J. Biochem. 128, 869–875 (2000)

Efficient Cloning and Engineering of Giant DNAs in a Novel *Bacillus subtilis* Genome Vector

Mitsuhiro Itaya,*[,1] Taeko Nagata,* Toshihiko Shiroishi,† Kyoko Fujita,* and Kenji Tsuge*

**Mitsubishi Kasei Institute of Life Sciences, 11 Minamiooya, Machida, Tokyo 194-8511; and †Mammalian Genetics Laboratory, National Institute of Genetics, 1111 Yata, Mishima, 411-8540*

Received July 21, 2000; accepted September 13, 2000

The Genome of *Bacillus subtilis* 168 was used for cloning and engineering of large-sized DNAs. A mouse genomic DNA of approximately 120 kb was cloned into a locus of the *B. subtilis* genome by ordered assembly of 20- to 50-kb mouse DNA segments. Cloned mouse DNA, maintained stably, was engineered through *B. subtilis* transformation and recombination. Creation of an I-*Ppo*I recognition sequence at both ends of the insert facilitated its isolation by pulsed field gel electrophoresis. The basic concept of genome vector technology is suited to the handling of DNAs larger than 100 kb.

Key words: genome vector, positive selection, recombination, transformation.

DNA cloning technology emerged in the late 1970s and led to great success in DNA transfer beyond the species barrier (*1*). Rapid and reliable methods for conducting genomic engineering of large-sized DNA are primarily useful for the genome project including mapping and contig constructions. They are also applicable for the construction and analysis of transgenic mutants. Several vectors were established in which the clonable size exceeded 100 kb, including YAC (yeast artificial chromosome) vectors in *Saccharomyces cerevisiae* (*2*), and BAC (*3*) and PAC (*4*) systems in *Escherichia coli* derived from F-plasmid and phage P1 replicons, respectively. In spite of the great contribution to the cloning of giant DNAs, two major areas remain to be improved. One is the stability of the cloned insert. The other is the method to engineer cloned inserts. The latter has been partially improved for BAC clones by introduction of Cre-lox, a site-specific recombination system (*5*).

We introduced a novel giant DNA manipulation system using the 4,215-kb genome of *Bacillus subtilis* 168 as a cloning vehicle (*6*). Examples of cloning, manipulation, and recovery of approximately 120-kb mouse genomic DNA using the *B. subtilis* genome-based manipulation vector, referred to hereafter as the BGM vector, are presented.

MATERIALS AND METHODS

1.1 Bacterial Transformation—Bacteria were grown at 37°C unless specified. Transformation of *E. coli* and *B. subtilis* was carried out as previously described (*7*). Selection conditions for both *E. coli* and *B. subtilis* were 15 μg/ml tetracycline (Tc), 5 μg/ml chloramphenicol (Cm), and 5 μg/ml erythromycin (Em). Neomycin-resistant *B. subtilis* transformants were selected at 3 μg/ml. Blasticidin S (BS)–resistant transformants were selected using concentrations of 150 μg/ml for *E. coli* and 500 μg/ml for *B. subtilis*. *B. subtilis* genomic DNA in liquid or in agarose plug was prepared as described previously (*7*).

1.2 B. subtilis Strain Specific for Cloning Non-Cognate DNA (BGM vector)—*B. subtilis* strain BEST6016 has two extra DNA segments that are described in Ref. *8*. One is a 4.3-kb sequence of *E. coli* plasmid pBR322. pBR322, opened at the *Pvu*II site, was inserted into the *Not*I site of the *proB* gene at the position of 1,378 kb (*8–10*). The pBR322 sequence indicated as a barcode in the figures remains at both ends of the cloned insert, providing integration sites to extend adjacent regions. The other segment is a neomycin resistance gene [Pr-neo] in the *yvfC* gene at the position of 3,516 kb (*8*, *10*). Expression of [Pr-neo] is regulated by a lambda Pr promoter, to which the lambda *cI* repressor gene product specifically binds to suppress transcription. BEST6016, a BGM vector, was derived from strain RM125, which had lost the restriction-modification system (*11*). BEST6016 derivatives in this study are collectively called Bsu genome clones.

1.3 E. coli Plasmids—The bacterial plasmids used in this study are listed in Table 1. pCISP303B, pCISP310B, and pCISP311B were constructed to deliver an LPS array (see "RESULTS") in the cloning locus of the genome vector (BEST6016). pCISP303B was derived from pBRCI-BS (*8*) as follows. Blasticidin S resistance gene (*bsr*) of pBRCI-BS was replaced by a spectinomycin resistance gene (*spc*) isolated from pBEST518 (*14*) to yield pCISP301. We refer to the 2.6-kb segment containing the *cI* gene and spectinomycin resistance gene as the *cI* cassette (shown as V in Fig. 1), hereafter designated as [cI]. The *cI* gene of pCISP301 had two *Hin*dIII sites and a *Bam*HI site. The first *Hin*dIII site was converted to AAACTT and the second one to AAGTTT by site-directed mutagenesis using ExSITE PCR-Based Site-Directed Mutagenesis Kit (Stratagene, La jolla, CA). In addition, the *Bam*HI site located between the Shine-Dalgano sequence and the initiation codon of the *cI* gene was converted to GGACTC, giving rise to pCISP303#6. These conversions do not alter the amino acid sequence of the cI gene product. pCISP303B was derived from pCISP303#6 by *Bam*HI digestion and re-ligation to remove a 375-bp sequence originated from pBR322. pCISP304 was constructed as follows. [cI] of pCISP303#6 obtained by *Eco*RI

