

Sequence Analysis of Replication Origin of Plasmid pCTP4 in γ -PGA-producing *Bacillus* Strain isolated from “Dan-douchi” in China

Iwamoto, Nobuhide

Microbial Genetics Division, Institute of Genetic Resources, Faculty of Agriculture, Kyushu University

Nakajima, Kazuhiko

Microbial Genetics Division, Institute of Genetic Resources, Faculty of Agriculture, Kyushu University

Saito, Hiroyuki

Microbial Genetics Division, Institute of Genetic Resources, Faculty of Agriculture, Kyushu University

Kaneko, Shinji

Microbial Genetics Division, Institute of Genetic Resources, Faculty of Agriculture, Kyushu University

他

<https://doi.org/10.5109/24076>

出版情報：九州大学大学院農学研究院紀要. 39 (3/4), pp.139-148, 1995-03. Kyushu University
バージョン：
権利関係：

Sequence Analysis of Replication Origin of Plasmid pCTP4 in *y*-PGA-producing *Bacillus* Strain isolated from “Dan-douchi” in China

Nobuhide Iwamoto, Kazuhiko Nakajima, Hiroyuki Saito,
Shinji Kaneko, Masahito Orima and Toshio Hara*

Microbial Genetics Division, Institute of Genetic Resources, Faculty of Agriculture,
Kyushu University, Fukuoka 812-81, Japan
(Received November 9, 1994)

The structure of a 1.6-kb *SphI*-*HindIII* DNA sequence necessary and sufficient for the replication of a 6.3-kb plasmid pCTP4 in *y*-PGA-producing *Bacillus* strain isolated from “dan-douchi” in China, which is responsible for *y*-polyglutamate production, has been characterized by using a trimethoprim-resistance gene derived from *B. subtilis* TTK24 chromosomal DNA as a selective marker. The 1.6-kb DNA sequence contains a *rep* gene encoding the protein (333 amino acids) essential for initiation of replication and a possible origin of replication. The predicted REP protein of pCTP4 has an overall homology with the REP proteins of pUH1 (73.9% identity), pLS11 (87.4%), pBAA1 (82.0%), and pFTB14 (76.6%) in *Bacillus* spp., pLP1 (46.5%) and pLAB1000 (32.4%) in *Lactobacillus* spp., and pUB110 (34.8%) and pC194 (21.9%) in *Staphylococcus aureus*, but has not any similarity with the REP protein of the staphylococcal plasmid pT181.

INTRODUCTION

“Douchi” is a native fermented product made from whole soybeans in China. There are two kinds of “douchi” as a raw fermented product : without or with salt, *i. e.*, “dan-douchi” and “xian-douchi.” “Xian-douchi” is made from steamed soybean and salt using the koji mold of *Aspergillus oryzae*, and it is found to greatly resemble “hamanatto” in Japan. “Dan-douchi,” the non-salty aging product of steamed soybean by the biological action of “natto” *Bacillus*-like microorganisms, is an adhesive fermented soybean with a noticeable odor of ammonia, and is, then considered to be the same as Japanese “natto” (Nakao, 1972). After fermentation, the raw “dan-douchi” is sun-dried, and then the dark-colored product is mixed with red pepper and ginger and aged for a while before being eaten. However, there are some striking and possesses some acidity, whereas “natto” is predominantly sticky and contains no acid (Bo, 1984).

We have reported that a 5.8-kilobase pair (kb) plasmid designated pUH1, which encodes the *y*-polyglutamate-production stimulating factor (*psf*) gene responsible for *y*-polyglutamate production, is distributed widely in a number of *Bacillus subtilis* (*natto*) strains isolated from a fermented soybean food, “natto” (Hara et al., 1981 and 1983). *B. subtilis* and *B. subtilis* (*natto*) should be considered as one species, but these two bacilli are found to be classified separately on the basis of whether biotin is

* Corresponding author.

essential for growth or not (Amaha et al., 1952; Kida et al., 1963). γ -Polyglutamate-producing *Bacillus* strain was isolated from "dan-douchi," which required biotin for growth and produced viscous substances, harbored a single plasmid pCTP4 (6.3 kb) (Hara et al., 1993). Recently, electron microscopy showed that a heteroduplex molecule between pUH1 and pLS11 of *Bacillus subtilis* IFO 3022 contains 1.11-kb and 2.14-kb double-stranded termini (Hara et al., 1986). The 2.0-kb *Bst*EII DNA fragment of pUH1, including the 1.11-kb double-stranded region, contains a 999-bp open reading frame, a promoter for the open reading frame, and a possible replication origin upstream of the promoter (Hara et al., 1991).

Analysis of the organization of various plasmids isolated from Gram-positive bacteria, such as pT181 (Projan and Novick, 1988), pC194 (Alonso and Taylor, 1987), pE194 (Dempsey and Dubnau, 1989), pFTB14 (Murai et al., 1987), pLP1 (Perez-Martin et al., 1988), pUB110 (Maciag et al., 1988), pBAA1 (Devine et al., 1989), and pLAB1000 (Josson et al., 1990), has shown that all of the information necessary for replication is located on fragments of about 1.5 kb in length. These fragments harbor a *rep* gene, encoding a protein essential for the initiation of replication (REP) and its corresponding target site.

This communication reports that the 1.6-kb fragment of pCTP4 contains a 999-bp open reading frame, a promoter for the open reading frame, and a possible replication origin upstream of the promoter. Significant homology was observed between the amino acid sequence predicted from the 999-bp open reading frame and those of similar putative REP proteins encoded by the other well-known Gram-positive replicons.

MATERIALS AND METHODS

Bacterial strains and plasmids. *Escherichia coli* JM101 and *B. subtilis* MI112 (*arg*-15, *leuB8*, *thr-5*, *recE4*) were used for a cloning host. The plasmids pCTP4 (Hara et al., 1993) and pATE1 (Hara et al., 1991) have been described previously, and plasmid pTL12 carrying the dihydrofolate reductase gene was kindly provided by Tanaka and Kawano (1980).

