



(86) Date de dépôt PCT/PCT Filing Date: 2007/06/25
(87) Date publication PCT/PCT Publication Date: 2007/12/27
(45) Date de délivrance/Issue Date: 2016/01/12
(85) Entrée phase nationale/National Entry: 2009/01/26
(86) N° demande PCT/PCT Application No.: JP 2007/062720
(87) N° publication PCT/PCT Publication No.: 2007/148819
(30) Priorité/Priority: 2006/06/23 (JPPCT/JP2006/312633)

(51) Cl.Int./Int.Cl. *C12N 15/82* (2006.01),
C12N 1/21 (2006.01), *A01H 5/00* (2006.01),
C12N 5/10 (2006.01)

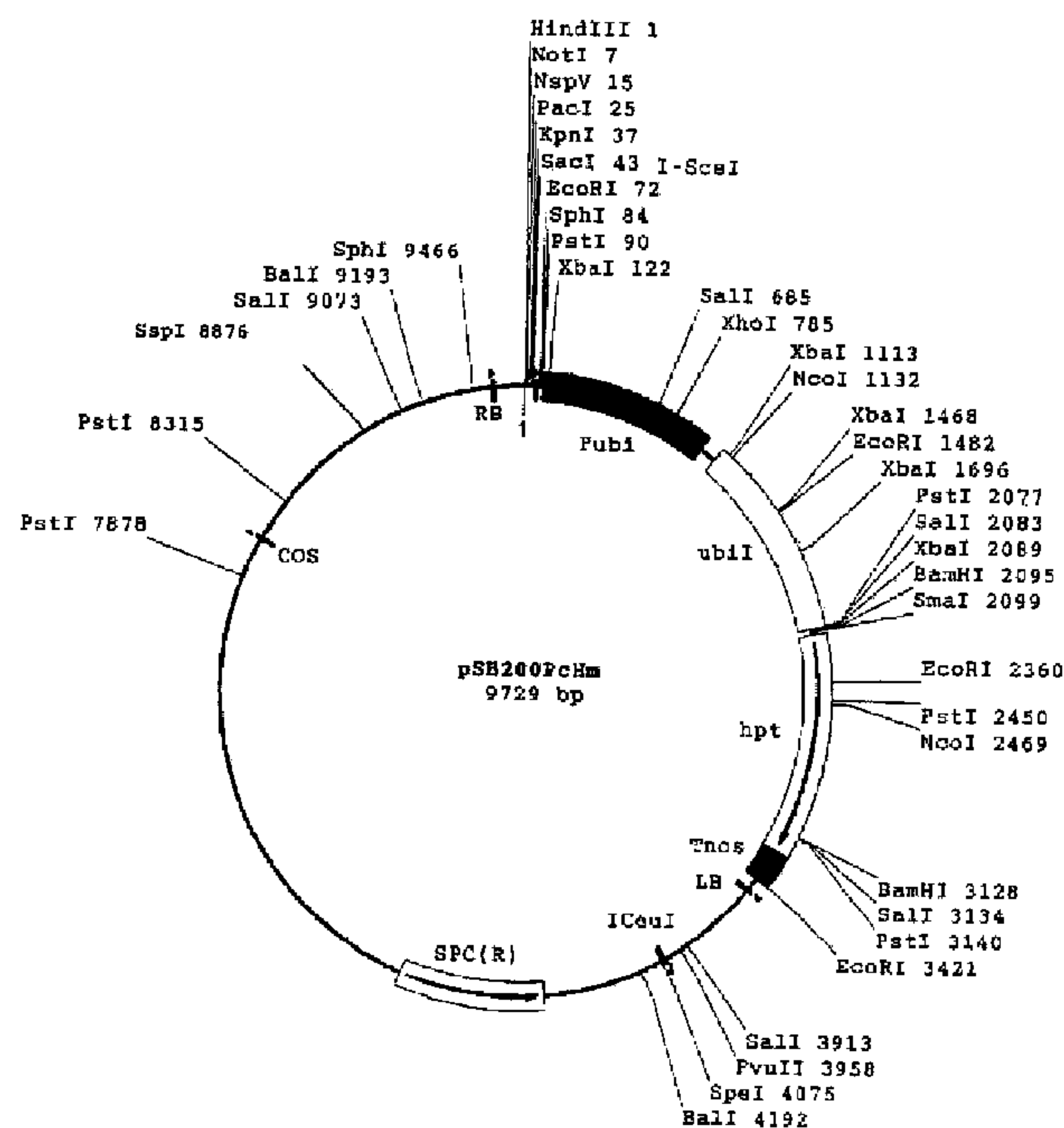
(72) Inventeurs/Inventors:
TAKAKURA, YOSHIMITSU, JP;
KOMARI, TOSHIHIKO, JP;
ISHIDA, YUJI, JP;
KOMORI, TOSHIYUKI, JP;
HIEI, YUKOH, JP;
MINE, TOSHIKI, JP;
IMAYAMA, TERUYUKI, JP

