Sindh Univ. Res. Jour. (Sci. Ser.) Vol.45 (4) 651-656 (2013)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

In silico Nucleotide Sequence Analysis of the Colicin E7-K317 plasmid: A Bioinformatics Approach

S. BANO[++], S. A. TUNIO, A. KEERIO*, A. B. MEMON**, N. AHMED, S. MANZOOR

Department of Microbiology, University of Sindh, Jamshoro, Pakistan

Abstract: Colicins are protein antibiotics that are synthesized by *E. coli* cells and act against related species. E colicins are plasmid-encoded antibacterial proteins produced by *E. coli* which employ the BtuB receptor for its reception into sensitive target cell. The objective of the current study was to undertake the *in silico* analysis of complete nucleotide sequences of pColE7-K317 plasmid and to explore the relationship of predicted open reading frames with other known colicin plasmids using bioinformatics tools. The nucleotide sequences of the pColE-7-K317 were analyzed using Gene Mark software, Vector NTI, MEGA6, and online gene analysis tools including BLAST. The data of bioinformatics analysis of the complete nucleotide sequences of pColE7-K317, comprising of 6077 bps, demonstrated seven open reading frames (ORFs) on the plasmid. Four ORFs, including three coding for ColE7 operon and one ORF for replication region have previously been characterized. In the present study, we report three new previously unrecognized putative OFRs. Of these three, one ORF located on positive strand displayed close homology with *mob*A gene of *E .coli* whereas two ORFs, one located on positive and other on negative strand, were identified to correspond with hypothetical proteins.

Keywords: Colicin E7; *Escherichia coli*-K317; GeneMark analysis, BLAST.

1. INTRODUCTION

Bacteriocins are protein antibiotics secreted by most species of bacteria and play an important role in mediating microbial interaction and maintaining microbial diversity. Colicins are plasmid-encoding bacteriocins that are produced by strains of *E. coli* and have narrow killing spectrum against *E. coli* and closely related species (Kirkup & Riley, 2004). They have been classified into groups on the basis of cell surface receptor on target cells to which they bind in the first stage of cell killing. E-group colicins use the vitamin B_{12} receptor protein, encoded by the chromosomal gene *btuB*, as their surface receptor (Di Masi *et al.*, 1973), and are further distinguished into E1 to E9 colicin types corresponding to their specific immunity groups (Cooper & James, 1984; Watson *et al.*, 1985). E. group colicins are exclusively encoded by plasmids. Several cytotoxic activities have been recognized in E colicins. Colicins E3, E5 and E6 are RNases; E3 and E6 cause cleavage of rRNA (Bowman *et al.*, 1971; Senior & Holland, 1971), while E5 specifically target tRNA (Masaki & Ogawa, 2002; Ogawa *et al.*, 1999). In contrast, colicins E2, E7, E8 and E9 of the DNase group degrade DNA non-specifically (Akutsu *et al.*, 1989; Curtis *et al.*, 1989; Schaller & Nomura, 1976; Toba *et al.*, 1988). An E colicin operon typically consists of three genes. Of these three, the first gene is always the colicin structural gene which encodes the colicin protein. The second gene of the E colicin operon is usually a specific immunity gene, that encodes the immunity protein, which provides resistance to the colicin-producing strains against the lethal effects of the colicin produced (Chak & James, 1986; Kuhar, 1999). The third and last gene in the operon is the lysis gene , that encodes the lysis protein (also called the Bacteriocin Release Protein), and assists in secretion of the colicin from the producing cell (van der Wal *et al.*, 1995).

The plasmid pColE7-K317 was first isolated from a strain of *E. coli* K12; CL317 (Males & Stocker, 1980). The characterization of this plasmid was previously made by cloning the colicin operon clusters in recombinant plasmids (Chak *et al.*, 1991). However, bioinformatics analysis of its complete sequence remained to be carried out. In the present study, we describe *in silico* analysis, construct the distance tree of the plasmid and produce the genetic map of pColE7-J plasmid through BLAST search.

