Microbiology (2000), **146**, 1671–1677 Printed in Great Britain

The newly characterized colicin Y provides evidence of positive selection in pore-former colicin diversification

M. A. Riley,[1] Luis Cadavid,[1] Marc S. Collett,[4]† Melody N. Neely,[4]‡ Mark D. Adams,[2] Cheryl M. Phillips,[2] James V. Neel[3] and David Friedman[4]

Author for correspondence: David Friedman. Tel: +1 734 763 3142. Fax: +1 734 764 3562. e-mail: davidfri@umich.edu

[1] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511, USA

[2] The Institute for Genomic Research, 9712 Medical Center Dr., Manassas, VA 20850, USA

[3,4] Dept of Human Genetics[3] and Dept of Microbiology and Immunology[4], University of Michigan, Ann Arbor, MI 48109, USA

Two evolutionary mechanisms have been proposed in the process of protein diversification of the large family of antimicrobial toxins of *Escherichia coli*, known as the colicins. Data from previous studies suggest that the relatively rare nuclease colicins appear to diversify primarily through the action of positive selection, whilst the more abundant pore-former colicins appear to diversify through the action of recombination. The complete DNA sequence of the newly characterized colicin plasmid, pCol-Let, isolated from a Yanomama Indian of South America, is presented here. This plasmid encodes a newly identified pore-former colicin, colicin Y. DNA and protein sequence comparisons of the colicin Y gene cluster and the encoded proteins with those of published pore-former colicins provide the first evidence that positive selection may also act to increase pore-former colicin diversity.

Keywords: colicins, positive selection, evolution, plasmid, Yanomama Indians

INTRODUCTION

The colicins of *Escherichia coli* are members of the large and diverse family of antimicrobial toxins known as bacteriocins. The characterized colicins, numbering over 20, can be divided into two major classes, the pore-former and nuclease colicins (James *et al*., 1991; Vuyst & Vandamme, 1994). Pore-former toxins kill by creating channels in the cytoplasmic membrane. Nuclease colicins kill by non-specific degradation of DNA or specific cleavage of rRNA. Members of both classes of colicins share a number of characteristics: (1) colicin gene clusters consist of three tightly linked genes: a colicin gene; an immunity gene, which encodes an immunity protein that provides specific protection against the encoded colicin; and a lysis gene, which encodes a protein that is involved in colicin release from the cell; (2) colicin gene clusters are encoded on plasmid replicons; and (3) colicin production is SOS mediated (Riley, 1998). Despite these shared characteristics, low levels of protein sequence similarity between pore-former and nuclease colicin proteins make it unclear whether these two classes share a common ancestor (Riley, 1993a, b).

Colicins have served as a model system for investigating the mechanisms of bacteriocin evolution and diversification (Lau *et al*., 1992; Riley & Gordon, 1992, 1995; Riley, 1993a, b, 1998; Riley *et al*., 1994; James *et al*., 1996). Much of this work has involved comparisons of DNA and protein sequences among colicins and their associated immunity and lysis genes, and encoded proteins, to infer evolutionary relationships and molecular mechanisms of diversification.

Two primary mechanisms of diversification have been implicated in colicin evolution: diversifying selection and recombination (Tan & Riley, 1997; Riley, 1998). The role of positive selection in generating colicin diversity was first proposed to explain an unusual pattern of divergence between two pairs of closely related nuclease colicin gene clusters (colicin pairs E3/E6 and E2/E9) (Riley, 1993a, b). DNA sequence comparisons reveal an apparent excess of substitutions in the immunity regions (i.e. the immunity gene and the

During the time this manuscript was being prepared for publication, our colleague James V. Neel passed away. The other authors wish to dedicate this work to his memory.

†**Present address:** Viropharma Incorporated, 405 Eagleview Blvd, Exton, PA 19341, USA.

‡**Present address:** Washington University Medical School, Molecular Biology, Campus Box 8230, St Louis, MO 62110, USA.

The GenBank accession number for the sequence reported in this paper is AF197335.

immunity binding domain of the colicin gene) of these two pairs of colicin gene clusters. To account for such an unusual clustering of substitutions, Riley (1993a, b) proposed that colicin gene clusters 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 region (for details see Riley, 1998). DNA sequence comparisons of the pore-former colicins reveal that recombination generates novel combinations of pore-former colicin functional domains and novel combinations of colicin, immunity and lysis genes (Riley, 1998). However, the pore-former gene cluster sequences are not similar enough to allow alignments in the immunity region, so that the impact of selection on this region of pore-former colicins cannot be assessed. The high frequencies with which pore-former colicins are recovered from nature argue that some sort of selection is operating but this hypothesis has not been formally tested.