[1] To whom correspondence should be addressed. Tel: +81-42-724-6254, Fax: +81-42-724-6316, E-mail: ita@libra.ls.m-kagaku.co.jp
Abbreviation: BAC, bacterial artificial chromosome.

digestion was cloned into the *Eco*RI site of pBR322 whose unique *Bam*HI site had been removed by T4 DNA polymerase-mediated blunting. Plasmid pCISP304, thus obtained, had a unique *Bam*HI site at the right end [cI]. This *Bam*HI site was converted to a *Bam*HI–I-*Ppo*I recognition sequence (5′-AGCTATGACTCTCTTAAGGTAGCCAAAGGATCC-3′, counterclockwise of pCISP303B in Fig. 1), resulting in pCISP305. pCISP306A was obtained by conversion of the unique *Eco*RI site residing left of the [cI] of pCISP305 to 5′-GAATTTGGCTACCTTAAGAGAGTCATGTTAACGAATTC-3′ (including I-*Ppo*I, *Hpa*I, and *Eco*RI sites in that order). Blasticidin S resistance gene (*bsr*) of pBEST402 (*12*) prepared by *Pst*I digestion and blunted by T4 DNA polymerase, was inserted into the *Pvu*II site of pCISP306A to yield pCISP307A. pCISP310B was obtained by insertion of a *cat* gene into the unique *Hpa*I site of pCISP307A. The *cat* gene was prepared from pBEST4 (*12*) using *Hin*cII. Similarly, insertion of *erm* (erythromycin re-