(73) Propriétaire/Owner:
JAPAN TOBACCO INC., JP

(74) Agent: ROBIC

(54) Titre : VECTEUR COSMIDE SERVANT A TRANSFORMER UN VEGETAL ET SON UTILISATION

(54) Title: COSMID VECTOR FOR TRANSFORMING PLANT AND USE THEREOF



(57) Abrégé/Abstract:

Disclosed is a novel vector for plant transformation. The vector is a cosmid vector having the entire length of 15 kb or shorter and also having the following properties: 1) the vector contains a replication initiation point for IncP plasmid but does not contain a replication initiation point for other plasmid; 2) the vector contains trfA1 gene for IncP plasmid; 3) the vector contains oriT for IncP plasmid; 4) the vector contains incC1 gene for IncP plasmid; 5) the vector contains a cos site for a lambda phage and the cos site is located outside of T-DNA; 6) the vector contains a drug-resistant gene which can be expressed in Escherichia coli and a bacterium belonging to the genus Agrobacterium; 7) the vector contains a right T-DNA-border sequence for a bacterium belonging to the genus Agrobacterium; 8) the vector contains a left T-DNA-border sequence for a bacterium belonging to the genus Agrobacterium; 9) the vector contains a selective marker gene for plant transformation, wherein the selective marker gene is located between the sequences mentioned in 7) and 8) and can be expressed in a plant; and 10) the vector contains a restriction enzyme recognition site for cloning of a foreign gene, wherein the restriction enzyme recognition site is located between the sequences mentioned in 7) and 8).

(12) 特許協力条約に基づいて公開された国際出願

訂正版

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2007 年12 月27 日 (27.12.2007)

PCT

(10) 国際公開番号
WO 2007/148819 A1

(51) 国際特許分類:

C12N 15/82 (2006.01)

C12N 1/21 (2006.01)

A01H 5/00 (2006.01)

C12N 5/10 (2006.01)

(21) 国際出願番号:

PCT/JP2007/062720

(22) 国際出願日:

2007 年6 月25 日 (25.06.2007)

(25) 国際出願の言語:

日本語

(26) 国際公開の言語:

日本語

(30) 優先権データ:

PCT/JP2006/312633

2006 年6 月23 日 (23.06.2006) JP

(71) 出願人 (米国を除く全ての指定国について): 日本たばこ産業株式会社 (JAPAN TOBACCO INC.) [JP/JP]; 〒1058422 東京都港区虎ノ門2丁目2番1号 Tokyo (JP).

(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 高倉 由光 (TAKAKURA, Yoshimitsu) [JP/JP]; 〒4380802 静岡県磐田市東原700番地 日本たばこ産業株式会社 植物イノベーションセンター内 Shizuoka (JP). 小鞠 敏彦 (KOMARI, Toshihiko) [JP/JP]; 〒4380802 静岡県磐田市東原700番地 日本たばこ産業株式会社 植物イノベーションセンター内 Shizuoka (JP). 石田 祐二 (ISHIDA, Yuji) [JP/JP]; 〒4380802 静岡県磐田市東原700番地 日本たばこ産業株式会社 植物イノベーションセンター内 Shizuoka (JP). 小森 俊之 (KOMORI, Toshiyuki) [JP/JP]; 〒4380802 静岡県

磐田市東原700番地 日本たばこ産業株式会社 植物イノベーションセンター内 Shizuoka (JP). 樋江井 祐弘 (HIEI, Yukoh) [JP/JP]; 〒4380802 静岡県磐田市東原700番地 日本たばこ産業株式会社 植物イノベーションセンター内 Shizuoka (JP). 峯 利喜 (MINE, Toshiki) [JP/JP]; 〒4380802 静岡県磐田市東原700番地 日本たばこ産業株式会社 植物イノベーションセンター内 Shizuoka (JP). 今山 輝之 (IMAYAMA, Teruyuki) [JP/JP]; 〒4380802 静岡県磐田市東原700番地 日本たばこ産業株式会社 植物イノベーションセンター内 Shizuoka (JP).

