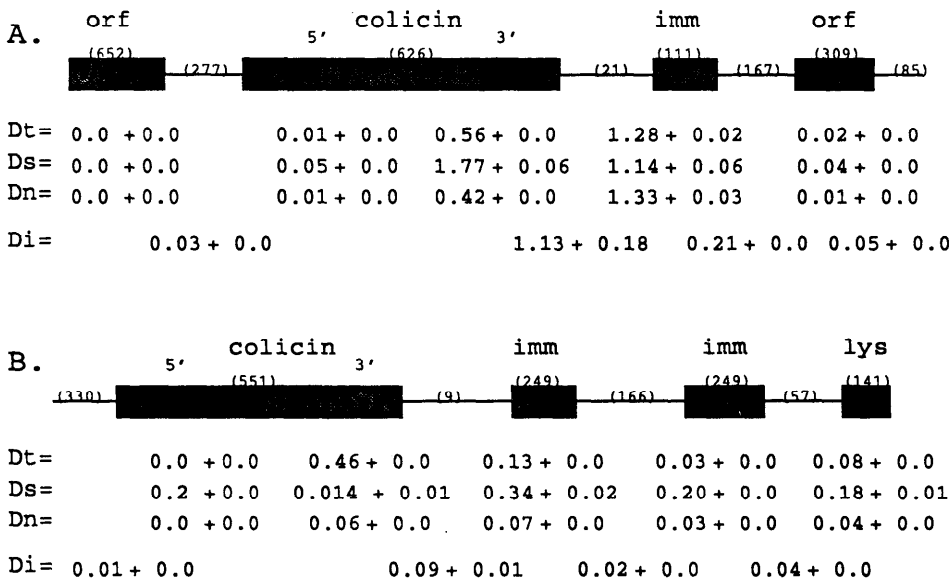

FIG. 4.—Comparison of substitution rates across colicin clusters. Dt = average no. of substitutions/site, averaged over all sites; Ds = no. of substitutions/site, averaged over synonymous sites; Dn = no. of substitutions/site, averaged over nonsynonymous sites; and Di = no. of substitutions/site, averaged over intergenic sites. All other abbreviations are as in fig. 1. Colicin clusters are not drawn to scale. Numbers in parentheses are the no. of base pairs compared (indels excluded). A, Colicin clusters Ia and Ib. B, Colicin clusters E3 and E6.

short period of time, the Ia and Ib immunity genes and the immunity-binding region of the colicin genes have diverged at a rate several orders of magnitude faster than has the remainder of this region (refer to fig. 4). For example, the level of synonymous substitution is roughly 30 times higher in the immunity gene than in either the 5′ half of the colicin gene or the 5′ and 3′ open reading frames (ORFs).

How do we account for these varying levels of substitution across the cluster? One hypothesis is that there is a relaxation of purifying selection in the immunity gene and in the immunity-binding region. However, this hypothesis does not account for the decreased level of substitution observed in noncoding sequences 5′ to the colicin gene and 3′ to the lysis gene, which should show a similar relaxation of function as is seen in the silent sites in both the immunity gene and the immunity-binding region of the colicin gene and in the noncoding region between them. Further, it does not make biological sense, as it has been shown that the immunity protein and the immunity-binding region of the colicin protein must interact to produce immunity—that is, functional constraints must exist in these proteins.

Closer examination of the Ia and Ib colicin cluster sequences reveals no sharp boundaries between these regions of high and low sequence divergence. Thus, a single recombination event introducing a novel sequence in one or both clusters is also not a sufficient explanation for the varying levels of divergence observed between the colicin and immunity proteins. Multiple recombination events could generate the observed gradual transition between conserved and divergent sequences. However, in light of the 99% DNA sequence identity observed between the 5′ and 3′ flanking ORFs, it is unlikely that there has been enough time for multiple recombination events to have occurred, unless conjugation and recombination rates in natural populations are at least three orders of magnitude higher than anticipated (Gordon, 1992). Further, it is also unlikely that some very restricted mutator phenomenon, acting over a narrow window, can account for the pattern observed, particularly as the identical pattern is observed in multiple, unrelated colicin clusters (see below).