2. MATERIALS AND METHODS

Bioinformatics analysis of the complete nucleotide sequence of pColE7-K317 based on homologous DNA sequences available at EMBL and GenBank (http://www.ncbi.nlm.nih.gov/genbank/) database libraries were performed using BLAST search available at http://blast.ncbi.nlm.nih.gov/. The distance tree of the complete nucleotide sequence (6077 bp) of pColE7-K317 was constructed by BLAST pairwise alignment available at www.ncbi.nlm.nih.gov/blast/treeview/treeView.cgi.
Tree was prepared using fast minimum evolution with 0.75 maximum sequence differences. The prediction of putative ORFs was carried out by using Gene Mark software (Besemer & Borodovsky, 1999) available at http://exon.gatech.edu/heuristic_hmm2.cgi. The similarity patterns of predicted protein encoded by pColE7-K317 was done using protein BLAST search available at

[++]Corresponding Author: S. BANO
*Institute of Mathematics and Computer Science, University of Sindh Jamshoro
**Department of Computer Science, Sindh University Laar campus @ Badin

http://blast.ncbi.nlm.nih.gov/. The physical genetic map of the pColE7-K317 plasmid was constructed using the Vector NTI program (Invitrogen).

3. RESULTS

3.1 *In silico* analysis of sequence of gene of ColE7-K317 plasmid

The complete sequence of pColE7-K317 was analyzed using BLAST nucleotide search and distance tree of whole pColE7-K317 was generated which is shown in Figure 1.The data of Gene Mark software demonstrated that six ORFs were on positive strand whereas one ORF was identified on the negative strand (Table 1). Three of the ORFs located on positive strand have previously been reported to code for colicin operon which consists of *ceaG*, *ceiG*, *celG* genes that encode colicin activity, Immunity and lysis proteins for ColE7, respectively. Moreover, one ORF located on positive strand has been suggested to encode Rep protein (Hiraga *et al*., 1994). Among the remaining three ORFs, one exhibited similarity with mobA gene, whilst, other two genes were found to correspond with hypothetical proteins. One gene coding for putative hypothetical proteins was present on negative strand. The single-strand initiation sequence (ssi) of pColE7-K317 plasmid (gtagcgctcgccgcagtctcatgaccgagcgtagcgagcgaatgagcgag gaagcgcaaaggcgtccggtg) exhibited 100% homology with ssiA sequence of pColE9-J (Bano *et al.*, 2012), and ColE2 plasmids (Nomura *et al.*, 1991).

Complete information regarding all identified ORFs including total gene length, and location on plasmid strand is given in **table 1.**

Table 1. The pColE7-K317 plasmid Gene Prediction data obtained by using GeneMark.hmm 2.0 .

Gene Order	DNA Strand	Left end	Right end	Gene Length (bps)	Class
1	+	222	1952	1731	1
2	+	1954	2217	264	1
3	+	2246	2422	177	1
4	+	2922	3836	915	1
5	-	4074	4439	366	1
6	+	4616	5542	927	1
7	+	5545	>6075	531	1

3.2 Amino Acid sequences of Predicted proteins

>gene_1|GeneMark.hmm|576_aa|+|222|1952
>pColE7-K317
MSGGDGRGHNSGAHNTGGNINGGPTGLGGNGG
ASDGSGWSSENNPWGGGSGSGVHWGGGSGHGN
GGGNSNSGGGSNSSVAAPMAFGFPALAAPGAGT
LGISVSGEALSAAIADIFAALKGPFKFSAWGIALY
GILPSEIAKDDPNMMSKIVTSLPAETVTNVQVSTL
PLDQATVSVTKRVTDVVKDTRQHIAVVAGVPMS
VPVVNAKPTRTPGVFHASFPGVPSLTVSTVKGLP
VSTTLPRGITEDKGRTAVPAGFTFGGGSHEAVIRF
PKESGQKPVYVSVTDVLTPAQVKQRQDEEKRLQ
QEWNDAHPVEVAERNYEQARAELNQANKDVAR
NQERQAKAVQVYNSRKSELDAANKTLADAKAEI
KQFERFAREPMAAGHRMWQMAGLKAQRAQTDV
NNKKAAFDAAAKEKSDADVALSSALERRKQKEN
KEKDAKAKLDKESKRNKPGKATGKGKPVNNKW
LNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKK
FWEEVSKDPELSKQFSRNNNDRMKVGKAPKTRT
QDVSGKRTSFELHHEKPISQNGGVYDMDNISVVT
PKRHIDIHRGK