Media. LB broth and Panassay broth (Difco) for both *B. subtilis* and *E. coli*, Spizizen minimal medium for *B. subtilis*, and M9 minimal medium for *E. coli* were the same as those described previously (Hara et al., 1991). The cells carrying trimethoprim-resistant (Tmp^r) plasmids were grown in AA medium (Tanaka and Sakaguchi, 1978) containing 1 mg/ml of Tmp.

DNA manipulations. Plasmid DNAs from *B. subtilis* were prepared and purified as described previously (Hara et al., 1991). Restriction endonucleases, T4 DNA ligase, and bacterial alkaline phosphatase were purchased from Takara Shuzo Co., Ltd., and used as recommended by the manufacture. Degradation of DNA with exonuclease Ba131 (Takara Shuzo Co., Ltd.) was done by the procedure of Legerski et al. (1978).

Methods for transformation and assessment of *ori* function. *E. coli* was transformed by the method of Morrison (1977), and *B. subtilis* was transformed by using protoplasted cells (Chang and Cohen, 1979).

DNA sequencing. DNA fragments were subcloned into plasmids pUC18 and

pUC19, and DNA sequencing was done by the dideoxy chain termination method (Sanger *et al.*, 1977) with Sequenase (United States Biochemical Corporation, Ohio, USA). Nucleotide and amino acid sequences were analyzed by the Hitachi DNASIS system.

RESULTS AND DISCUSSION

Delimitation of the replication origin

To facilitate the identification of the replication region of pCTP4, we used the trimethoprim-resistant (T^{mp}^r) dihydrofolate reductase gene of *B. subtilis* 168. A schematic presentation of the constructed plasmids is given in Fig. 1. The source of the dihydrofolate reductase-coding gene was a T^{mp}^r strain, TTK24, of *B. subtilis* 168 (Tanaka and Sakaguchi, 1978) and has been cloned in the pBR322 plasmid of *E. coli*. DNAs from pTL12, carrying the T^{mp}^r dihydrofolate reductase gene, which was constructed by Tanaka and Kawano (1980), and pBR322 were both treated with *Eco*RI and *Hind*III, mixed and ligated by T4 ligase, then pATE1 was constructed. Plasmid pCTP4 was digested with *Sph*I and *Hind*III, and then the ends were filled in with the Klenow fragment to generate blunt ends. The DNA fragments were mixed and ligated to the *Aat*I site of pATE1 by T4 ligase, and then added to *B. subtilis* MI112 protoplasts. Several T^{mp}^r colonies were obtained on AA agar plates containing Tmp (1 mg/ml) and one of them was used for further study. A plasmid, pNH3, carried in such a T^{mp}^r colony had a molecular size of 8.2kb (Fig. 1). The physical map of pNH3 using various restriction enzymes is shown in Fig. 1.

To define the boundaries of a functional unit of pCTP4 replication, the 1.6-kb *Sph*I-*Hind*III fragment of pCTP4 was digested with selected restriction endonucleases to obtain a set of overlapping DNA fragments. The digests filled in with Klenow fragment were ligated with pATE1, introduced into *E. coli* by transformation, and selected for ampicillin resistance. The plasmid DNA preparations containing each generated fragment were tested for replication in *B. subtilis*. The results are summarized in Fig. 2. The constructed plasmid with the 1.4-kb *Sph*I-*Hae*III fragment (fragment 2 in Fig. 2) could replicate in the *B. subtilis* host, but the recombinant plasmid preparations containing the small fragments (fragment 3 and 4 in Fig. 2) could not replicate.

Nucleotide sequence of the replication region

The nucleotides of the 1.6-kb *Sph*I-*Hind*III fragment were sequenced by the method of Sanger (1977). Though the strategy is not shown, the nucleotides of both strands were sequenced using numerous restriction fragments to give enough overlapping regions (Fig. 3). Looking for possible open reading frames (ORFs), we found only one large frame (Fig. 2) designated rep, which consisted of 999 bp and encoded a protein molecule with 333 amino acids with a M_r of 38,582. Several conserved regulatory sequences similar to the *E. coli* and *Bacillus* consensus promoter sequences (McLaughlin *et al.*, 1981) were observed 5' upstream to the *rep* ORF.

Plasmid pNH3 was digested with *Bgl*II and then treated for 20 min with exonuclease Ba131 under the conditions in which about 50 bp per min were removed from

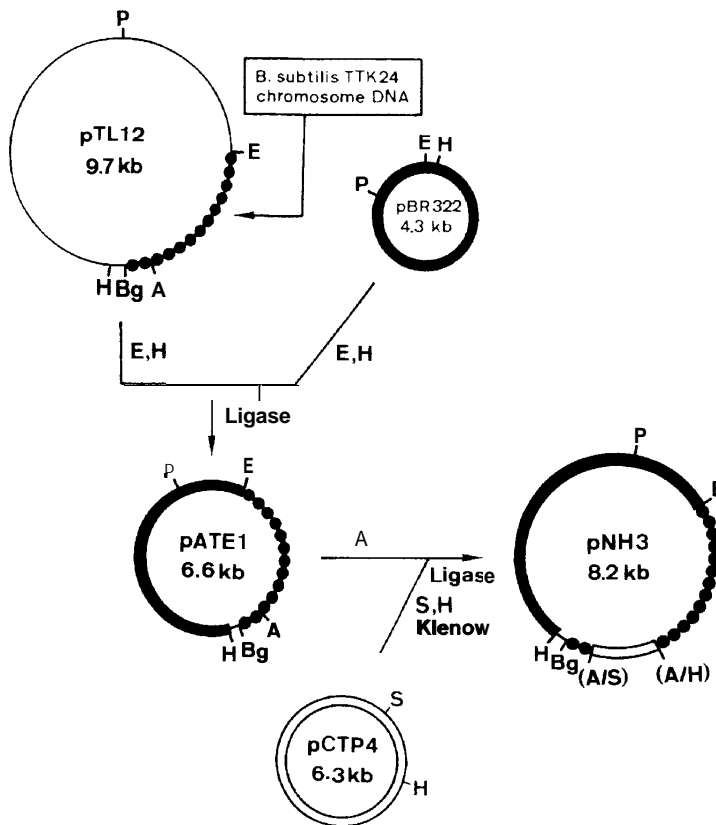


Fig. 1. Derivation of plasmids used in this study.