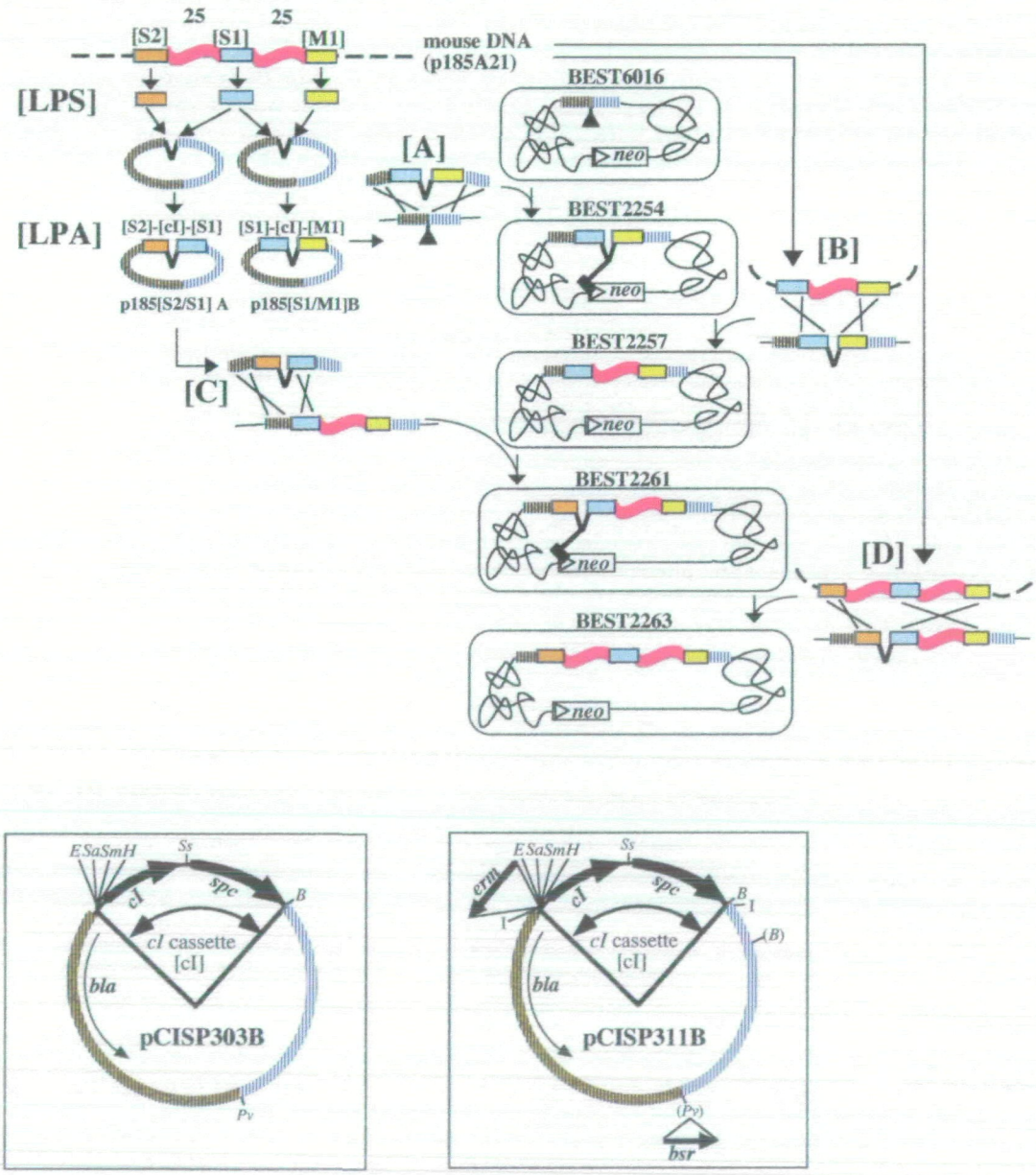

Fig. 1. **Cloning principles in the *B. subtilis* genome vector (BGM vector).** Mouse DNA between Landing Pad Sequence (LPS) [S2] and [M1] is shown on top. Two LPSs are aligned in pCISP vector (shown as a circular barcode). The LPA (*LP*S *A*rray) was transferred in BEST6016 genome *via* homologous recombination with pBR322 in the genome (bar code). Only the cloning locus and [Pr-neo] locus are shown in BEST strains. The structure of transformation intermediates [A], [B], [C], and [D], where expected homologous recombinations are indicated by X, is shown. The hammer in BEST2254 and BEST2261 indicates repression of neomycin resistance gene (*neo*) by a CI product encoded by the *cI* gene. Details are described in the text. Insert: Structure of pCISP303B and pCISP311B. pCISP310B contains *cat* instead of *erm* in pCISP311B. Restriction enzyme sites are; *B*, *Bam*HI; *E*, *Eco*RI; *H*, *Hin*dIII; *I*, I-*Ppo*I; *Sa*, *Sac*I; *Sm*, *Sma*I; *Ss*, *Sse*8387I, and *Pv*, *Pvu*II. *cI* gene (*cI*) plus spectinomycin resistance gene (*spc*), designated as *cI*-cassette [cI], are illustrated as V in other figures. pBR322 regions, the left and right half divided by the *Pvu*II site, correspond to those in barcode in other figures.

TABLE I. ***E. coli*** **plasmids constructed in this study.**

Plasmids	Construction or features[a]	
pCISP303B	ApR, SpR	(This study, section 2.3)
pCISP310B	ApR, SpR, CmR, BSR	(This study, section 2.3)
pCISP311B	ApR, SpR, EmR, BSR	(This study, section 2.3)
pBAC108L	BAC vector (7.3 kb), CmR	(*3*)
p185A21	A BAC clone	(Shiroishi, unpublished. see section 2.4)
pBAC108Lbsr	*bsr*-marked pBAC108L	(This study, section 2.3)
pBR::BAC101		(This study, section 2.3)
pBR::BAC102		(This study, section 2.3)
LPS clones in pCISP303B isolated from p185A21		
p185N1	3.2-kb *Bam*HI fragment [N1] carries a *Not*I	
p185N201	5.0-kb *Bam*HI fragment [N2] carries a *Not*I	
p185S1	9.0-kb *Bam*HI fragment [S1] carries a *Sal*I	
p185S2	5.5-kb *Bam*HI fragment [S2] carries a *Sal*I	
p185M1[c]	2.6-kb *Bam*HI fragment [M1] carries a *Mlu*I	
LPS clone derivatives		
p185N1Cm	*cat* inserted in the *Not*I site of p185N1.	
p185N2(I)	1.9-kb internal *Hin*dIII fragment from p185N201 inserted in *Hin*dIII site of pCISP310B.	
p185N202	a 6.6-kb *Eco*RI–[b]*Eco*RV fragment including the [cI] from p185N201 cloned in *Eco*RI–*Eco*RV of pBR322.	
p185S1T	*tet* cassette (1.9-kb/*Sma*I) from pBEST306 (*13*) into *Eco*RV site of p185S1.	
p185S101	a 2.0-kb internal *Bgl*II fragment of p185S1 cloned into *Bam*HI site of pCISP303B.	
p185S201	a 4.4-kb *Bam*HI–*Bgl*II fragment from p185S2 cloned into *Bam*HI site of pBEST4C (*12*) to convert 4.4 kb *Eco*RI fragment.	
p185S210	a 4.4-kb *Bam*HI–*Bgl*II fragment from p185S2 cloned into *Bam*HI site of pCISP311B.	
p185M201	2.6-kb *Bam*HI fragment from p185M1 into *Bam*HI site of pBEST4C (*12*) to convert 2.6 kb *Eco*RI fragment.	
LPA plasmid[d]		
p185[S1/M1]B	internal 3.7-kb *Eco*RI fragment of p185S1 cloned into *Eco*RI site of p185M1.	
p185[M1/N2]A	a 2.6-kb *Eco*RI fragment from p185M201 into *Eco*RI site of p185N202 (4.0 kb).	
p185[S2/S1]A	a 4.4-kb *Eco*RI fragment of p185S201 cloned into *Eco*RI site of p185S101.	
p185S2BACL′	a 3.8-kb *Hin*dIII fragment of pBR::BAC101 cloned into *Hin*dIII site of p185S210.	

[a]ApR, ampicillin resistance; SpR, spectinomycin resistance; CmR, chloramphenicol resistance; BSR, blasticidin S resistance; EmR, erythromycin resistance. [b]*Eco*RV site came from the mouse genome, residing approximately 0.7 kb inside from the right junction of the BAC vector. [c]The 2.6-kb *Bam*HI fragment shortened. See "RESULTS" section 3.7. [d]/ indicates the [cI] cassette.

sistance gene) into the *Hpa*I site of pCISP307A resulted in pCISP311B (Fig. 1). *erm* was prepared from pBEST701 (*14*) using *Pst*I and blunted. The *bsr* of pCISP310B and pCISP311B was used to discriminate plasmid integration into the *B. subtilis* genome (*12*).