>gene_2|GeneMark.hmm|87_aa|+|1954|2217
>pColE7-K317
MELKNSISDYTEAEFVQLLKEIEKENVAATDDVL
DVLLEHFVKITEHPDGTDLIYYPSDNRDDSPEGIV
KEIKEWRAANGKPGFKQG

>gene_3|GeneMark.hmm|58_aa|+|2246|2422
>pColE7-K317
MTGILSQQGVVMKKITGIILLLLAAIILAACQANYI
RDVQGGTVSPSSTAELTGVETQ

>gene_4|GeneMark.hmm|304_aa|+|2922|3836
>pColE7-K317
MSAALQYFDENLPHRPYHTDDLAFGLRISGKGRA
LLARYIQQNQPHAQFWLVFDVDREGAAIDWSDR
NAPAPNITVKNPVNGHAHLLYALNIAVRTAPDSS
VKALKYAAAIERSLCEKLCADVNYSGLICKNPFH
LEWQVMEWREEAYTLDELADYLDLSASARRSID
KHYGMGRNCHLFEMTRKWAYRAMRQGWPEFSQ
WLDAVIQRVEMYNASLPVPLSPPECRAIGRSIAKY
THRNFTPETFAQYVADTHTPEIQAARGRKGGSKS
KRSTVATSARTLKPWEALGISRAWYYQLKKRGL
VE

>gene_5|GeneMark.hmm|121_aa|-|4074|4439
>pColE7-K317
MRISNIEWLKKRIGFIRKLGEQTARQRQIIDLLDNE
AGLTEQERKLPPRTATAEKNDLQAQESERKQAVQ
KRIEGKKQRRERNHRLFLAAGLLIEAGLVDTKTG
ELCYKKDRILQALKELK

>gene_6|GeneMark.hmm|308_aa|+|4616|5542
>pColE7-K317
VATYHLSVKFGGKGQAANHADYIERKGEYAKRK
DLEYTEHGNMPEWARDNPSHFWQAADQFERAN
GSTYRELEIALPRELTPEQRLELVQDFVRQEAGER
HAWSFAIHNPKASIDGGEQPHAHIMMSQRVNDGI
DRTPEQYFRRYNARYPERGGAKKDSGSLTLTQQK
EQLRELRKRWEVKHNEHMRKHGFERGFIDCRTL
KEQGIERRPEVHLGFEAAGRLDDAQRHDIMQKRS
EPDYSRHLEQARTVGHDIREWLDTPVKQHELVRR
LQHDVAREHGFDMRDRVSRQDRTEPERVKSPSR
GFSMGF

>gene_7|GeneMark.hmm|177_aa|+|5545|6075
>pColE7-K317

MASTEEKLLTLLLSVEHLQHTAMEQQKAVADSL
AVARETLRGYRDISDGVAKQVRDEVRNEVQRLD
VAGLIGERISQDFQRLEESVANMKANADALNRDI
ATIHSNIVQEYNTLKGYSWKWLGGVSVAFLVMF
SVLGWGMKTILDRNYNATAAVYQKVVTLEDKL
AEKPVTGKSRKN

3.3 Bioinformatics analysis of all ORFs of plasmid pColE7-K317

To determine the homologous sequences and similarity of pColE7-K317 sequences, we undertook BLAST analysis of all predicted proteins present on pColE7-K317 plasmid. Since the results of Gene Mark analysis were in the form of protein sequences, we used protein BLAST analysis, of all ORFs which were obtained from Gene Mark analysis of complete sequence pColE7-K317. In Figure 2 (A), the BLAST data of ORF for predicted protein 1 of pColE7-K317, showed a putative conserved domain corresponding to cloacin super family indicating the protein encoded by the gene is for ColE7 structural protein which was validated by distance tree generated by BLAST. Similarly, the BLAST search result of amino acid sequence of ORFs 2, 3 and 4 of pColE7-K317 demonstrated a putative conserved domain for ORF 2 which is colicin-pyocin superfamily, therefore it belonged to colicin operon but not with structural protein for ColE7 (Fig. 2B). Further analysis of the BLAST search result together with the distance tree demonstrated that the protein was related to colicin immunity protein. However, the putative conserved domain for ORF 3 protein is lysis-col super family; therefore this protein also belonged to colicin operon for ColE7 (Fig 2C). In Figure 2D, the BLAST search data of the gene 4 of pColE7-K317 showed a putative conserved domain corresponding to Replicase super family, suggesting that this ORF encodes Rep protein that is important for replication initiation of the plasmid. In Figure 2E, the BLAST search result of gene 6 of pColE7-K317 (Fig. 2E) showed that there is a putative conserved domain for this protein which is MobA_MobL super family. Further analysis of the BLAST search result together with the distance tree generated by BLAST demonstrated that the predicted protein sequence was for Mob protein The BLAST search result of amino acid sequence of ORFs 5 and 7, of pColE7-K317 showed no putative conserved domain for these proteins. However, further analysis of the BLAST search results demonstrated that the protein sequence corresponding to ORFs 5 and 7 belonged to hypothetical proteins.