The chain of solid circles in the diagram indicates the DNA segment containing the Tmp^r gene of *B. subtilis* TTK24. Heavy and thin lines represent the regions of pBR322 and pTL12, respectively. Double lines represent the DNA fragment of pCTP4. A, *AatI*; Bg, *BglII*; E, *EcoRI*; H, *HindIII*; P, *PstI*; S, *SphI*.

each end of the DNA molecule. After ligation by T4 ligase, the DNA was transformed into *E. coli* and subsequently introduced into *B. subtilis* by protoplast transformation. Tmp^r transformants, which contain 1.4-kb fragment 5 in Fig. 2, were obtained at high efficiencies with a 168-bp deletion plasmid generated with Bal31 (fragment 5 in Fig. 2), while no transformants were obtained with a similar 255-bp deletion plasmid (fragment 6 in Fig. 2). It suggests that the putative replication origin of pCTP4 is between position 168 and 255. To discover the sequences essential for replication, a homology search was done to find whether there are sequences within this region conserved in the registered Gram-positive replicons. A 34-bp sequence was conserved in four plasmids, pCTP4, pLS11, pUH1, pBAA1 and pFTB14 in *Bacillus* spp. (Fig. 4), and the 14-bp sequence within this region was conserved in all eight plasmids in Gram-positive bacteria including *Lactobacillus* spp. and *Staphylococcus* spp. This conserved 14-bp sequence is found in the 55-bp region of pC194 shown by Gros *et al.* (1987) to

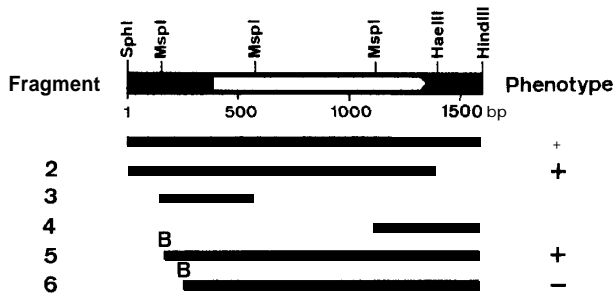


Fig. 2. Structure and replication activity of the derived fragments of the 1.6-kb *ori* fragment.

The open arrow in the restriction map indicates an open reading frame found in the sequence data (see Fig. 3). + and - indicate, respectively, ability and inability to replicate in the *B. subtilis* host. Derivatives were made by deletion with the indicated restriction enzyme or by Ba131 digestion (B) from the *Bgl*II site of *pNH3*.

have origin activity. Within this 14-bp sequence, in addition, the sequence CTTGATA is the sequence at which nicking of the plus-strand occurs in the initiation of replication of the coliphage ϕ X174 (Shlomag and Kornberg, 1980), and this conserved sequence was found in a hairpin region of ϕ X174.

Amino acid sequence homology of *pCTP4* REP protein with different Gram-positive replication proteins

The amino acid sequence of the REP protein coding region of *pCTP4* was compared with a number of amino acid sequence of proteins registered in GenBank by the homology search system GENAS (Kuhara *et al.*, 1984). Homologies between the predicted amino acid sequence of the REP protein of *pCTP4* and those of the REP protein of *pUH1*, *pLS11*, *pBAA1*, *pFTB14*, *pLP1*, *pLAB1000*, *pUB110* and *pC194* are illustrated in Fig. 5. The REP protein encoded on *pCTP4* showed a substantial degree of homology to three REP proteins in *Bacillus* plasmids: 87.4% identity with *REPPpLS11*, 82.0% with *REPPpBAA1*, 76.6% with *REPPpFTB14*, and 73.9% with *REPPpUH1*. The *REPPpCTP4* is similar to *REPPpLP1* (46.5% identity) and *REPPpAB1000* (32.4%) from lactic acid bacteria but also to *REPPpUB110* (34.8%) and *REPPpC194* (21.9%) from *S. aureus*. The REP protein of *pFTB14* stretches for 1,017 bp, a promoter region for *rep* expression, and a possible replication origin for *rep* expression, which is upstream of the promoter. The *rep* product is trans-active and essential for plasmid replication (Murai *et al.*, 1987). The tyrosine, which acts as an active site of REP in the rolling circle mechanism (Van Mansfeld *et al.*, 1986), was conserved in all of the Gram-positive initiation proteins (Fig. 5). Khan *et al.* (1982) identified the start site of *pT181* DNA synthesis within a 127-bp segment and showed that a 168 bp segment containing the replication start site is enough to initiate unidirectional replication. Furthermore, like REP protein of *pT181* (Projan and Novick, 1988), the protein of the *E. coli* plasmid *R6K* (Stalker *et al.*, 1982 ; Germino and Bastia, 1982) does not