The BAC vectors supplying plasmids were constructed as follows. A *bsr* gene (0.65 kb) prepared by *Hin*dIII digestion from pBEST402 (*12*) was inserted in the *Hin*dIII site of pBAC108L to give pBAC108Lbsr. pBAC108Lbsr was opened with *Eco*RV and inserted into the *Sma*I site of pBREm (*9*) to produce pBR::BAC101 and pBR::BAC102 with different orientations. These plasmids were constructed to provide BAC vector sequence as LPA. BAC vector sequence was used to transfer the end of the BAC insert in BGM vector. For example, a 3.8-kb fragment designated as [L] of p185S2BACL′ shown in Fig. 2 was obtained from this by use of *Hin*dIII.

All *E. coli* plasmids used in this study were prepared from 50-ml *E. coli* cultures and purified by CsCl gradient centrifugation in the presence of ethidium bromide (*1*). Type II restriction enzymes and T4 DNA ligase were obtained from Toyobo (Tokyo), except *Not*I (Takara Shuzo, Kyoto). Southern hybridization was done as described previously (*7*).

1.4 Mouse DNA—A BAC clone p185A21 was obtained by screening a BAC library of mouse strain 129. This clone was mapped in the distal mouse chromosome 11 (Shiroishi, T., unpublished observations) and had approximately 120 kb of the insert. Details are shown in Fig. 2.

RESULTS

2.1 Principle for Cloning in the B. subtilis Genome Vector (BGM Vector)—Basic schemes are presented in Fig. 1. BGM vector was developed based on a different concept from that of plasmid-based vectors. In contrast to plasmid vectors, to which fragmented DNA is directly ligated, cloning in the BGM vector requires flanking DNA segments, called LPSs (*L*anding *P*ad *S*equences). LPSs should first be prepared from target DNA by a regular cloning method. For example, the LPSs used in this study were *Sal*I, *Not*I, and *Mlu*I-linking clones, because the location of these restriction enzyme sites in the insert of p185A21 was suitable with respect to appropriate distances (25–50 kb, shown at the top of Fig. 2). They were isolated from a *Bam*HI DNA library of p185A21 constructed in the *Bam*HI site of *E. coli* vector pCISP303B (Fig. 1) and listed in Table I. The two LPSs cloned on both sides of the [cI] of pCISP vector are collectively called the LPA (*LPS Array*). The LPA defines the target DNA to be cloned in the BGM vector. The LPA was integrated into the cloning site of BGM vector by transformation of *B. subtilis*. DNA uptake by *B. subtilis* competent cells has unique features (*15*) that are absent in *E. coli*. Briefly, competent-related genes of *B. subtilis* are expressed at the late stage of cell growth, and their products are assembled in cell membrane. The transformation