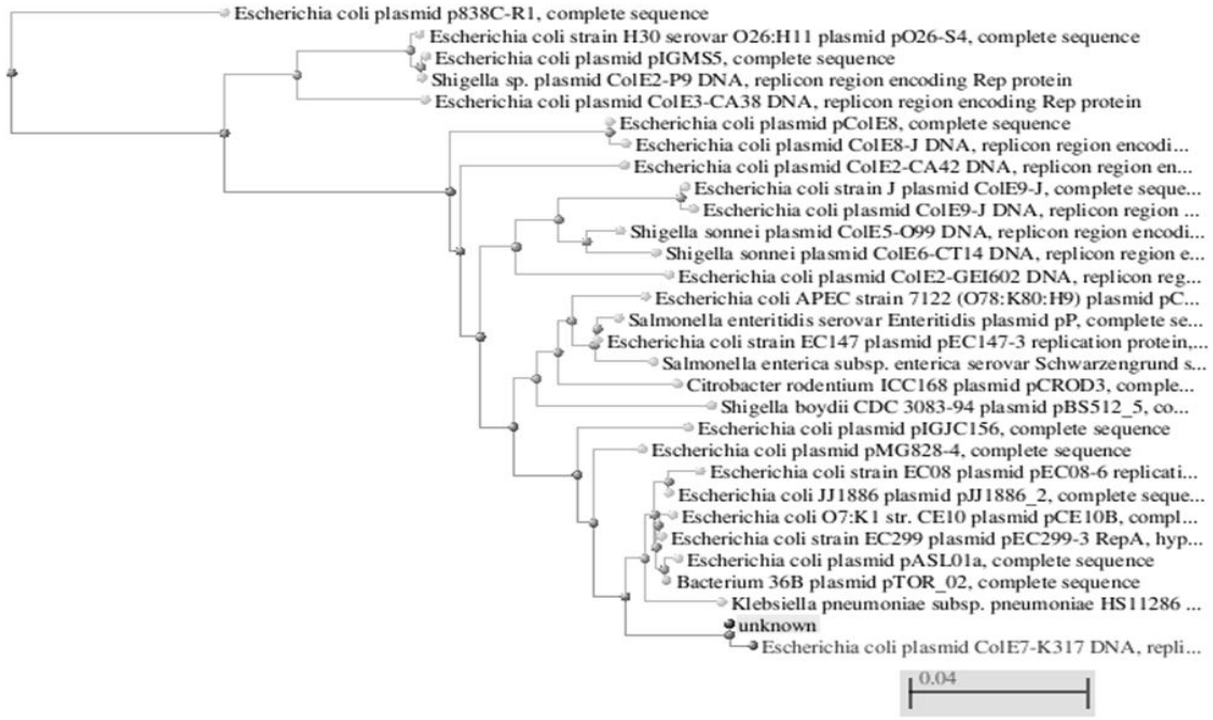


Fig. 1. The evolutionary distance tree of pColE7-K317 generated using BLAST pairwise alignment