We report here the complete DNA sequence and analysis of a colicin-producing plasmid (pCol-Let). Based on this analysis, we conclude that pCol-Let is related to the recently identified plasmid pColU (Smajs *et al*., 1997). The pore-former colicins encoded by these plasmids have precisely the same pattern of substitution described for nuclease colicins. This observation suggests that diversifying selection may play a similarly important role in the diversification of both nuclease and pore-former colicins.

METHODS

Plasmid isolation and characterization. Thirteen *E. coli* strains, isolated from a collection of *E. coli* strains with unique O antigens obtained from Amerindians residing in two isolated villages located in Northern Brazil near the Venezuelan border (Neel, 1970, 1994) and identified by Eveland *et al*. (1971) as non-typable, were tested for plaque formation following UV treatment. Bacteria were grown in LB broth overnight, diluted in fresh LB broth and grown to late exponential phase. Bacteria were sedimented by centrifugation, resuspended in 0·01 M $MgSO_4$ and irradiated under a UV lamp (90 s treatment with 20×10^{-7} J s^{-1} cm^{-2}). Dilutions of irradiated bacteria were plated on a lawn of *E. coli* strain K37 (Friedman *et al*., 1976). Strain K339 (originally labelled 03K-1055) consistently produced zones of lysis following UV irradiation on lawns formed from K37 and was chosen for further study. A zone of lysis was observed when K339 not treated with UV was placed on a bacterial lawn formed from K37. However, cell-free extracts derived from UV-irradiated K339 did not form plaques on a lawn formed from K37. Therefore, we concluded that lysis was not due to the induction of a prophage and most likely was due to release of a bacteriocin (colicin). Plasmid DNA isolated from K339 was used to transform K37(StrR). The isolated plasmid was named pCol-Let. Derivatives of K37 carrying pCol-Let were obtained following transformation by selecting colonies on an LB-streptomycin plate spread with the supernatant from a UV-irradiated culture of K339. The derivatives of K37 selected by this procedure, unlike their K37 parent, produced a lytic factor. However, the lytic factor did not lyse K339, indicating that K339 was immune to the lytic factor. A plasmid identical to that isolated from K339 was isolated from one of the K37 derivatives. Based on these observations, we concluded that K339 carried a colicin-producing plasmid.

Media. See Miller & Friedman (1980) for details on the media used in this study.

DNA sequence determination of pCol-Let. pCol-Let plasmid DNA, isolated using the Qiagen Midi Prep (Qiagen), was digested with *Hin*dIII and subcloned into the pUC19 cloning vector (Yanisch-Perron *et al*., 1985). DNA sequencing was initiated with the universal primer designed for pUC19 and successive rounds of primer walking were used to sequence across both strands of the insert DNA. Sequence assembly and manual inspection were done using the ABI Autoassembler.

DNA and protein sequence analysis. The entire pCol-Let DNA sequence was subjected to a DNA BLAST search employing standard methods (Altschul *et al*., 1990). Translations from all three reading frames and both orientations of the entire pCol-Let DNA sequence were subjected to a protein BLAST search using standard methods (Altschul *et al*., 1990).

Regions of significant protein or DNA sequence similarity were subjected to more detailed alignments using the CLUSTAL W algorithm of the LASERGENE program (DNASTAR). Distance and parsimony methods were employed to infer evolutionary relationships (Saitou & Nei, 1987; Swofford, 1997). In all cases, both methods yielded similar or identical tree topologies. Only parsimony-based phylogenetic inferences are reported here. The robustness of inferred tree topologies was assessed with bootstrapping (Felsenstein, 1988). Bootstrap values greater than 80 % (of 500 replicates) are reported here.