SphI 50
 5'-GCATGCTGCGGAGATAACACGGGCTGCTGCGCAATGATGCGGGCATGGGTGCGTTGATCGTGCCGAGACTGAAA
 100 150
 AGCCGATTTTTGAAAAAAAATCTCCCCCCCCGCGAAGAATGGTTTTGATCTTTGGTTTTAGGTTTTAAAAAAGCC
 200
 GGCTGTTTTGAGCCCGATTTTTCGATTTTGGCGGCAAAATCGGTCCTTTTCTATCTTGACTATATAGAAAACACAT
 250 (-35) 300 (-10)
 CATTPTTCAAATCAGTTCAAAGCTTGTGTGCAAGGTTAGATCGGTTTTCACAGGTAATAAACCTCTCTCTGCTATT
 350 (sp)
 TTAAGGTGTCGAATCAAATAATAGAAATGCTAGAGAACTAGCTCAGTAGGAGCTTTTTTTTGGTATTATTATCATCTGAAA
 400 450
 ATG ATT ATA GCA TCC TCG AAG ATA AAA CCG CAA CAG GTA AAA AGC GGG ATT GGA AGG GGA
 Met Ile Ile Ala Ser Ser Lys Ile Lys Pro Gln Gln Val Lys Ser Gly Ile Gly Arg Gly
 500
 AAA AGA GAC GTA CGA ACC TCA TGG CCG ACT TAC GAA GCG TTA GAG AGT AAG ATT GGG GCA
 Lys Arg Asp Val Arg Thr Ser Trp Arg Thr Tyr Glu Ala Leu Glu Ser Lys Ile Gly Lys Ala
 550
 CCT TAC TAT GGC AAA AAG GCT GCC AAA CTA ATT AGT TGT GCA GAG TAT CTT TCG TTT AAG
 Pro Tyr Tyr Gly Lys Lys Ala Ala Lys Leu Ile Ser Cys Ala Glu Tyr Leu Ser Phe Lys
 600
 AGA GAC CGG AGA CGC GGC AAG TTA AAA CTG TAT CAA GCC CAT TTT TGT AAA GTG AGG TTA
 Arg Asp Arg Arg Arg Gly Lys Leu Lys Leu Tyr Gln Ala His Phe Cys Lys Val Arg Leu
 650
 TGC CCG ATG TGT GCG TGG CGT AGG TCG TTA AAA ATT GCT TAT CAC RAT AAG TTG ATC GTA
 Cys Pro Met Cys Ala Trp Arg Arg Ser Leu Lys Ile Ala Tyr His Asn Lys Leu Ile Val
 700 750
 GAG GAA GCC AAT CCG CAG TAC GGT TCT GGA TGG ATT TTT CTC ACG CCG ATG GTT CCG AAT
 Glu Glu Ala Asn Pro Gln Tyr Gly Ser Gly Trp Ile Phe Leu Thr Pro Met Val Arg Asn
 800
 GTC GAG GGA GAA CCG CTG AAG CCA CAA ATT TCT GAG ATG CAT AAA GGA TTT AGG AAA CTG
 Val Val Gly Glu Arg Leu Lys Pro Gln Ile Ser Glu Met His Lys Gly Phe Arg Lys Leu
 850
 TTC CAG TAC AAA AAR GTA AAA ACT TCG GTP GCT GGA TTT TTC AGA GCT TTA GAG ATT ACC
 Phe Gln Tys Lys Lys Val Lys Thr Ser Val Ala Gly Phe Phe Arg Ala Leu Glu Ile Thr
 900
 AAA AAT CAT GAA GAA GAT ACA TAC CAT CCT CAT TTT CAT GAG TTG ATA CCA GTA AGG AAA
 Lys Asn His Glu Glu Asp Thr Tyr His Pro His Phe His Glu Leu Ile Pro Val Arg Lys
 950
 AAT TAT TTT GGG AAA AAC TAT ATT AAG CAG TCG GAG TGG ACG AGC CTT TGG AAA AAG TCG
 Asn Tyr Phe Gly Lys Asn Tyr Ile Lys Gln Ser Glu Trp Thr Ser Lys Trp Lys Ser
 1000 1050
 GTG AAA TTG GAT TAC CCT CCG ATT GTT GAT ATT CGA AGA GIG AAAGCTAAAGCC AAG ATT
 Val Lys Leu Asp Tyr Pro Pro Ile Val Asp Ile Arg Arg Val Lys Gly Lys Ala Leu Ile
 1100
 GAC GCT GAA CTG ATT GAA AGC GAT GTG CCG GAA GCA ATG ATG GAG CAA AAA GCT GTT CTC
 Asp Ala Glu Leu Ile Glu Ser Asp Val Arg Glu Ala Met Met Glu Gln Lys Ala Val Leu
 1150
 GAA ATT TCT AAA TAT CCG GTT AAG GAT ACG GAT GTT GTG CCG GGC AAT AAG GTG ACT GAA
 Glu Ile Ser Lys Tyr Pro Val Lys Asp Thr Asp Val Val Arg Gly Asn Lys Val Thr Glu
 1200
 GAC ATT CTG AAC GCG GTG CTT TAT TTG GAT GAT GCG TTG GCA GCT CGA AGG TTA ATT GGA
 Asp Ile Leu Asn Ala Val Leu Tyr Leu Asp Asp Ala Leu Ala Ala Arg Arg Leu Ile Gly
 1250
 TTC GGT GGC ATT TTG AAG GAG ATA CAT AAA AGA CTG AAT CTT GGA GAT CCG GAG GAC GGC
 Phe Gly Gly Ile Leu Lys Glu Ile His Lys Arg Leu Asn Leu Gly Asp Arg Glu Asp Gly
 1300 1350
 GAT CTG GTC AAG ATT GAG GAA GAA GAT GAC GAG ATT GCA AAC GGC GCA TTT GAG GTT ATG
 Asp Leu Val Lys Lys Ile Glu Glu Glu Asp Asp Glu Ile Ala Asn Gly Ala Phe Glu Val Met
 1400 HaeIII
 GCT TAT TGG CAT CCT GGC ATT AAA AAT TAC ATA ATC AGA TAA AAAGCAGGCCCTTGTCCIGCTTT
 Ala Tyr Trp His Pro Gly Ile Lys Asn Tyr Ile Ile Arg *
 1450
 TTTTATACTCTAATAGTCAAATGTGTTGAAATATGTTTTGTGCTTTTTTATCTTTGTAGCTTTTTATGATTAATTGTC
 1500 1550
 TATGATAACCTAGGAATCTTTACATTGTCTGTGTTGAATAGTTACATAAACTGCTTGTCTTAGGATAGTTAAATA
 HindIII 1607
 TACTGTGTCATTGCTCCGAAAACAAAAGCTT-3'

Fig. 3. Nucleotide sequence of the 1.6-kb *SphI*-*HindIII* DNA fragment.