(74) 代理人: 社本 一夫, 外(SHAMOTO, ICHIO et al.); 〒1000004 東京都千代田区大手町二丁目2番1号 新大手町ビル206区 ユアサハラ法律特許事務所 Tokyo (JP).

(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD,

[続葉有]

(54) Title: COSMID VECTOR FOR PLANT TRANSFORMATION AND USE THEREOF

(54) 発明の名称: 植物形質転換用コスミドベクター及びその利用法

(57) Abstract: Disclosed is a novel vector for plant transformation. The vector is a cosmid vector having the entire length of 15 kb or shorter and also having the following properties: 1) the vector contains a replication initiation point for IncP plasmid but does not contain a replication initiation point for other plasmid; 2) the vector contains trfA1 gene for IncP plasmid; 3) the vector contains oriT for IncP plasmid; 4) the vector contains incC1 gene for IncP plasmid; 5) the vector contains a cos site for a lambda phage and the cos site is located outside of T-DNA; 6) the vector contains a drug-resistant gene which can be expressed in Escherichia coli and a bacterium belonging to the genus Agrobacterium; 7) the vector contains a right T-DNA-border sequence for a bacterium belonging to the genus Agrobacterium; 8) the vector contains a left T-DNA-border sequence for a bacterium belonging to the genus Agrobacterium; 9) the vector contains a selective marker gene for plant transformation, wherein the selective marker gene is located between the sequences mentioned in 7) and 8) and can be expressed in a plant; and 10) the vector contains a restriction enzyme recognition site for cloning of a foreign gene, wherein the restriction enzyme recognition site is located between the sequences mentioned in 7) and 8).

(57) 要約: 本発明は、植物形質転換用の新規ベクターを提供することを目的とする。本発明のベクターは、下記の特徴を有する全長15kb以内のコスミドベクター。1) IncPプラスミドの複製開始点を含み、他種のプラスミドの複製開始点を含まない; 2) IncPプラスミドのtrfA1遺伝子を含む; 3) IncPプラスミドのoriTを含む; 4) IncPプラスミドのincC1遺伝子を含む; 5) ラムダファージのcos部位を含み、かつ該cos部位はT-DNAの外側に位置する; 6) 大腸菌ならびにアグロバクテリウム属細菌において発現する薬剤耐性遺伝子を含む; 7) アグロバクテリウム属細菌のT-DNAの右ボーダー配列を含む; 8) アグロバクテリウム属細菌のT-DNAの左ボーダー配列を含む; 9) 7)と8)の間に配置される、植物において発現する、植物形質転換用選抜マーカー遺伝子を含む; そして10) 7)と8)の間に配置される、外来遺伝子をクローニングするための制限酵素認識部位を含む。

WO 2007/148819 A1

WO 2007/148819 A1



SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(48) この訂正版の公開日: 2008 年12 月11 日

(15) 訂正情報:
2008 年12 月11 日の更新情報 (Notice) を参照

添付公開書類:

- 国際調査報告書
- 電子形式により別個に公開された明細書の配列表部分、請求に基づき国際事務局から入手可能

- 1 -

SPECIFICATION

COSMID VECTOR FOR TRANSFORMING PLANT AND USE THEREOF

TECHNICAL FIELD

[0001] The present invention relates to novel cosmid vectors for transformaing plant and use thereof.

BACKGROUND ART

[0002] Various vectors have been previously developed for the purpose of plant transformation.

[0003] Recently, the entire genome sequences of *Arabidopsis thaliana* and rice (*Oryza sativa*) were elucidated, which moved the focus of plant genome studies from the accumulation of nucleotide sequence information to the elucidation of gene functions. For the elucidation of gene functions, experiments are absolutely necessary in which cloned DNA is transferred into a plant to analyze changes in the phenotype. If large DNA could be transferred in this operation, the study efficiency would be dramatically improved.