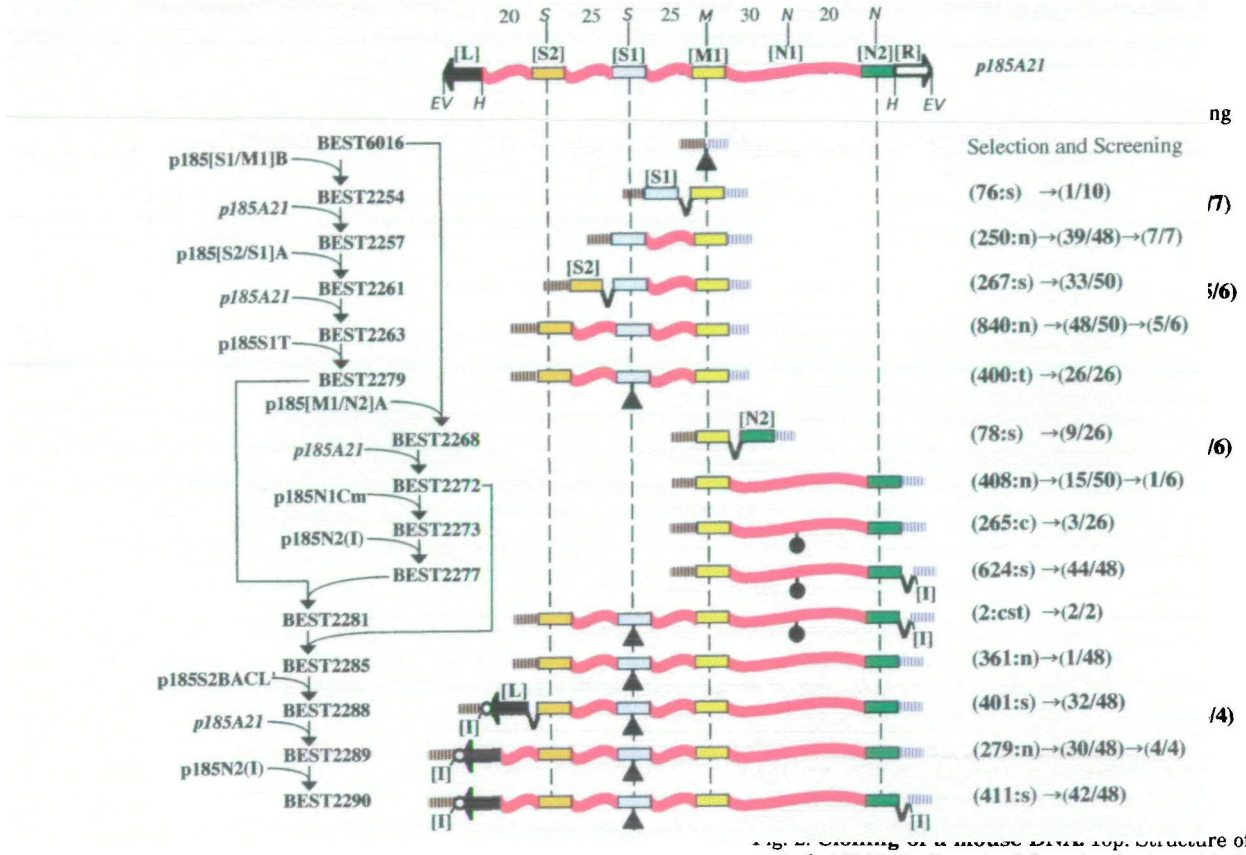

Fig. 2. **Cloning of a mouse DNA** [top]. Structure of the 120-kb insert of p185A21 in linearized form, beginning with the *Eco*RV site of the BAC vector pBAC108L. BAC vector is represented by closed arrows [L] and open arrows [R]. *cat* marker is carried in [I]. *S*, *M*, and *N* indicate *Sal*I, *Mlu*I, and *Not*I with distances in kb. *EV* and *H* stands for *Eco*RV and *Hin*dIII. Left: Transformation of BEST strains by [LPA] plasmid or p185A21. Center: Insert structure in the cloning locus of BGM vectors. Details of the top 5 strains, from BEST6016 to BEST2263, are presented in Fig. 1. Isolation of BEST2281 and BEST2285 is described in detail in Fig. 3. Antibiotic markers *erm*, *tet*, and *cat* are represented by ○, ▲, and ●, respectively. Right: Number of transformants selected by spectinomycin (s), neomycin (n), tetracycline (t), or chloramphenicol (c) in the first column. Results of the screening are indicated in the second column. Numbers of positive clones after Southern analysis are indicated in the third column.

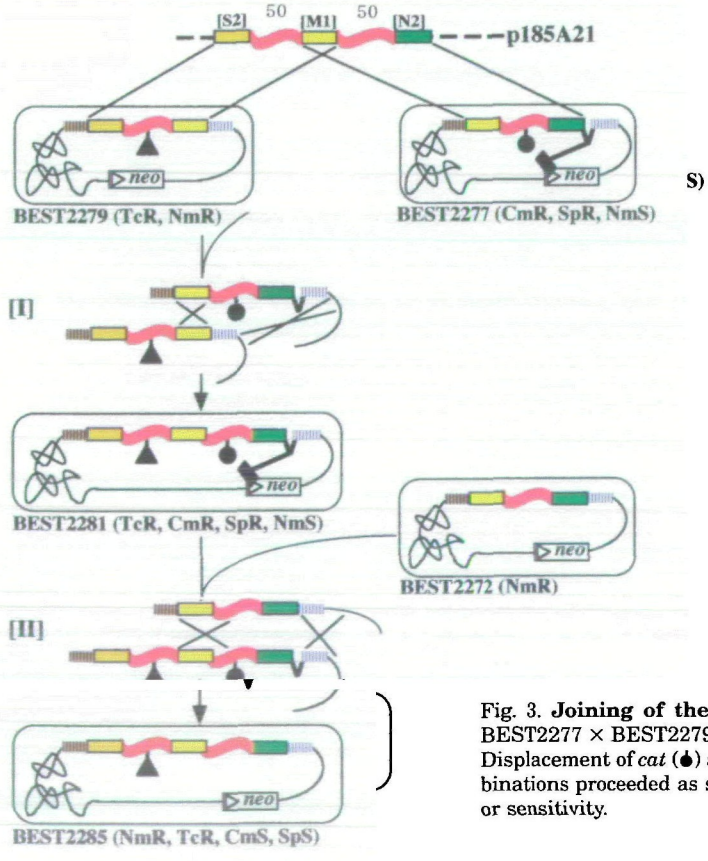

Fig. 3. **Joining of the separately cloned inserts.** [I]: Two recombinations in the cross of BEST2277 × BEST2279, one in the [M1] and the other the BGM vector beyond the pBR322. [II]: Displacement of *cat* (●) and [cI] (V) in the cross of BEST2272 × BEST2281. Two expected recombinations proceeded as shown by X. R or S after the antibiotic abbreviation indicates resistance or sensitivity.

machinery non-specifically binds and cleaves DNA. One strand from the cleaved site is actively transferred into the cytoplasm, yielding linearized recombinogenic DNA as shown in [A]–[D] of Fig. 1. Plasmid DNAs prepared in *E. coli*, normally covalently closed circular DNA form, are linearized when being taken up by *B. subtilis* competent cells. Thus, undigested plasmid DNA is used to transform *B. subtilis* in this study.

To integrate the approximately 25-kb segment between the LPSs [S1] and [M1], an LPA ([S1]-[cI]-[M1]) was constructed in plasmid pCISP303B as described in Table I. The LPA of p185[S1/M1]B was transferred into the BEST6016 genome as indicated in structure [A] of Fig. 1. This transfer was mediated by homologous recombination between pBR322 sequences in the genome and the pCISP vector, which is essentially identical to pBR322. BEST2254 was obtained through the screening of 76 colonies on spectinomycin, and then 10 colonies for neomycin sensitivity. The remaining 9 colonies were regarded as the result of Campbell-type integration by a single cross-over (*12*). This selection process is summarized in the right of Fig. 2 as (76:s)→(1/10). Repression of the Pr promoter by the *cI* gene product in BEST2254 is illustrated by a hammer in Fig. 1. p185A21 DNA added to competent BEST2254 caused homologous recombinations with the [S1] and [M1] of the BEST2254 genome. As a result of the intermediate structure [B] of Fig. 1, contiguous DNA from [S1] to [M1] displaced the [cI], and the *neo* gene was derepressed accordingly. The expected recombinant was selected on neomycin. BEST2257 having the expected 25-kb mouse DNA was obtained through selection of 250 neomycin-resistant colonies, followed by the screening of 48 giving 39 spectinomycin sensitives, and 7 strains had the correct DNA out of 7 examined by Southern hybridization using p185A21 as a probe. These steps are indicated in the right of Fig. 2 as (250:n)→(39/48)→(7/7). False positives were observed as reported in Ref. *8*, and in the steps outlined below. One group of neomycin-resistant false positives appeared by spontaneous base change in the *cI* gene (Tsuge and Itaya, unpublished). This group can be discriminated by the presence of the spectinomycin resistance gene. We noticed that clones showing expected antibiotic resistance appeared in which, however, DNA regions covering the *cI* and the spectinomycin resistance gene were merely deleted. These deletions, the formation mechanism of which remains undetermined, were detected in the screening of genome structure.