RESULTS

Colicin gene cluster

pCol-Let was isolated because it confers a colicin-like killing phenotype on the host bacterium. The complete DNA sequence of the 5857 bp pCol-Let plasmid has been determined. The sequence has been deposited in GenBank (accession no. AF197335). Protein sequence searches of GenBank (Altschul *et al*., 1990) were conducted against all putative ORFs. Significant protein sequence similarity was revealed with nine previously characterized plasmid-encoded proteins (Table 1). Three colicin-related genes (colicin, immunity and lysis) can be identified in the pCol-Let DNA sequence based

Table 1. Inferred functional regions of the pCol-Let plasmid based upon DNA and protein sequence comparisons with other characterized colicin plasmids

Function	Position (bp)
Colicin	1019–2902
Immunity	3460–2939
Lysis	3568–3717
RNA I	3778–3884
RNA II	3775–4328
Exc I	356–742
Exc II	5698–269
Rom	4736–4924
Bom	4431–4729

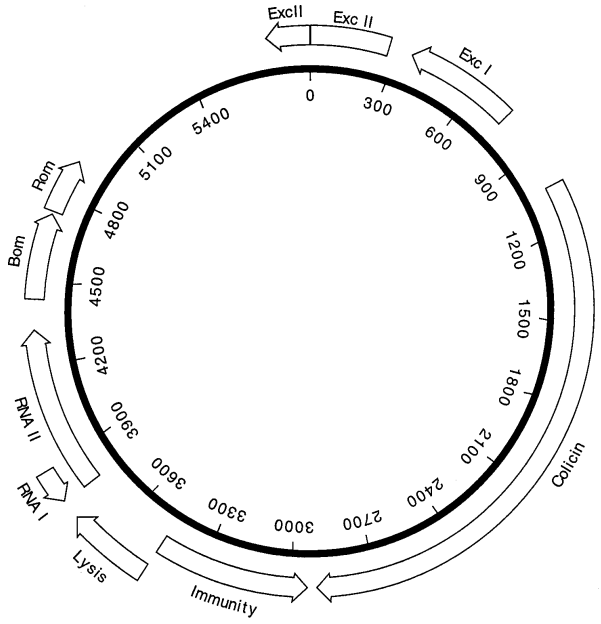

Fig. 1. pCol-Let plasmid map. Arrows indicate functional regions assigned based upon sequence similarity with characterized colicin-encoding plasmids.

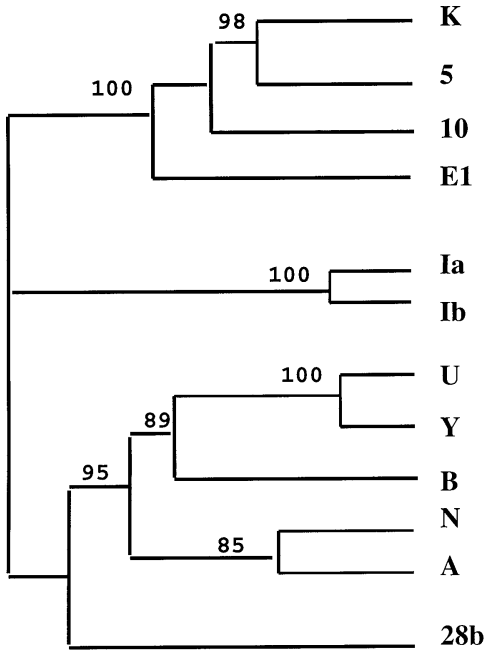

Fig. 3. Phylogenetic relationships inferred for the C-terminal region of colicin proteins using parsimony-based methods. Numbers on branches correspond to bootstrap values over 80%.

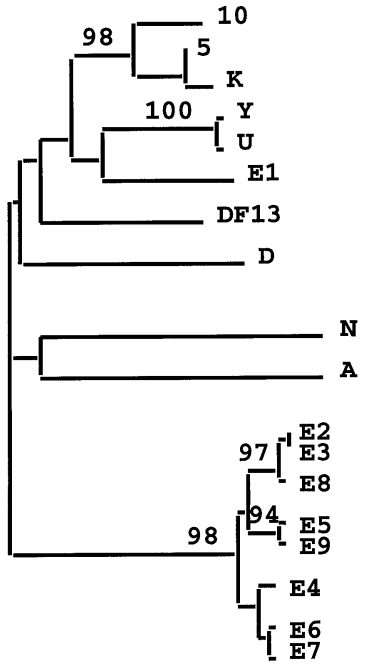

Fig. 2. Phylogenetic relationships inferred for colicin lysis genes using parsimony-based methods. Numbers on branches correspond to bootstrap values over 80%.

upon high levels of DNA sequence similarity with previously characterized colicin gene clusters (Table 1, Fig. 1). Of the three colicin-related genes, the lysis gene exhibits the highest degree of sequence conservation relative to published colicin gene clusters. The pCol-Let lysis gene is comprised of 153 bp and is predicted to encode a protein of 50 amino acids, based upon comparisons to published lysis genes and encoded proteins. Levels of pCol-Let lysis protein sequence similarity range from 87% (colicin U lysis) to 41% (colicin A lysis).