The observed function of the immunity protein is to bind to the immunity-binding region of the colicin protein, providing specific immunity to the host cell. One can envision a form of positive selection that would preserve novel immunity variants as they arise in newly diverging colicin clusters. For example, in the early stages of divergence between Ia and Ib, mutations that generate a distinct immunity function in the Col Ia plasmid and/or the Col Ib plasmid would be selected for. Subsequent mutations in both the immunity gene and immunity-binding region of the colicin gene, mutations that exaggerate differences in immunity function and ensure proper binding to the colicin, would likewise be selected for. I will call these "clusters" of selected sites, as it is essential that they occur almost simultaneously to ensure proper immunity function. This form of positive selection would likely be quite strong during the early stages of diversification, when the colicin clusters are most similar, resulting in a very rapid accumulation of clusters of substitutions at nonsynonymous sites in both the immunity gene and immunity-binding region. Further, mutations that occur at synonymous and noncoding positions in this region will accumulate at elevated levels because of to a hitchhiking effect; that is, as clusters of mutations are held in the population by positive selection, neutral mutations in the region will accumulate because of their tight linkage to the selected sites (Strobeck 1983). This form of accelerated divergence is analogous to the diversifying selection hypothesized at histocompatability loci in mammals (Hughes and Nei 1988, 1989). This form of se-

lection cannot simply involve selection for rare variants, as this process would result in decreased, rather than increased, levels of divergence at silent and noncoding sites.

A second example of this same form of positive selection is illustrated by colicin clusters E3 and E6, which are 93% identical over a region including the colicin, immunity, and lysis genes. The E3/E6 sequences display the same pattern of increased levels of divergence in the immunity gene and immunity-binding region as are observed for Ia and Ib (fig. 4B). Indeed, the remaining E-series colicin clusters reveal the same pattern of divergence (not shown).

This pattern of accelerated divergence of the immunity gene and the immunity-binding region of the colicin gene suggests that some form of positive selection has played a role in the origin and diversification of colicin clusters. Further, because of the presence of the conserved sequences flanking the regions under selection, it seems reasonable to assume that this process of divergence occurs quite rapidly. Thus, a combination of (*a*) frequent recombination events generating novel combinations of colicin functional domains and colicin clusters and (*b*) their subsequent, rapid divergence through the action of positive selection may help to explain the diversity of colicins in natural populations of *E. coli.*

Acknowledgments

We thank M. Feldgarden, R. Dorit, R. Welch, W. Fitch, and an anonymous reviewer for their advice and comments on previous versions of the manuscript. This work was supported by NIH grant GM47471-01 to M.A.R.

LITERATURE CITED

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

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**:6430–6436.

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.

BRODA, P. 1979. Plasmids. W. H. Freeman, San Francisco.

CAVARD, D., R. LLOUBES, J. MORLON, M. CHARTIER, and C. LAZDUNSKI. 1985. Lysis protein encoded by plasmid ColA-CA31. Mol. Gen. Genet. **199**:95–100.

CHAK, K., and R. JAMES. 1984. Localization and characterization of a gene on the ColE3-CA38 plasmid that confers immunity to colicin E8. J. Gen. Microbiol. **130**:701–710.

———. 1986. Characterization of the ColE9-J plasmid and analysis of its genetic organization. J. Gen. Microbiol. **132**:61–71.

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

CHAN, P. T., H. OHMORI, J. TOMIZAWA, and J. LEBOWITZ. 1985. Nucleotide sequence and gene organization of colE1 DNA. J. Biol. Chem. **260**:8925–8935.

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.

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.

DAYHOFF, M. O. 1978. Surveys of new data and computer methods of analysis. Pp. 2–8. *in*

M. O. DAYHOFF, ed. Atlas of protein sequence and structure. Vol. **5**, Supp. 3. National Biomedical Research Foundation, Silver Spring, Md.

DREHER, R., V. BRAUN, and B. WITTMANN-LIEBOLD. 1985. Functional domains of colicin M. Arch. Microbiol. **140**:343–346.

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.

GILSON, L., H. K. MAHANTY, and R. KOLTER. 1990. Genetic analysis of an MDR-like export system: the secretion of colicin V. EMBO J. **9**:3875–3884.

GORDON, D. 1992. The rate of plasmid transfer among *Escherichia coli* strains. 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. Comput. Appl. Biosci. **8**:189–191.

HUGHES, A. L., and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatability 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 in the germfree mouse intestine of colicinogenic and colicin-sensitive microorganisms. Proc. Soc. Exp. Med. **130**:1280–1284.

JAMES, R., M. JARVIS, and D. F. BARKER. 1987. Nucleotide sequence of the immunity and lysis region of the ColE9-J plasmid. J. Gen Microbiol. **133**:1553–1562.

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

KOCK, J., T. OLSCHLAGER, R. M. KAMP, and V. BRAUN. 1987. Primary structure of colicin M, an inhibitor of murein biosynthesis. J. Bacteriol. **169**:3358–3361.

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

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.