[0004] Thus, a number of vectors intended to transfer large DNA fragments into plants were developed. As typical examples, cosmid vectors for plant transformation were prepared, such as pOCA18 (Olszewski et al., 1988, *Nucleic Acids Res.* 16: 10765-10782) and pLZO3 (Lazo et al., 1991, *Bio/Technology* 9: 963-967). The use of a cosmid has the advantage that a lambda phage packaging reaction can be used, which allows easy cloning of relatively large genomic fragments (Sambrook J. and Russell D.W. 2001. *Molecular Cloning, A Laboratory Manual*, 3rd edn. Cold Spring Harbor

- 2 -

Laboratory Press, Cold Spring Harbor, New York, USA.). In cloning with a cosmid vector and a packaging reaction, the total size of the vector and the insert fragment is 40 kb-50 kb so that the size of the insert fragment is restricted within a certain range by the size of the vector and the sizes of the vector and the insert fragment inversely correlate with each other.

[0005] Vectors such as pOCA18 and pLZO3 contain elements for plant transformation such as T-DNA border sequences and a selectable marker (kanamycin resistance gene) in pRK290 (Ditta et al., 1980, Proc. Natl. Acad. Sci. USA 77: 7347-7351) which is a typical vector having an origin of replication (oriV) of an IncP plasmid that is functional in both E.coli and Agrobacterium. These vectors per se had a size of 24.3-30.1 kb, and therefore, the size of DNA that can be cloned using a packaging reaction was about 20 kb (pOCA18), or about 13-22 kb (pLZO3) on average. These vectors have an origin of replication (oriV) of an IncP plasmid, but other vectors such as pCIT103 and pCIT104 (Ma et al. 1992 Gene 117: 161-167) have an origin of replication from ColE1 in addition to an origin of replication (oriV) of an IncP plasmid. On the other hand, pC22 (Simoens et al. 1986 Nucleic Acids Res 14: 8073-8090) is a vector having an origin of replication from ColE1 and an origin of replication from an Ri plasmid. Other cosmid vectors capable of plant transformation include pMON565 (Klee et al. 1987 Mol Gen Genet 210: 282-287) and pCLD04541 (Bent et al. 1994 Science 265: 1856-1860), but they are not suitable for cloning DNA fragments of 25 kb or more because their own

- 3 -

sizes are 24 kb and 29 kb, respectively. Other examples such as pE4cos (16 kb, Klee et al. 1987 Mol Gen Genet 210: 282-287), pMON565, pLZ03, pOCA18, pCLD04541, pC22 and the like had a structure containing a cos site within the T-DNA.

[0006] Subsequently, the BIBAC vector (binary bacterial artificial chromosome, Hamilton U.S. Patent No. 5733744, Hamilton et al., 1996, Proc. Natl. Acad. Sci. USA 93:9975-9979, Hamilton, 1997, Gene 200:107-116) was developed, which is capable of cloning DNA fragments of up to about 150 kb and transferring them into plants. This vector is based on a BAC vector capable of carrying large DNA fragments and further contains elements for plant transformation such as T-DNA border sequences and a selectable marker as well as an origin of replication for Agrobacterium. The TAC vector (transformation-competent bacterial artificial chromosome) pYLTAC7 (Liu et al., 1999, Proc. Natl. Acad. Sci. USA 96: 6535-6540.) was also developed, which is capable of cloning DNA of up to about 80 kb and transferring it into plants. This vector is based on a high-capacity PAC vector (P1-derived artificial chromosome) using the replication mechanism of P1 phage and contains elements for plant transformation such as T-DNA border sequences and a selectable marker as well as an origin of replication for Agrobacterium. These vectors contain an origin of replication of (ori) from a plasmid existing as a single copy per cell in E.coli and Agrobacterium for the purpose of stably maintaining a large foreign gene. That is, they use an F factor ori (BIBAC) or a P1 phage ori (TAC) as ori for E.coli and an Ri ori from Agrobacterium

- 4 -

rhizogenes (both BIBAC and TAC) as ori for Agrobacterium. However, the use of an origin of replication from a single-copy plasmid is not necessarily essential, and vectors having an origin of replication (oriV) of an IncP plasmid known to exist as a few copies per cell such as pSLJ1711 and pCLD04541 were reported to be capable of stably maintaining plant genomic DNA fragments of more than 100 kb in size (Tao and Zhang (1998) Nucleic Acids Res 26: 4901-4909). In addition, pBIGRZ was also reported, which contains Ri ori in the versatile binary plasmid vector pBI121 (JPA Hei-10-155485).