A phylogenetic tree was inferred for all characterized lysis genes based upon parsimony methods (Swofford, 1997) (Fig. 2). With the exception of DF13, there is a division in the gene tree between lysis genes associated with pore-former colicins (E1, 5, 10, K, U, D, N and A) and nuclease colicins (E2–E9). DF13 is a nuclease colicin; however, its associated lysis gene clusters with the pore-former lysis genes. pCol-Let falls within the pore-former cluster of branches of the lysis tree. Based upon sequence similarity and branch lengths, the closest known relative to pCol-Let lysis is the lysis protein of colicin U (see Fig. 2).

The pCol-Let immunity gene is predicted to be 525 bp and encodes a protein of 174 amino acids, based upon sequence similarity with other characterized immunity genes. The immunity gene is apparently transcribed in the opposite direction to that of the lysis and colicin genes (refer to Fig. 1). This reverse orientation of the immunity gene relative to the remainder of the colicin gene cluster is a common characteristic of pore-former immunity genes (Pugsley & Oudega, 1987). In contrast

Table 2. Estimates of synonymous (K_s) and non-synonymous (K_n) substitutions between pore-former colicin gene clusters

	K_s	K_n
pCol-Let vs colicin U		
Colicin	0·390	0·072
5′ half	0·269	0·054
3′ half	0·524	0·088
Immunity	0·426	0·161
Lysis	0·324	0·092
Colicin 5 vs colicin 10		
Colicin	0·108	0·035
5′ half	0·000	0·000
3′ half	0·229	0·068
Immunity	0·501	0·204
Lysis	0·289	0·057

to the lysis protein, the pCol-Let immunity protein does not show extensive protein sequence similarity with other characterized immunity proteins. The highest levels of sequence similarity are detected between pCol-Let immunity protein and the immunity proteins associated with colicins U, A and B. Levels of pair-wise protein sequence similarity within this group of immunity proteins range from 69% (between colicin U and pCol-Let) to 31% (between colicin B and pCol-Let). As was the case for the lysis protein, the closest relative to pCol-Let immunity protein is the immunity protein of colicin U.

The pCol-Let colicin gene is predicted to be 1890 bp, which encodes a protein of 629 amino acids, based upon sequence similarity with other characterized colicin genes. The most similar colicin proteins are colicins U, A and B. Levels of protein sequence similarity within this group of colicin proteins range from 87% (between pCol-Let and colicin U) to 41% (between pCol-Let and colicin A). The reduced levels of protein sequence similarity among most characterized pore-former proteins preclude inference of a pore-former colicin protein phylogeny. However, if the protein sequence comparison is restricted to just the C-terminal domain of the protein, levels of protein sequence similarity exceed 20%. Fig. 3 provides the phylogeny inferred from the C-terminal domain of pore-former colicin proteins. Again, the closest relative to the pCol-Let colicin protein is colicin U. In this analysis, the colicin gene was divided into two blocks of sequence of equal length. This was done because previous investigations of nuclease colicin substitution patterns suggest that the killing and immunity domains of the colicin proteins (encoded in the 3′ end of the gene) and the immunity protein accumulate substitutions more rapidly than the 5′ end of the colicin protein (Riley, 1998).