Nucleotide residues are numbered in the 5'-to-3' direction, beginning with the 5'-end residue originated from the *SphI* site. The deduced amino acid sequence is given below the nucleotide sequence. Putative promoter elements (-35, -10 and the ribosome binding site) are underlined.

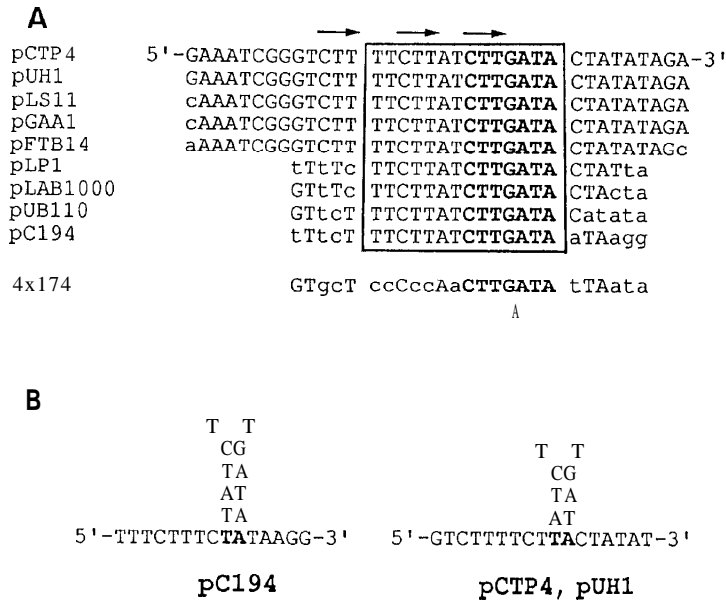


Fig. 4. (A) Homology at the origin for plus-strand synthesis among several plasmids in Gram-positive microorganisms. The nucleotides in boxes are conserved well among them. The sequence CTGATA, which is the site at which nicking of the plus strand occurs in phage ϕ X174, are boldface. The DNA bond between nucleotides G and A will be nicked by the REP protein and is indicated by an arrow. (B) Hairpin structure of the pC194 origin region proposed by S. D. Ehrlich and the predicted structure of origin region of plasmids pCTP4 and pUH1.

have any significant homology in its amino acid sequences with those of the REP proteins of pCTP4, pUH1, pLS11 and also pFTB14 deduced from the ORFs (data not shown). Comparison of dendrograms made on the basis of 16S rRNA hybridizations and the percentage of homology on REP protein are illustrated in Fig. 6. The degree of homology of REP_{pCTP4} with the REP proteins of pLP1 and pLAB1000 in *Lactobacillus* spp. strains was almost the same as those with the REP proteins of pUB110 and pC194 in *S. aureus*. Taxonomic studies (based on the 16S rRNA similarities) showed that *Bacillus* spp. and *S. aureus* strains are more closely related to each other than *Staphylococcus* spp. and *Lactobacillus* spp. (Stackebrandt and Teuber, 1988). The plasmid homologies suggest an exchange of plasmid replicons by recent horizontal transfer through the different genera, including *B. subtilis* (*natto*).