[0007] Such vectors can be used to clone large DNA fragments far exceeding 50 kb, but involve complicated cloning operations. Cloning of large DNA requires skilled techniques and a considerable amount of time and labor. Transformation with BIBAC requires special Agrobacterium cells overexpressing virG or the like and results in a much lower transformation efficiency (the number of selected calli/inoculated leaf section) for fragments of 150 kb as compared with those of normal small vectors (Hamiltin et al., 1996, Proc Natl Acad Sci USA 93: 9975-9979, Shibata and Liu, 2000, Trend Plant Sci 5: 354-357). Thus, transformation of large fragments into plants with BIBAC or TAC is limited to a few specific examples of large fragments (e.g., Hamiltin et al., 1996, Proc Natl Acad Sci USA 93: 9975-9979, Liu et al., 1999, Proc Natl Acad Sci USA 96: 6535-6540, Lin et al., 2003, Proc Natl Acad Sci USA 100: 5962-5967, Nakano et al., 2005, Mol Gen Genomics 273: 123-129).

[0008] As described above, pCLD04541 is a cosmid of 29 kb

- 5 -

in size, and therefore, the size of DNA fragments that can be cloned using a lambda phage packaging reaction is 10-20 kb. If cloning of larger DNA fragments is intended, a packaging reaction cannot be used as described above, and thus complicated cloning operations and a considerable amount of time and labor are required.

[0009] Recently, genetic markers based on DNA sequence polymorphisms or so-called DNA markers are used more and more frequently with the advance in genome studies of higher plants. Many attempts have been made to clone unknown genes of higher plants known only by their phenotypes on the basis of genetic map information using DNA markers, i.e., so-called map-based cloning. Generally, the basic protocol of map-based cloning is as follows.

[0010] 1. Examine a relatively small segregating population with a set of DNA markers widely used for rough mapping of a candidate region on a chromosome.

[0011] 2. Screen a large segregating population with a set of DNA markers newly designed for the particular region of the genome to narrow down the candidate region.

[0012] 3. Determine the nucleotide sequence of the genetic region and guess a candidate gene.

[0013] 4. Transfer a DNA fragment containing the candidate gene into a plant and determine the effect/function of the gene on the basis of the phenotype.

[0014] Many previous successful cases often involve narrowing down the genetic region to about 1-3 genes in step 3 and transferring several DNA fragments of several kilobases or

- 6 -

less in step 4. However, it is not always easy to narrow down the genetic region. For example, it is often impossible to narrow down the genetic region to 150 kb or less in chromosomal regions near centromeres because of the low frequency of genetic recombination upon cross-hybridization. Even cases where narrowing down is possible often require repeating the operation of step 2 and therefore enormous amounts of time. Even if narrowing down to about 50 kb were possible, it would be very difficult to guess a candidate gene without strong information linking the phenotype to the gene sequence in step 3.

[0015] Thus, map-based cloning is relatively easy until the step of defining a candidate region including one to a few DNA fragments cloned by a BAC vector by narrowing down to some extent (to 50 kb to several hundreds of kilobases), but it is often technically difficult to further pursue the analysis to practically identify a gene, and even if it is possible, enormous amounts of labor and time are often required.

References:

Patent Publication No.1: U.S. Patent No. 5733744

Patent Publication No.2: Japanese Patent Laid-open
Publication No. H10-155485

Patent Publication No.3: WO2005/040374

Non-patent Publication No.1: Olszewski et al., 1988, Nucleic
Acids Res. 16: 10765-10782

Non-patent Publication No.2: Lazo et al., 1991,
Bio/Technology 9: 963-967

Non-patent Publication No.3: Ditta et al., 1980, Proc. Natl.