Based upon colicin, immunity and lysis comparisons, the closest known relative to the pCol-Let colicin gene cluster is colicin U. These colicins are similar enough to allow unambiguous alignments of each of the three colicin-related genes. Levels of synonymous and non-synonymous substitutions between pCol-Let and colicin U colicin, immunity and lysis genes are given in Table 2. There is a significant difference in the levels of synonymous substitutions that have accumulated across the pCol-Let and colicin U colicin-related genes ($G = 23{\cdot}476$, d.f. $= 3$, $P << 0{\cdot}005$). The immunity gene has an elevated level of synonymous substitution, with lower levels of substitution observed in the colicin and lysis genes. When the colicin gene is divided into 5′ and 3′ halves, the 3′ half of the colicin gene has an elevated level of synonymous substitutions. There is also a significant difference in the levels of non-synonymous substitutions that have accumulated across the pCol-Let and colicin U colicin-related genes ($G = 73{\cdot}52$, d.f. $= 3$, $P << 0{\cdot}005$). The immunity gene has an elevated level of non-synonymous substitution, with lower levels of substitution detected in the 5′ half of the colicin gene and in the lysis genes.

pCol-Let plasmid organization

Several colicin (Col) plasmids have been sequenced completely, including ColE1 and ColA plasmids (Chan *et al.*, 1985; Morlon *et al.*, 1988). Levels of plasmid sequence similarity were assessed between pCol-Let, ColE1 and ColA (Table 3). All three Col plasmids show

Table 3. Percentage levels of DNA sequence similarity between colicin plasmids pCol-Let, ColE1 and ColA in plasmid maintenance and mobilization-related functions versus colicin-related functions

NA, Not applicable, as ColA does not have a detectable Rom region.

Comparison	Plasmid maintenance genes				Colicin-related genes		
	bom	*rom*	*RNAII*	*exc2*	*col*	*imm*	*lys*
pCol-Let vs ColA	58·0	NA	85·1	67·4	48·6	48·0	48·5
pCol-Let vs ColE1	73·9	59·8	93·8	69·4	28·1	20·1	41·4
ColA vs ColE1	60·8	NA	80·0	54·7	16·9	19·0	38·3

high levels of conservation in the RNA I and II regions, involved in plasmid replication, the Bom and Rom regions, involved respectively in plasmid mobility and replication, and the Exc region, involved in plasmid entry exclusion. Much lower levels of protein sequence similarity are observed in the colicin, immunity and lysis proteins (Table 3).

The primary differences in size among these three related Col plasmids involve length differences in the Mob region, which encodes the information required for plasmid mobilization. The Mob region of the ColE1 plasmid constitutes one-third of the plasmid length and includes four essential genes (*mobC*, *A*, *B* and *D*) required for mobilization. The ColA plasmid has the same four essential *mob* genes; however, there are numerous insertions and deletions in two of the *mob* genes (*mobA* and *mobB*) and the Rom region is absent (Morlon *et al*., 1988). Low levels of sequence similarity are detected between ColE1 and pCol-Let for one *mob* gene (*mobC*, data not shown). The remainder of the corresponding Mob region in pCol-Let is unalignable with either ColE1 or ColA. This portion of the pCol-Let plasmid is quite short, about 500 bp.

DISCUSSION

pCol-Let is a low molecular mass plasmid (5847 bp) isolated from the *E. coli* of a Yanomama Indian of South America. Examination of the complete DNA sequence of pCol-Let suggests that the plasmid encodes a number of functions required for plasmid DNA replication and stable plasmid maintenance (Fig. 1). These functions include DNA sequences involved in the regulation of plasmid DNA replication (the RNA I and II regions located near the plasmid origin of replication), the plasmid incompatibility function mediated by the entry exclusion proteins Exc 1 and 2, and a Rom region involved in plasmid replication.

In addition to these plasmid maintenance functions, pCol-Let also encodes a colicin 'killing' phenotype (Riley, 1993a). Comparisons of inferred protein sequences of the pCol-Let encoded colicin, immunity and lysis proteins with previously characterized colicin-related proteins reveal that pCol-Let encodes a newly identified member of the pore-former colicin family. Following bacteriocin nomenclature protocols, we have named this colicin gene cluster colicin Y. The recently characterized colicin U gene cluster represents the closest known relative of the colicin Y gene cluster (Smajs *et al*., 1997). This close evolutionary relationship is revealed most clearly with a comparison of the lysis and colicin proteins, which retain high levels of protein and DNA sequence similarity (Figs 2 and 3). The immunity protein encoded in the colicin Y gene cluster, although highly divergent from other characterized immunity proteins, retains unambiguous protein sequence similarities with the colicin U immunity protein (Table 2).