REFERENCES

- Alonso, J. C. and R. H. Taylor 1987 Initiation of plasmid pC194 replication and its control in *Bacillus subtilis*. *Mol. Gen. Genet.*, 210: 476-484


```

REPPCTP4      10      20      30      40      50      60
                M I A S S K I K P O Q V K S G I G K G K R D V R T S W M T Y E A L E S K I G A P Y
                * * * * *
REPPUH1      M I S L L K R T P P Q Q V K S G I G G K R D R I L W L S T M K L Y R K L V L V L
                * * * * *
REPPLS11     M I S I L K I K P O Q V K S G I G K G K R D G T S W M T Y E A L E S K I G A P Y
                * * * * *
REPPBAA1     M A S H Y E A L E S K I G A P Y
                * * * * *
REPPFTB14   M Y S E N D Y R I L E D K T A T G K R D W R K R R A N L M A E H Y E A L E K R I G A V Y
                * * * * *
REPPPL1     M S E I P E D K T E N G K V P W R R K I E N V P Y A E Y L A I L E F K R A N G O P
                * * * * *
REPPLAB1000 M S K K I L K D V S R N R K E R P W R R K I E N I Q Y A E Y L A I L N F K A N R V K
                * * * * *
REPPUB110   M G V S F N I M C P N S I F D E K S R V L V D K T K S G R V P W R R K I A N V D Y F E L L H I L E F K A E R V I
                * * * * *
REPPC194    M C Y N M E Y T E K K Q R Q V Q F K I K R H I
                * * * * *
REPPCTP4      70      80      90      100     110     120
                Y G K K A A K I T S C A E Y I S F K R D R R R K I K L Y Q A H F C V K L C P M C A W R R S L K I A V N K L I V E E
                * * * * *
REPPDH1     T M A K K L N C A V V N N V F L N E T R R R K L K L Y Q A H F C V K L C P M C A W R R S L K I A V N K L I V E E
                * * * * *
REPPLS11     Y G K K A E L I S C A E Y I S F K R D R R P G K L K L Y Q A H F C V K L C P M C R W R R S L K I A V N K L I V E E
                * * * * *
REPPBAA1     Y G K K A E L I S C A E Y I S F K R D P E T G R L K L Y Q A H F C V K L C P M C A W R R S L K I A V N K L I V E E
                * * * * *
REPPFTB14   Y G K K A E L S E C A E H L S F K R D P E T G R L K L Y Q A H F C V K L C P M C A W R R S L K I A V N K L I V E E
                * * * * *
REPPPL1     G C G E V L F R K I G E H ----- L K L Y Q A H F C V K L C P M C A W R R S L K I A V N K L I V E E
                * * * * *
REPPLAB1000 E C G E V L F R V A D D E ----- G R L R L Y Q T W C F S R L C L C N W R R S M G O S M Q L M W I D E
                * * * * *
REPPUB110   D C A E I L E Y K Q N R E T ----- G E R K L Y R V W F C K S R L C P M C N W R R A M H K I Q S O K V V A E
                * * * * *
REPPC194    G E N Q M D L V E D C N T F L S F V A D K T L E R K Q L Y K A N S K N R F C P V C A W R K A R K D A L G L S M H M O Y
                * * * * *
REPPCTP4      130     140     150     160     170     180
                A N R Q Y C G G W I - F L T L T V R N V - V G E R L K P Q I S E M H M G F R K L - P O Y K V K T S V L G F P R A L E I T N
                * * * * *
REPPDH1     A N R Q Y C G G W I - F L T L T V R N V - E G D G L K P I A D M H M G W N R L - P O Y K V K V A T L G Y F R A L E I T N
                * * * * *
REPPLS11     A N R Q Y C G G W I - F L T L T I R N V - K G E R L K N O I S A M H M G F R K L - P O Y K V K T S V L G F P R A L I A K N
                * * * * *
REPPBAA1     A N R Q Y C G G W I - F L T L T V R N V - K G E R L K P Q I S E M H M G F R K L - P O Y K V K T S V L G F P R A L E I T N
                * * * * *
REPPFTB14   A N R Q Y C G G W I - F L T L T V R N V - K G E R L K P Q I S E M H M G F R K L - P O Y K V K T S V L G F P R A L E I T N
                * * * * *
REPPPL1     A N R Q Y G K G R - F L T L T V K N A S A E L K V L S L A L T K A P N K L K T R Y K V T K N L L G V I R S T E I T V N
                * * * * *
REPPLAB1000 A H K Q R K T G R - F L T L T A S N A - S G E N L K O E V R K M G R A I S K L - P O Y K P A N M L L G V R S T E I T I N
                * * * * *
REPPUB110   V I K Q P T V K W L P L T L T V K N V D G E L M K L S D M A G S F R M - M O Y K I K N N L V G F R M A T E V T I N
                * * * * *
REPPC194    I K Q E K K E P I - F L T L T P F N V M S O E - L E N E I K R Y N S E F R K L - I E R K K V G S V I K G V R K L E I T Y N
                * * * * *
REPPDH1     H E E D T Y P H P H E L L P V R K N Y F - - G N Y I Q A E W T S L W K K A M L D Y T P I V D I R V R V G - - K A L I D A
                * * * * *
REPPUH1     H E E D T Y P H P H V L L P V R K S I F - - T H N Y I Q S E W S L W E R A M K L D Y T P I V D I R V R V G - - K A R I D A
                * * * * *
REPPLS11     H E E D T Y P H P H V L L P V R K N Y F - - G N Y I Q A E W T S L W K K A M L D Y T P I V D I R V R V G - - K A V L D A
                * * * * *
REPPBAA1     H E E D T Y P H P H V L L P V R K N Y F - - G N Y I Q A E W T S L W K K A M L D Y T P I V D I R V R V G - - K A K I D A
                * * * * *
REPPFTB14   H E S D T Y P H P H V L L P V R K N Y F - - G N Y I Q A E W T S L W K R A M K L D Y T P I V D I R V R V G - - R A K I D A
                * * * * *
REPPPL1     E O D G S V N Q H L H V L L P V K S S Y F K N S N Y L A G A E M A K L W Q K A L K V D Y P V V H V Q A V K A - N E R K G D T
                * * * * *
REPPLAB1000 - K N G T Y H Q H M H V L F V K P T Y K D S N Y N I G A E S S L M Q R A K K L D Y P I V R V E A V R S - N K A K C K N
                * * * * *
REPPUB110   N K D N S Y N Q H M H V L V C V P T Y F K N T E N Y N G K W I Q P W K A K K L D Y D P N D K V G M I R P K N K Y S D -
                * * * * *
REPPC194    K K D N S Y N Q H M H V L I A V N K S Y F D K R Y I S Q E W L E L W - - - - - R V T G I S E I T O V O V O K I R O N N N
                * * * * *
REPPCTP4      250     260     270     280     290
                E L I E S D V R E A M H M Q K A V L E I S K Y P V K D T D V V R G N K V T E D I L N A V L Y L D D A L A A R R L I G Y G
                * * * * *
REPPUH1     E Q I E N D V R E A M H M Q K A V L E I S K Y P V K D T D V V R G N K V T E D N L N T V L Y L D D A L A R R L I G Y G
                * * * * *
REPPLS11     E Q I E N D V R E A M H M Q K A V L E I S K Y P V K D T D V V R G N K V T E D N L N T V L Y L D D A L A R R L I G Y G
                * * * * *
REPPBAA1     D L I E N G V R E A M H M Q K A V L E I S K Y P V K D T D V V R G N K V T E D N L N T V L Y L D D A L A R R L I G Y G
                * * * * *
REPPFTB14   E Q I E S D V R E A M H M Q K A V L E I S K Y P V K D T D V V R G S V T P D N L N T V F Y L D D A L A R R L I G Y G
                * * * * *
REPPPL1     S L ----- Q A S A E E T A K Y E V K S A D Y M T A D D - - E R N L V V I K N L E Y A L A G T P Q I S Y G
                * * * * *
REPPLAB1000 S L ----- I A S A Q E T A K Y Q V K S D I I T W O D - - E R D L Q W S E D L E Q G L A G A R Q I S Y G
                * * * * *
REPPUB110   ----- I Q S A I D E T A K Y P V K D T D - - F M T D O E R K N K R L S D L E B G L H R K R L I S Y D
                * * * * *
REPPC194    K E L ----- Y E M A K Y S G K S D Y L I - N K S K L
                * * * * *
REPPCTP4      310     320     330     340     350
                G I L K E I H K R L M L G D R E D G D L V K I E E E D D E V - - A N G A F E V M A Y W H P G I K N Y I I R
                * * * * *
REPPDH1     G I L K E I H K E L M L G D A E D G D L V K I E E E D D E V - - A N G A F E V M A Y W H P G I K N Y I I R
                * * * * *
REPPLS11     G M L K E I H K E L M L G D A E D G D L V K I E E E D D E V - - Q N J A F E V M A Y W H P G I K N Y I I K
                * * * * *
REPPBAA1     G I L K E I H K E L M L G D A E D G D L V K I E E E D D E V - - A N G A F E V M A Y W H P G I K N Y I I K
                * * * * *
REPPFTB14   G I L K E I H K E L M L G D A E D G D L V K I E E E D D E V - - A N G A F E V M A Y W H P G I K N Y I I K
                * * * * *
REPPPL1     G L L K Q I K O D L E D V E N G D L V H V G D E D Y T E Q M E A E R V V A K W D N P K Q N Y F W
                * * * * *
REPPLAB1000 G L P K E I R N Q L Q L E D V D A H L E N V - - D D D K Y I D E V Y E V V A K W D N P K Q N Y F W
                * * * * *
REPPUB110   G L L K E I H K M L M L D D T E G D L - - I H T D D E K A E D G - F S I I A W N W E R K N Y F I K E
                * * * * *
REPPC194