- 7 -

- Acad. Sci. USA 77: 7347-7351
- Non-patent Publication No.4: Ma et al. 1992 Gene 117: 161-167
- Non-patent Publication No.5: Simoens et al. 1986 Nucleic
Acids Res 14: 8073-8090
- Non-patent Publication No.6: Klee et al. 1987 Mol Gen Genet
210: 282-287
- Non-patent Publication No.7: Bent et al. 1994 Science 265:
1856-1860
- Non-patent Publication No.8: Hamilton et al., 1996, Proc.
Natl. Acad. Sci. USA 93:9975-9979,
- Non-patent Publication No.9: Hamilton, 1997, Gene 200:107-116
- Non-patent Publication No.10: Liu et al., 1999, Proc. Natl.
Acad. Sci. USA 96: 6535-6540
- Non-patent Publication No.11: Tao and Zhang, 1998, Nucleic
Acids Res 26: 4901-4909
- Non-patent Publication No.12: Shibata and Liu, 2000, Trend
Plant Sci 5: 354-357
- Non-patent Publication No.13: Lin et al., 2003, Proc Natl
Acad Sci USA 100: 5962-5967,
- Non-patent Publication No.14: Nakano et al., 2005, Mol Gen
Genomics 273: 123-129
- Non-patent Publication No.15: Pansegrau et al. (1994) J Mol
Biol 239: 623-663
- Non-patent Publication No.16: Knauf and Nester 1982 Plasmid
8: 45-54
- Non-patent Publication No.17: Komari et al. 1996 Plant J
10:165-174
- Non-patent Publication No.18: Zambryski et al. 1980 Science

- 8 -

209: 1385-1391

Non-patent Publication No.19: Schmidhauser and Helinski, J.
Bacteriol. 164:446-455, 1985

Non-patent Publication No.20: Winans et al. 1986 Proc. Natl.
Acad. Sci. USA 83: 8278-8282

Non-patent Publication No.21: Pazour et al. 1992 J.Bac
174:4169-4174

Non-patent Publication No.22: Ward et al. (1988) J Biol Chem
263: 5804-5814

Non-patent Publication No.23: Frame et al. 2002 Plant Physiol
129: 13-22

Non-patent Publication No.24: Hansen et al. 1994 ProNAS
91:7603-7607

Non-patent Publication No.25: Ishida et al. 1996 Nat
Biotechnol 14:745-50

Non-patent Publication No.26: Close et al. 1984 Plasmid 12:
111-118

Non-patent Publication No.27: Jin et al. 1987 J Bacteriol
169: 4417-4425

Non-patent Publication No.28: Wang et al. 2000 Gene 242: 105-
114

Non-patent Publication No.29: Okumura and Kado (1992) Mol Gen
Genet 235: 55-63

Non-patent Publication No.30: Christensen et al. 1992 Plant
Mol Biol 18: 675-689

Non-patent Publication No.31: Bilang et al. (1991) Gene 100:
247-250

Non-patent Publication No.32: Hirsch and Beringer 1984

- 9 -

Plasmid 12: 139-141

Non-patent Publication No.33: Konieczny and Ausubel 1993

Plant Journal 4: 403-410

Non-patent Publication No.34: Hiei et al. (1994) Plant J
6: 271-282

Non-patent Publication No.35: Ishida et al. (2003) Plant
Biotechnology 20:57-66

Non-patent Publication No.36: Hiei and Komari (2006)
Plant Cell, Tissue and Organ Culture 85: 271-283

10 Non-patent Publication No.37: Komori et al. (2004) Plant
J 37: 315-325

Non-patent Publication No.38: Kazama and Toriyama (2003)
FEBS lett 544: 99-102.

DISCLOSURE OF THE INVENTION

PROBLEMS TO BE SOLVED BY THE INVENTION

[0016] WO2005/040374, discloses a method for efficiently
selecting and preparing a number of genomic DNA fragments
capable of improving traits expressed in heterosis or
quantitative traits as cloned DNA fragments. We have
20 selected large genomic DNA fragments capable of
introducing agriculturally useful mutations by using the
method described in WO2005/040374. However, the success
rate of transferring clones carried in E.coli into
Agrobacterium was about 80%. Moreover, only about 60% of
Agrobacterium strains harboring clones was able to
transform plants. In view of this result, we examined
whether or not the efficiency of the method of
WO2005/040374 could be significantly improved by changing the

- 10 -

vector used. However, the efficiency of this method could not be improved by any vector ever known.

[0017] Thus, an object of the present invention is to provide a novel vector capable of improving the efficiency of selecting and cloning relatively large genomic DNA fragments, e.g., in the method described in WO2005/040374.