Extensive DNA and protein sequence comparisons have revealed that the two families of colicin gene clusters (pore-former and nuclease) experience both kinds of mutational events. But the evolutionary significance varies (Riley, 1998). It has been suggested that pore-former diversity results from frequent recombination between existing pore-former types, generating additional killing phenotypes. With few exceptions, such as colicins Ia and Ib (Riley, 1993a), pore-former proteins are highly divergent in protein sequence. However, detailed protein sequence comparisons often reveal relatively short stretches of highly similar protein sequence interspersed within longer stretches of highly divergent sequence (Tan & Riley, 1997). These restricted regions of high sequence similarity have been interpreted as resulting from recombination events.

By contrast, nuclease colicins are similar in protein sequence. With one exception (colicin E2; Tan & Riley, 1997), amino acid sequence comparisons among nuclease colicins have revealed no evidence of the process of recombination-mediated diversification proposed for the pore-former colicins. When pairs of closely related nuclease colicin gene clusters are aligned, there is an elevated level of sequence divergence clustered in the C-terminal region of the colicin protein and in the immunity protein (Tan & Riley, 1997). However, this clustered divergence has been interpreted as resulting from the action of positive selection for novel immunity phenotypes, rather than resulting from recombination (Tan & Riley, 1997).

The process of selection-mediated nuclease colicin diversification is envisioned to require two steps. First, one or a few mutations in the immunity gene results in changes in the immunity function that broadens the immunity of the host strain against several additional colicins. In environments with multiple colicins segregating, this function will be strongly selected for and thus will be retained in the population long enough for a second, paired mutation to produce a 'super-killer' phenotype. The super-killer phenotype results in a strain of *E. coli* that is immune to its own colicin and to its ancestor's colicin. However, the ancestral strain will be immune to self, but not immune to the newly evolved colicin phenotype. Thus, the 'super-killer' strain will rapidly invade the ancestral population. Repeated rounds of this sort of diversifying selection will result in the accumulation of differences in the immunity proteins and in the immunity binding regions of the colicin proteins (reviewed by Riley, 1998). The result of diversifying selection is a large and homogeneous class of nuclease colicins (save for the rapidly evolving immunity regions).

The high frequency with which pore-former colicins are encountered in nature argues that, in addition to recombination generating new pore-former colicin phenotypes, positive selection acts to increase the frequency of these new phenotypes (Riley, 1998). However, due to the high levels of sequence divergence between the pore-former proteins characterized to date, the importance of selection and the molecular details of this process in the generation of pore-former diversity have been imposs-

ible to assess. The colicin Y gene cluster encoded in the pCol-Let plasmid provides the first example in which closely related pairs of pore-former colicin gene clusters have been examined for the signature of diversifying selection. As indicated in Table 2, colicin Y and its closest known relative, colicin U, have an elevated level of substitution in the immunity gene and in the 3′ half of the colicin gene, relative to the remainder of the gene cluster. This pattern of substitution corresponds to the pattern of divergence previously revealed only for nuclease colicins (Riley, 1998). Recently, a second pair of closely related pore-former gene clusters has been characterized, colicins 5 and 10 (Pilsl & Braun, 1995). Table 2 provides a summary of the level of substitution between these two closely related pore-former colicin gene clusters. There is an elevated level of substitution in the immunity genes and in the 3′ region of the colicin genes. In both these pore-former comparisons, there is a significantly elevated level of DNA sequence divergence in the immunity region, a pattern of substitution predicted by the diversifying selection hypothesis.

The correlation between higher substitutions in the immunity gene and the 3′ half of the colicin gene found for nuclease colicins appears easily explained, since these proteins are known to interact (Riley, 1998). However, the similar correlation observed here between higher substitutions in the immunity region gene and the 3′ half of the colicin gene for pore-former colicins is not as easily explained, since it is not clear yet how these proteins interact. However, it is difficult to explain immunity specificity without assuming some type of interaction between the immunity and colicin protein.