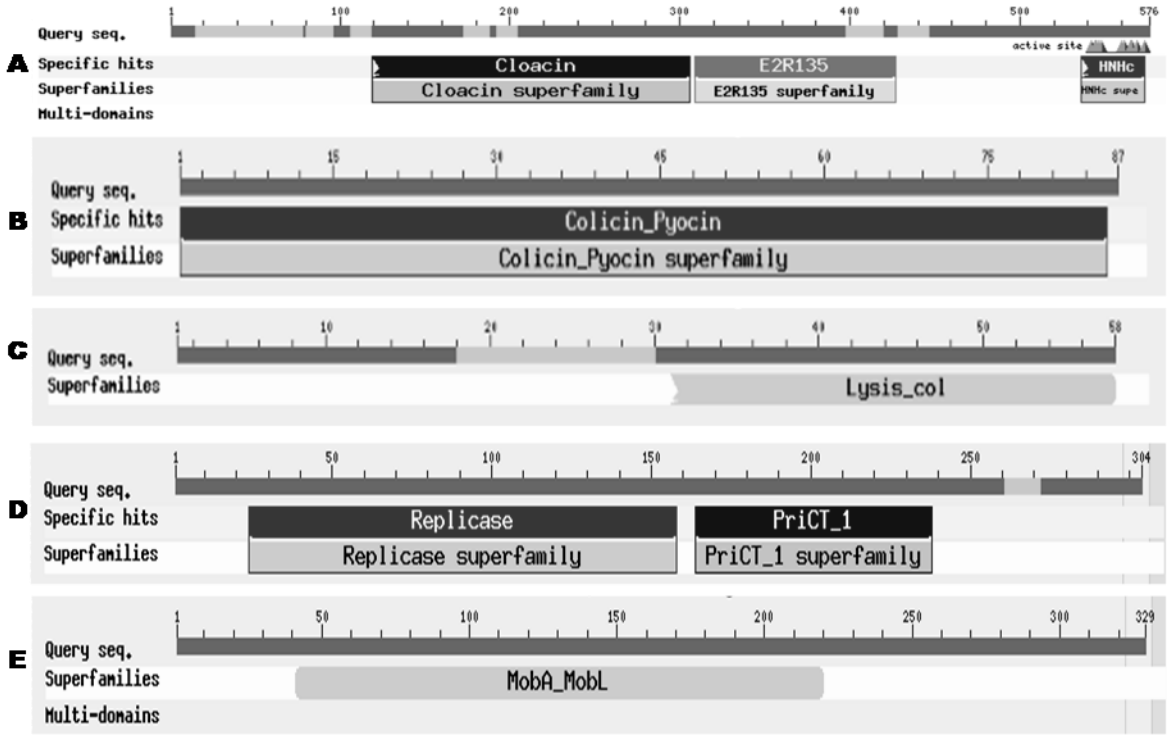


Fig. 2. Graphical presentation of BLAST analysis of Putative conserved domains of predicted ORFs of plasmid ColE7 -. a) ORF 1, b) ORF-2, c) ORF3, d) ORF4, e) ORF-6

Physical map of complete nucleotide sequences of pColE7-K317 was generated based on the bioinformatics analysis of the sequences using the results obtained from gene predictions by GeneMark and the BLAST search for putative conserved domains of all ORFs present on pColE7-K317 (Fig 3). The position and orientation of functional genes on the basis of sequence comparisons are shown in the form of arrows. The promoter of the colicin operon is indicated as *ColE7p* whereas the genes encoding colicin activity, immunity and lysis proteins of ColE7 are shown as *ceaG*, *ceiG*, *ceGI*, respectively. Various sites for restriction endonucleases enzymes are also shown in the map.

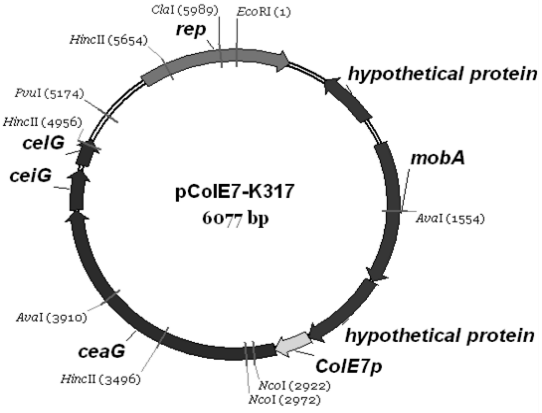


Fig 3. Physical and genetic map of plasmid pColE7-K317

4.0 Conclusion

The present study has, for the first time, described bioinformatics analysis of complete nucleotide sequences of the pColE7-K317 plasmid. The data demonstrated that total seven ORFs were present on the pColE7-K317 plasmid. Three ORFs belonged to ColE7 structural protein, ColE7 immunity protein, ColE7 lysis protein whereas the fourth ORF represented Rep protein. On the basis of homology analysis, one gene was identified to encode putative MobA protein and two ORFs were presumed as hypothetical proteins. The data of current study will facilitate the future genetic studies and further characterization of the plasmid.