LLOUBES, R. P., M. J. CHARTIER, A. M. JOURNET, S. G. VARENNE, and C. J. LAZDUNSKI. 1984. Nucleotide sequence of the gene for the immunity protein to colicin A: analysis of codon usage of immunity proteins as compared to colicins. Eur. J. Biochem. **144**:73–78.

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.

MOCK, M., and A. PUGSLEY. 1982. The BtuB group Col plasmids and homology between the colicins they encode. J. Bacteriol. **150**:1069–1076.

MORLON, J., R. LLOUBES, S. VARENNE, M. CHARTIER, and C. LAZDUNSKI. 1983. Complete nucleotide sequence of the structural gene for colicin A, a gene translated at a non-uniform rate. J. Mol. Biol. **170**:271–285.

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

OCHMAN, H., and A. C. WILSON. 1987. Evolutionary history of enteric bacteria. Pp. 1649–1654 *in* F. C. NEIDHARDT, ed. *Escherichia coli* and *Salmonella typhimurium:* cellular and molecular biology. American Association of Microbiologists, Washington, D.C.

OHNO-IWASHITA, Y., and K. IMAHORI. 1980. Assignment of the functional loci of colicin E2 and colicin E3 by the characterization of proteolytic fragments. Biochemistry **19**:652–659.

PARKER, M. W., F. PATTUS, A. D. TUCKER, and D. TSERNOGLOU. 1989. Structure of the membrane-pore-forming fragment of colicin A. Nature **337**:93–96.

PATTUS, F., D. CAVARD, R. VERGER, C. LAZDUNSKI, and C. LEVINTHAL. 1983. Pp. 407–413 *in* G. SPACH, ed. Physical chemistry of transmembrane ion motions. Elsevier, New York.

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.

———. 1987. Nucleotide sequencing of the structural gene for colicin N reveals homology between the catalytic, C-terminal domains of colicins A and N. Mol. Microbiol. **1**:317–325.

———. 1988. The immunity and lysis genes of colN plasmid pCHAP4. Mol. Gen. Genet. **211**: 335–341.

PUGSLEY, A., and M. SCHWARTZ. 1984. Colicin E2 release: lysis, leakage or secretion? possible role of phospholipase. EMBO J. **3**:2392–2397.

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.

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.

SCHRAMM, E., J. MENDE, V. BRAUN, and R. KAMP. 1987. Nucleotide sequence of the colicin B activity gene *cba:* consensus pentapeptide among TonB-dependent colicins and receptors. J. Bacteriol. **169**:3350–3357.

SCHRAMM, E., T. OLSCHLAGER, W. TROGER, and V. BRAUN. 1988. Sequence, expression and localization of the immunity protein for colicin B. J. Gen. Genet. **211**:176–182.

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.

STROBECK, C. 1983. Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics **103**:545–555.

SWOFFORD, D. L. 1989. PAUP: phylogenetic analysis using parsimony, version 3.01. Illinois Natural History Museum, Champaign.

THUM, G., T. OLSCHLAGER, and V. BRAUN. 1988. Plasmid pColBM-Cl139 does not encode a colicin lysis protein but contains sequences highly homologous to the D protein (resolvase) and the *oriV* region of the miniF plasmid. Plasmid **20**:75–82.

TOBA, M., H. MASAKI, and T. OHTA. 1986. Primary structures of the ColE2-P9 and ColE3-CA38 lysis genes. J. Biochem. **99**:591–596.

———. 1988. Colicin E8, a DNase which indicates an evolutionary relationship between colicins E2 and E3. J. Bacteriol. **170**:3237–3242.

VAN DEN ELZEN, P. J. M., H. H. B. WALTERS, E. VELTKAMP, and H. J. J. NIJKAMP. 1983. Molecular structure and function of the bacteriocin gene and bacteriocin protein of plasmid CloDFB. Nucleic Acids Res. **11**:1465–2477.

VLASUK, G. P., S. INOUYE, H. ITO, K. ITAKURA, and M. INOUYE. 1983. Effects of the complete removal of basic amino acid residues from the signal peptide on secretion of lipoprotein in *Escherichia coli.* J. Biol. Chem. **258**:7141–7148.

WATERS, V. L., and J. H. CROSA. 1991. Colicin V virulence plasmids. Microbiol. Rev. **55**:437–450.

WU, H. C., and M. TOKUNAGA. 1986. Biogenesis of lipoproteins in bacteria. Curr. Top. Microbiol. Immun. **125**:127–158.

WALTER M. FITCH, reviewing editor

Received October 5, 1992; revision received April 30, 1993

Accepted April 30, 1993