The pCol-Let plasmid was originally characterized with the expectation that understanding the evolutionary relationships of the encoded colicin-associated proteins might provide additional insight into the degree of isolation of the microbial flora of Yanomama Indian. pCol-Let encodes a colicin phenotype that has not previously been described, as might be expected for an isolated *E. coli* population. However, the colicin Y gene cluster is quite closely related to the colicin U gene cluster. Colicin U was isolated in 1989 from a strain of *Shigella boydii* isolated in Prague (Horak, cited in Smajs *et al*., 1997). Thus, both significant geographical space and time separate these two colicin gene cluster isolations. Previous DNA sequence studies of colicin E2 gene clusters have revealed that a ColE2 plasmid isolated from Australia was more closely related to a ColE2 plasmid isolated over 15 years earlier from France, than to 12 other ColE2 plasmids isolated at the same time from Australia. These observations argue that *E. coli* migration rates (or at least plasmid transfer and subsequent migration rates) are substantial and may not reflect the same patterns of isolation as do the host mammals from which they are isolated.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Science Foundation and National Institutes of Health to M. A. R. and the National Institutes of Health to D. F. M. N. N. was supported in part by NIH training grant GM08353 and M. S. C. was supported in part by NIH training grant GM07315.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Chan, P. T., Ohmori, H., Tomizawa, J. & Lebowitz, J. (1985). Nucleotide sequence and gene organization of ColE1 DNA. *J Biol Chem* **260**, 8925–8935.

Eveland, W. C., Oliver, W. J. & Neel, J. V. (1971). Characteristics of *Escherichia coli* serotypes in the Yanomama, a primitive Indian tribe of South America. *Infect Immun* **4**, 753–756.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* **22**, 521–565.

Friedman, D. I., Baumann, M. & Baron, L. S. (1976). Cooperative effects of bacterial mutations affecting λ N gene expression. *Virology* **73**, 119–127.

James, R., Lazdunski, C. & Pattus, F. (1991). Bacteriocins, microcins and lantibiotics. NATO ASI series, vol. 65. New York: Springer.

James, R., Kleanthous, C. & Moore, G. R. (1996). The biology of E colicins: paradigms and paradoxes. *Microbiology* **142**, 1569–1580.

Lau, P. C. K., Parsons, M. & Uchimura, T. (1992). Molecular evolution of E colicin plasmids with emphasis on the endonuclease types. In *Bacteriocins*, *Microcins and Antibiotics*. Edited by R. James, C. Lazdunski & F. Pattus. Berlin: Springer.

Miller, H. I. & Friedman, D. I. (1980). An *E. coli* gene product required for λ site-specific recombination. *Cell* **20**, 711–719.

Morlon, M., Sherratt, D. & Lazdunski, C. (1988). Identification of functional regions of the colicinogenic plasmid ColA. *Mol Gen Genet* **211**, 223–230.

Neel, J. V. (1970). Lessons from a ''primitive'' people. *Science* **170**, 815–822.

Neel, J. V. (1994). *Physician to the Gene Pool*. New York: Wiley.

Pilsl, H. & Braun, V. (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. & Oudega, B. (1987). Methods for studying colicins and their plasmids. In *Plasmids: a Practical Approach*, pp. 105–161. Edited by K. G. Hardy. Oxford: IRL Press.

Riley, M. A. (1993a). Molecular mechanisms of colicin evolution. *Mol Biol Evol* **10**, 1380–1395.

Riley, M. A. (1993b). Positive selection for colicin diversity in bacteria. *Mol Biol Evol* **10**, 1048–1059.

Riley, M. A. (1998). Molecular mechanisms of bacteriocin evolution. *Annu Rev Genet* **32**, 255–278.

Riley, M. A. & Gordon, D. (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. & Gordon, D. (1995). Ecology and evolution of bacteriocins. *J Ind Microbiol* **17**, 155–158.

Riley, M. A., Tan, Y. & Wang, J. P. (1994). Nucleotide polymorphism in colicin E1 and Ia plasmids from natural isolates of *Escherichia coli*. *Proc Natl Acad Sci USA* **91**, 11276–11280.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Smajs, D., Pilsl, H. & Braun, V. (1997). Colicin U, a novel colicin produced by *Shigella boydii*. *J Bacteriol* **179**, 4919–4928.

Swofford, D. (1997). *Phylogenetic Analysis using Parsimony* (*PAUP* 4.0**). Sunderland, MA: Sinauer.

Tan, Y. & Riley, M. A. (1997). Positive selection and recombination: major molecular mechanisms in colicin diversification. *Trends Ecol Evol* **12**, 348–351.

Vuyst, L. D. & Vandamme, E. (1994). Lactic acid bacteria and bacteriocins: their practical importance. In *Bacteriocins of Lactic Acid Bacteria*, pp. 1–11. Edited by L. De Vuyst. London: Blackie Academic and Professional.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

Received 9 November 1999; revised 6 April 2000; accepted 17 April 2000.