Similarly, to integrate the region between [M1] and [N2], an LPA ([M1]-[cI]-[N2]) was constructed in pCISP303B as described in Table I. Neomycin-sensitive BEST2268 was obtained upon transfer of LPA of p185[M1/N2]A to BEST6016. Recombinant BEST2272 that integrated 50-kb mouse DNA from [M1] to [N2] was obtained through a similar process to that described for the isolation of BEST2257.

2.2 Elongation to the Next LPS—To elongate the insert of BEST2257 to [S2], an LPA ([S2]-[cI]-[S1]) was constructed in pCISP303B as described in Table I. The LPA of p185[S2/S1]A was transferred to BEST2257, so as to add the segment [S2]-[cI] to the left end of the [S1] of BEST2257. This addition was achieved by two homologous recombinations in [S1] and a pBR322 sequence (brown barcode), shown by X in intermediate structure [C] of Fig. 1. BEST2261 was isolated as 1 of 33 neomycin-sensitives from 50 of the 267 spectinomycin-resistant colonies examined. In competent BEST2261, p185A21 DNA recombined within [S2] and the region left of [S1] indicated by an X in intermediate structure [D] of Fig. 1. BEST2263, obtained through a selection procedure shown in the right of Fig. 2, had the expected 50-kb contiguous DNA.

2.3 Joining of the Two Separately Cloned Inserts—The 50-kb mouse DNA from [S2] to [M1] was cloned in BEST2263, while the 50-kb from [M1] to [N2] was cloned in BEST2272. Because [M1] resides in both strains, we joined the two inserts in one BGM vector. To facilitate this selection, the insert of BEST2263 was labeled with the tetracycline resistance gene (*tet*), and that of BEST2272 with the chloramphenicol resistance gene (*cat*) and spectinomycin resistance gene (*spc*). The labeling was done using plasmid p185S1T, p185N1Cm, and p185N2(I), respectively. The construction of these plasmids is described in Table I. The structures of BEST2279 (*tet* labeled) and BEST2277 (*cat* and *spc* labeled) are indicated in Figs. 2 and 3. A joining reaction was initiated by adding liquid genomic DNA prepared from BEST2277 to competent BEST2279 at a concentration causing saturation (3 μg/ml). Two recombinants were selected by the use of three antibiotics. Having the same 100-kb contiguous DNA from [S2] to [N2], the two probably underwent the homologous recombination indicated in intermediate structure [I] of Fig. 3. One representative BEST2281 was used in the next step.

To use the *cI*-mediated selection system in the next elongation, the [cI] remaining in BEST2281 must be eliminated. This was achieved by replacing it with the non-labeled insert of BEST2272. Upon crossing BEST2272 with BEST2281, BEST2285 was obtained in which the expected homologous recombination (X) occurred outside the [cI] and *cat* as described in [II] of Fig. 3.

2.4 Completion of the Transfer of 120-kb Mouse DNA—An appropriate LPS was not obtained at the left end of the p185A21 insert. Thus, we used the BAC vector as LPS to extend the remaining sequence of the 20 kb left of [S2]. The BAC vector indicated as [L] of Fig. 2 was prepared from pBR::BAC101 by *Hin*dIII digestion and used to construct an LPA ([L]-[cI]-[S2]) *via* p185S210 in pCISP311B (Table I). The LPA of p185S2BACL′ was added to the left end of BEST2285 by a process similar to that used to obtain BEST2261 from BEST2257. BEST2288, an expected recombinant, integrated the erythromycin resistance gene and the I-*Ppo*I site closely linked to the *cI* cassette, because p185S2BACL′ was constructed using pCISP311B. pCISP311B is illustrated in Fig. 1. After transformation of BEST2288 with p185A21 DNA, BEST2289 was obtained by screening. The transfer of the 120-kb mouse DNA of p185A21 into the BGM vector was confirmed by Southern hybridization as shown in Fig. 4.

2.5 Recovery of the Insert—To recover the cloned 120-kb insert, the use of a 22-base sequence recognized by I-*Ppo*I nuclease was examined. As BEST2289 already had an I-*Ppo*I site at the left end, another I-*Ppo*I site was introduced by p185N2(I) at the right end. p185N2(I) has an I-*Ppo*I site derived from pCISP310B that was used to construct the plasmid (Table I). Spectinomycin-resistant BEST2290 was obtained upon transformation of BEST2289 with p185N2(I). I-*Ppo*I digestion of the BEST2290 genome produced an expected fragment that was resolved by CHEF gel electrophoresis as shown in Fig. 5.