```

Fig. 5. Amino acid sequence comparison of the REPPCTP4 protein with different Gram-positive replication proteins.

The original names of these proteins have been replaced by REP, followed by a letter code indicating the plasmid from which they originate. The amino acid identities with several plasmids in Gram-positive microorganisms with REPPCTP4 are indicated (*). Amino acid numbers follow the sequence of the REP protein from the amino-terminal methionine to the carboxy terminus. Gaps have been inserted to gain maximum matching. The one-letter amino acid code has been used. The tyrosine residue (marked with arrow) probably involved in the binding of the REP protein to the DNA is indicated by an arrow.

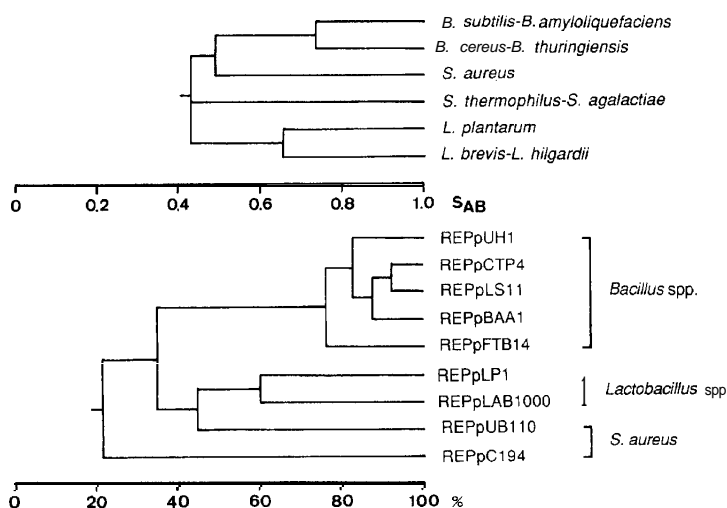


Fig. 6. Comparison of dendrograms made on the basis 16S rRNA hybridizations (A) and the percentages of homology on Rep proteins among several plasmids in Gram-positive microorganisms (B).

The microorganisms from which the plasmids originate are shown. Dendrograms were made by the unweighted average pair group method (Sneath and Sokal, 1973).

- Amaha, M. and K. Sakaguchi 1952 *Bacillus zoku saikin no hanshoku saibou narabini houshi no eiyououkyu nituite*. *Nippon Nougakagaku Kaishi*, 26 : 353-359 (in Japanese)
- Bo, T.A. 1984 Origin of douchi, and its production technique. Part 2. *Nippon Jozo Kyokai Zasshi*, 79 : 395-402 (in Japanese)
- Chang, S. and S. N. Cohen 1979 High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. *Mol. Gen. Genet.*, 168: 111-115
- Dempsey, L. A. and D. A. Dubnau 1989 Localization of the replication origin of plasmid pE194. *J. Bacteriol.*, 171 : 2866-2869
- Devine, K. M., S. T. Hogan, D. G. Higgins and J. McConnell 1989 Replication and segregational stability of *Bacillus* plasmid pBAA1. *J. Bacteriol.*, 171 : 1166-1172
- Germino, J. and D. Bastia 1982 Primary structure of the replication initiation protein of plasmid R6K. *Proc. Natl. Acad. Sci., U. S. A.*, 79: 5475-5479
- Gros, M. F., H. te Riele and S. D. Ehrlich 1987 Rolling circle replication of single-stranded DNA pC194. *EMBO J.*, 6 : 3863-3869
- Hara, T., A. Aumayr and S. Ueda 1981 Characterization of plasmid deoxyribonucleic acid in *Bacillus natto*: evidence for plasmid-linked PGA production. *J. Gen. Appl. Microbiol.*, 27 : 299-305
- Hara, T., A. Ishizaki and S. Ueda 1986 Formation of heteroduplex molecules between plasmids pUH1 and pLS11 in polyglutamate-producing *Bacillus* strains. *Agric. Biol. Chem.*, 50 : 2391-2394
- Hara, T., S. Nagatomo, S. Ogata and S. Ueda 1991 Molecular structure of the replication origin of a *Bacillus subtilis* (*natto*) plasmid, pUH1. *Appl. Environ. Microbiol.*, 57 : 1838-1841
- Hara, T., K. Nakajima, H. Saito, A. Ishizaki, S. Ogata and S. Ueda 1992 Sequence analysis of replication origin of plasmid pLS11 of *Bacillus subtilis* IFO 3022. *Biosci. Biotech. Biochem.*, 56 : 223-227
- Hara, T., S. Ogata and S. Ueda 1993 Plasmid distribution in γ^- polyglutamate-producing *Bacillus* strains isolated from "dan-douchi," a "Natto"-like non-salty fermented soybean food in China.