REFERENCES:

Akutsu, A., H. Masaki & 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.

Bano, S., M. Vankemmelbeke, S. A. Tunio, C. Penfold & R. James (2012). Complete nucleotide sequencing and molecular characterization of the pColE9-J plasmid. *Sindh Uni Res Jour (Sci Ser)* (**44)**, 463-486.

Besemer, J. & M. Borodovsky (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Res* (**27)**, 3911-3920.

Bowman, C. M., J. E. Dahlberg, T. Ikemura, J. Konisky & M. Nomura (1971). Specific inactivation of 16S ribosomal RNA induced by colicin E3 in vivo. *Proc Natl Acad Sci U S A* (**68)**, 964-968.

Chak, K. F. & R. James (1986). Characterization of the ColE9-J plasmid and analysis of its genetic organization. *J Gen Microbiol* (**132)**, 61-70.

Chak, K. F., W. S. Kuo, F. M. Lu & R. James (1991). Cloning and characterization of the ColE7 plasmid. *J Gen Microbiol* (**137)**, 91-100.

Cooper, P. C. & R. James (1984). Two new E colicins, E8 and E9, produced by a strain of Escherichia coli. *J Gen Microbiol* (**130)**, 209-215.

Curtis, M. D., R. James & 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.

Di Masi, D. R., J. C. White, C. A. Schnaitman & C. Bradbeer (1973). Transport of vitamin B12 in Escherichia coli: common receptor sites for vitamin B12 and the E colicins on the outer membrane of the cell envelope. *J Bacteriol* (**115)**, 506-513.

Hiraga, S., T. Sugiyama & T. Itoh (1994). Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J Bacteriol* (**176)**, 7233-7243.

Kirkup, B. C. and M. A. Riley (2004). Antibiotic-mediated antagonism leads to a bacterial game of rock-paper-scissors in vivo. *Nature* (**428)**, 412-414.

Kuhar, I. (1999). Transcription Regulation of the Colicin K cka Gene Reveals Induction of Colicin Synthesis by Differential Responses to Environmental Signals. *J Bacteriol* (**181)**, 7373-7380.

Males, B. M. & B. A. Stocker (1980). Escherichia coli K317, formerly used to define colicin group E2, produces colicin E7, is immune to colicin E2, and carries a bacteriophage-restricting conjugative plasmid. *J Bacteriol* (**144)**, 524-531.

Masaki, H. & T. Ogawa (2002). The modes of action of colicins E5 and D, and related cytotoxic tRNases. *Biochimie* (**84)**, 433-438.

Nomura, N., H. Masai, M. Inuzuka & other authors (1991). Identification of eleven single-strand initiation sequences (ssi) for priming of DNA replication in the F, R6K, R100 and ColE2 plasmids. *Gene* (**108)**, 15-22.

Ogawa, T., K. Tomita, T. Ueda, K. Watanabe, T. Uozumi & H. Masaki (1999). A cytotoxic ribonuclease targeting specific transfer RNA anticodons. *Science* (**283)**, 2097-2100.

Schaller, K. & M. Nomura (1976). Colicin E2 is DNA endonuclease. *Proc Natl Acad Sci U S A* (**73)**, 3989-3993.

Senior, B. W. & I. B. Holland (1971). Effect of colicin E3 upon the 30S ribosomal subunit of Escherichia coli. *Proc Natl Acad Sci U S A* (**68)**, 959-963.

Toba, M., H. Masaki & T. Ohta (1988). Colicin E8, a DNase which indicates an evolutionary relationship between colicins E2 and E3. *J Bacteriol* (**170)**, 3237-3242.

Van der Wal, F. J., J. Luirink & B. Oudega (1995). Bacteriocin release proteins: mode of action, structure, and biotechnological application. *FEMS Microbiol Rev* (**17)**, 381-399.

Watson, R. J., T. Vernet & L. P. Visentin (1985). Relationships of the Col plasmids E2, E3, E4, E5, E6, and E7: restriction mapping and colicin gene fusions. *Plasmid* (**13)**, 205-210.