To estimate the stability of mouse DNA in the BGM vec-

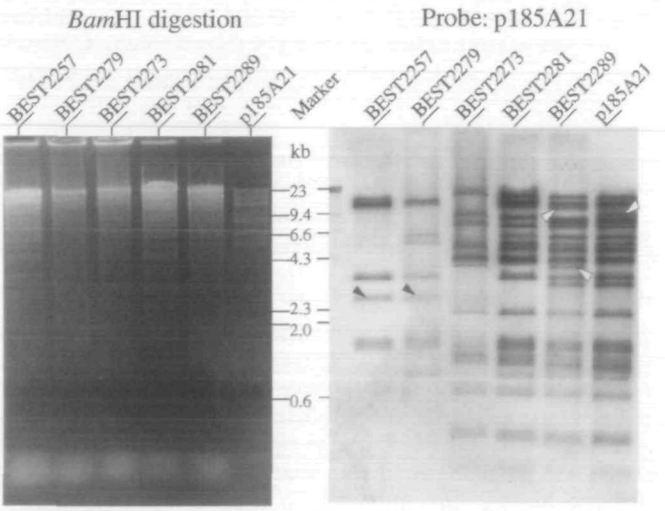

Fig. 4. **Progressive elongation of mouse DNA in the BGM vector.** Complete *Bam*HI digests of BGM clones are hybridized with p185A21 as a probe. Gel picture is shown at left. Comparison of Southern bands of the complete clone BEST2289 with the p185A21 lane revealed different bands. They all are indicated by (◺). BAC vector only appeared in the p185A21 lane. (▽) right junction and (◸) *tet* insert [S1] appeared only in BEST2289. LPS [M1] appeared as a 2.6-kb *Bam*HI fragment only in BEST2257 and BEST2279 (◣). This is argued in the text. Other identical bands were not investigated further.

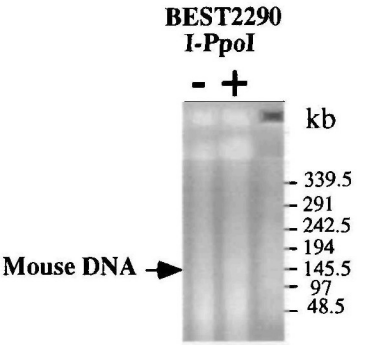

Fig. 5. **Cloned insert resolved by pulsed field gel electrophoresis.** BEST2290 produced a fragment of approximately 130 kb on I-*Ppo*I digestion. The band indicated by the arrow includes mouse genome DNA (120 kb), *erm*, (1.0 kb erythromycin resistance gene), [L] (3.8 kb), *tet* (1.9 kb), and [cI] (2.6 kb). Running conditions: 3 V cm^{-1}, 60 s pulse time and 36 h running time at 14°C. Smears observed below 100 kb were DNA degraded during preparation in agarose plugs. The amount varies from experiment to experiment.

tor, eight independent colonies of BEST2290 were grown in LB media without antibiotics, and the structure of each mouse insert was examined by Southern analysis. They were all identical to that of BEST2290 (data not shown). This observation, together with the high yield of expected strains in all steps employed in this study, indicated that the mouse DNA, up to 120 kb, faithfully replicated as part of the *B. subtilis* genome. It should be stressed that the stability is such that the cloned DNA can be engineered using the *B. subtilis* transformation/recombination system. The tetracycline resistance marker in the [S1] locus of BEST2290 is being replaced by a selection maker for mouse (M. Itaya, unpublished).

2.6 Discrimination of LPS Orientation by B. subtilis—The orientation of the LPSs other than [N2] had not been determined. We prepared four possible LPAs of [S1]-[cI]-[M1], with two [M1] orientations and two [S1] orientations. Expected recombinant BEST2257 was obtained only from BEST2254, one of the four strains examined having each LPA. Similarly, two LPAs of [S2]-[cI]-[S1] were prepared, given by two orientations for one [S1] orientation of BEST2257. Expected (elongated) recombinant BEST2263 was obtained only from BEST2261, one of the two strains examined. These results strongly indicated that integration definitely depends on tandem alignment of the two LPSs, and that the *B. subtilis* transformation/recombination system discriminates a priori the correct array. This selection by *B. subtilis* is advantageous, particularly when the relative orientation of the LPS can be hardly determined.

2.7 Repair Function by B. subtilis Transformation—*Bam*HI fragments of p185A21 were faithfully transferred into BEST2289 as indicated in Fig. 4, except for a 2.6-kb [M1] fragment subcloned in pCISP303B. The 2.6-kb *Bam*HI fragment observed in BEST2257 and BEST2279 appeared neither in BEST2273 or BEST2283 (joined recombinant) nor in p185A21. We speculate that the originally longer *Bam*HI fragment of M1 was shortened during the subcloning process in *E. coli*. The 2.6-kb fragment, however, was sufficient to conduct recombination. The shortened part, which probably resides at the right end of [M1], was repaired to give BEST2272, and the repaired region was later transmitted to the BGM clones.

DISCUSSION

The cloning and manipulation system using the *B. subtilis* genome was significantly improved from the prototype that used an *E. coli* prophage lambda DNA (48.5 kb) (*6*). Three recent findings contribute to this: (i) DNA larger than 50-kb can be transferred between *B. subtilis* genomes by a normal transformation procedure (*16*), (ii) LPS of 2 kb functioned to integrate a segment of 26.2 kb (*16*), and (iii) a positive selection system that allows rapid screening for recombinant *B. subtilis* integrating DNAs having no selection markers (*8*). Cloned DNAs are stably maintained as a single copy in the genome even when the host is cultured in non-selective medium. The most important characteristic of the BGM vector system is flexibility to engineer transferred DNAs. The flexibility is such that most DNA manipulations can be applied including insertion, deletion, replacement, and even addition of an adjacent region.