- J. Gen. Appl. Microbiol.*, **39** : 75-82
- Hara, T., J. R. Zhang and S. Ueda 1983 Identification of plasmids linked with polyglutamate production in *Bacillus subtilis* (*natto*). *J. Gen. Appl. Microbiol.*, **29**: 345-354
- Josson, K., P. Soetaert, F. Michiels, H. Joos and J. Mahillon 1990 *Lactobacillus hilgardii* plasmid pLAB1000 consists of two functional cassettes commonly found in other Gram-positive organism. *J. Bacteriol.*, **172** : 3089-3099
- Kahn, S. A., G. K. Adler and R. P. Novick 1982 Functional origin of replication of pT181 plasmid DNA is contained within a 168-base-pair segment. *Proc. Natl. Acad. Sci., USA*, **79**: 4580-4584
- Kida, S., W. Hashida and S. Terumoto 1963 *Natto oyobi nattokin nikansuru eiyougakuteki kenkyu. HakkokogakuKaishi*, **34**: 542-546
- Kuhara, S., F. Matsuo, S. Futamura, A. Fujita, T. Shinohara, T. Takagi and Y. Sakaki 1984 GENAS : a database system for nucleic acid sequence analysis. *Nucleic Acids Res.*, **12** : 89-99
- Legerski, R. J., J. L. Hodnett and H. B. Gray Jr. 1978 Extracellular nucleases of *Pseudomonas* BAL31. III. use of the double-strand deoxyribonuclease activity as the basis of a convenient method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. *Nucleic Acids Res.*, **5**: 1445-1464
- Maciag, I. E., J. -F. Viret and J. C. Alonso 1988 Replication and incompatibility properties of plasmid pUB110 in *Bacillus subtilis*. *Mol. Gen. Genet.*, **212**: 232-240
- McLaughlin, J. R., C. L. Murray and J. C. Rabinowitz 1981 Unique features in the ribosome binding site sequence of the Gram-positive *Staphylococcus aureus* β -lactamase gene. *J. Biol. Chem.*, **256** : 11283-11291
- Morrison, D. A. 1977 Transformation in *Escherichia coli* : cryogenic preservation of competent cells. *J. Bacteriol.*, **132** : 349-351
- Murai, M., H. Miyashita, H. Araki, T. Seki and Y. Oshima 1987 Molecular structure of the replication origin of *Bacillus amyloliquefaciens* plasmid pFTB14. *Mol. Gen. Genet.*, **210** : 92-100
- Nakao, S. 1972 *Ryori no Kigen*. NHK Press, Tokyo, pp.121 (in Japanese)
- Perez-Martin, J., G. H. del Solar, A. G. de la Campa and M. Espinosa 1988 Three regions in the DNA of plasmid pLS1 show sequence-directed static bending. *Nucleic Acids Res.*, **16** : 9113-9126
- Projan, S. J. and R. Novick 1988 Comparative analysis of five related staphylococcal plasmids. *Plasmid*, **19** : 203-221
- Sanger, F., S. Nicklen and A. R. Coulson 1977 DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci., U. S. A.*, **74**: 5463-5476
- Shlomai, J. and A. Kornberg 1980 An *Escherichia coli* replication protein that recognizes a unique sequence within a hairpin region in ϕ X174 DNA. *Proc. Natl. Acad. Sci., U. S. A.*, **77** : 799-803
- Sneath, P. A. and R. R. Sokal 1973 *Numerical Taxonomy. The Principle and Practice of Numerical Classification*. W. H. Freeman & Co., San Francisco
- Stackebrandt, E. and M. Teuber 1988 Molecular taxonomy and phylogenetic position of lactic acid bacteria. *Biochimie*, **70** : 317-324
- Stalker, D. M., R. Kolter and D. R. Helinski 1982 Plasmid R6K DNA replication. *J. Mol. Biol.*, **161**: 33-43
- Tanaka, T. and N. Kawano 1980 Cloning vehicles for the homologous *Bacillus subtilis* host-vector system. *Gene*, **10**: 131-136
- Tanaka, T., M. Kuroda and K. Sakaguchi 1977 Isolation and characterization of four plasmids from *Bacillus subtilis*. *J. Bacteriol.*, **129** : 1487-1494
- Tanaka, T. and K. Sakaguchi 1978 Construction of a recombinant plasmid composed of *B. subtilis* leucine gene and a *B. subtilis* (*natto*) plasmid: its use as cloning vehicle in *B. subtilis* 168. *Mol. Gen. Genet.*, **165**: 269-276
- Van Mansfeld, A. B. M., H. A. A. M. Teeffelen, P. D. Baas and H. S. Janz 1986 Two juxtaposed tyrosyl-OH groups participate in ϕ X174 gene a protein catalyzed cleavage and ligation of DNA. *Nucleic Acids Res.*, **14** : 4229-4238