One fragile manipulation at present is an integration of LPA sequence as represented in [C] of Fig. 1. Alternative recombination between blue barcodes instead of that between [S1]s yields clones showing the same antibiotic resistance phenotype as those of BEST2261. There was no effective method to discriminate the two structures, but one candidate clone investigated was BEST2261. It was reported that frequency of double cross-over decreases as the interval distance increased in *B. subtilis* transformation (*9*, *16*). It is likely that double crossover flanking the short interval ([S2]-[cI], approximately 7.5 kb) preferentially occurred to that flanking the insert of BEST2257 (approximately 25 kb). We think this interpretation is supported by the similar observation that a candidate clone arbitrarily

selected in each LPA integration step indicated in Fig. 2 was BEST2277 and BEST2288.

The efficiency of integration seems to be affected by a number of factors such as grade of competent cell, amount and quality of DNA, length of LPS, and length of target segment, *etc.* The amount of DNA used is 3–5 μg/ml, which is the concentration causing saturation (*16*). Because cloning of long contiguous mouse genomic DNA using *B. subtilis* was unprecedented, we examined this starting with a 25-kb interval and finally succeeded in cloning of 50-kb DNA in single transformation. As regards the length of LPS, about 10% of the target DNA length has been empirically adopted. Whether of not this empirical rule is extended to larger-sized DNA than those of this study will be investigated.

BEST2289 showing an almost identical doubling time as BEST6016 had no obvious impairment of growth. Expression of the mouse genomic gene in BGM vector seems unlikely. The BGM vector should have broad applications in the assembly of DNA and recovery after appropriate engineering of the cloned insert. To elucidate maximum clonable size in the 4.2 Mb *B. subtilis* genome is critical for BGM technology. A *B. subtilis* strain whose genome size increased by several hundreds kb caused by duplication of part of the genome was reported (*17*). This indicated that the *B. subtilis* genome has plasticity to increase the size regardless of possible instability due to long tandem duplication. We are planning to assemble non-cognate DNA longer than 200 kb in the BGM vector. It should be addressed whether LPS can be obtained by a PCR-mediated DNA amplification method from species in which genome sequencing is completed. Construction of a transgenic mouse using BGM vectors is currently being investigated in order not only to show examples using the BGM vector but also to establish systems to detect subtle mutations possibly introduced during cloning and manipulations.

We thank Dr. H. Yanagawa for his helpful discussions and continuous encouragement.

REFERENCES

1. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
2. Burke, D.T., Carle, G.F., and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, **236**, 806–812
3. Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA*, **89**, 8794–8797
4. Pierce, J. and Sternberg, N.L. (1992) Using bacteriophage P1 system to clone high molecular weight genomic DNA in *Methods in Enzymology* Vol. 216, pp. 549–574, Academic Press, New York
5. Yang, X.W., Model. P., and Heintz, N. (1997). Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat. Biotechnol.* **15**, 859–865
6. Itaya, M. (1995) Toward a bacterial genome technology: Integration of the *Escherichia coli* prophage lambda genome into the *Bacillus subtilis* 168 chromosome. *Mol. Gen. Genet.* **248**, 9–16
7. Itaya, M. and Tanaka, T. (1991) Complete physical map of the *Bacillus subtilis* 168 chromosome constructed by a gene-directed mutagenesis method. *J. Mol. Biol.* **220**, 631–648
8. Itaya, M. (1999) Effective cloning of unmarked DNA fragments in the *Bacillus subtilis* 168 genome. *Biosci. Biotechnol. Biochem.* **63**, 602–604
9. Itaya, M. (1993) Integration of repeated sequences (pBR322) in the *Bacillus subtilis* 168 chromosome without affecting the genome structure. *Mol. Gen. Genet.* **241**, 287–297
10. Kunst, F., Ogasawara, N. Mozar, I. *et al.* (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256
11. Itaya, M. and Tanaka, T. (1997) Predicted and Unsuspected alterations of the genome structure of genetically defined *Bacillus subtilis* 168 strains. *Biosci. Biotechnol. Biochem.* **61**, 56–64
12. Itaya, M., Yamaguchi, I., Kobayashi, K., Endo, T., and Tanaka, T. (1990) The blasticidin S resistance gene (*bsr*) selectable in a single copy state in the *Bacillus subtilis* chromosome. *J. Biochem.* **107**, 799–801
13. Itaya, M. (1992) Construction of a novel tetracycline resistance gene cassette useful as a marker on the *Bacillus subtilis* chromosome. *Biosci. Biotechnol. Biochem.* **56**, 685–686
14. Toda, T., Tanaka, T., and Itaya, M. (1996) A method to invert DNA segments of the *Bacillus subtilis* 168 genome by recombination between two homologous sequences. *Biosci. Biotechnol. Biochem.* **60**, 773–778
15. Dubnau, D. (1999) DNA uptake in bacteria. *Ann. Rev. Microbiol.* **53**, 217–244
16. Itaya, M. (1999) Genetic transfer of large DNA inserts to designated loci of the *Bacillus subtilis* 168 genome. *J. Bacteriol.* **181**, 1045–1048
17. Anagnostopoulos, C. (1990) Genetic rearrangements in *Bacillus subtilis* in *The Bacterial Chromosome* (Drlica, K. and Riley, M., eds.) pp. 361–367, American Society for Microbiology, Washington, DC