[0018] Another object of the present invention is to provide a vector preferably fulfilling all of the requirements below:

- it allows efficient cloning of DNA fragments of about 25-40 kb in size;
- it is stably maintained in E.coli and Agrobacterium cells ;
- it can be efficiently introduced into Agrobacterium;
- the copy number per cell in E.coli and Agrobacterium is 4-5; and
- it allows efficient transfer of only cloned DNA fragments of interest into plants, preferably monocotyledons.

[0019] Still another object of the present invention is to provide a gene transfer method for transferring a gene into a plant at a very high efficiency using such a vector.

[0020] Still another object of the present invention is to provide a method for rapidly narrowing down a gene region for completing map-based cloning with ease in a short time using such a vector.

[0021] Still another object of the present invention is to provide a plasmid capable of further improving the transformation efficiency by combining it with said vector.

- 11 -

MEANS FOR SOLVING THE PROBLEMS

[0022] Cosmid vectors

The cosmid vectors of the present invention are vectors having a full length of 15 kb or less satisfying all of the following criteria (hereinafter referred to as "pLC vectors"):

[0023] 1) they contain an origin of replication (oriV) of an IncP plasmid, but do not contain any origin of replication of other plasmid groups;

2) they contain the trfA1 gene of an IncP plasmid;

10 3) they contain an origin of conjugative transfer (oriT) of an IncP plasmid;

4) they contain the incC1 gene of an IncP plasmid;

5) they contain a cos site of lambda phage and the cos site is located outside the T-DNA;

6) they contain a drug resistance gene expressed in E.coli and a bacterium of the genus *Agrobacterium*;

7) they contain a T-DNA right border sequence of a bacterium of the genus *Agrobacterium*;

20 8) they contain a T-DNA left border sequence of a bacterium of the genus *Agrobacterium*;

9) they contain a selectable marker gene for plant transformation located between 7) and 8) and expressed in a plant; and

10) they contain restriction endonuclease recognition site(s) located between 7) and 8) for cloning a foreign gene.

- 11a -

The present invention also concerns a method for transforming a plant, comprising transforming the plant with a bacterium of the genus *Agrobacterium* harboring an expression vector containing a nucleic acid fragment of a plant inserted into the cosmid vector of the invention.

The present invention also concerns a map-based cloning method comprising the steps of:

1) partially or completely digesting BAC clones containing candidate genes responsible for a plant phenotype with a restriction endonuclease;

2) subcloning DNA fragments obtained in step 1) using the cosmid vector of any one of claims 1 to 5 to construct a library; and

3) individually transferring clones constituting the library into a plant to evaluate the phenotypes of transformed plants.

The present invention also concerns a method for transforming a plant, comprising transforming the plant with a bacterium of the genus *Agrobacterium* harboring the cosmid vector of the invention and a plasmid vector characterized in that:

it contains an origin of replication of an IncW plasmid, but does not contain any origin of replication of other plasmid groups;

it contains the repA gene necessary for the replication of an IncW plasmid;

it contains a drug resistance gene expressed in *E.coli* and a bacterium of the genus *Agrobacterium*; and

- 11b -

the *virG* gene of a bacterium of the genus *Agrobacterium*.

[0023.1] The present invention also concerns a method for transforming a plant, comprising transforming the plant with a bacterium of the genus *Agrobacterium* harboring the cosmid vector of the invention and a plasmid vector characterized in that:

1) said plasmid vector contains an origin of replication of an IncW plasmid, but does not contain any origin of replication of other plasmid groups;

2) said plasmid vector contains the *repA* gene necessary for the
10 replication of an IncW plasmid;

3) said plasmid vector contains a drug resistance gene expressed in *E.coli* and a bacterium of the genus *Agrobacterium*; and

4) said plasmid vector contains the *virG* gene of a bacterium of the genus *Agrobacterium*.

[0024] The vectors of the present invention are cosmid vector containing a cos site of lambda phage. This allows cloning of relatively large genomic fragments by a packaging

- 12 -

reaction (Sambrook J. and Russell D.W. 2001. Molecular Cloning, A Laboratory Manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.). In cloning using a packaging reaction of a cosmid vector, the total size of the vector and the insert fragment is around 40 kb-50 kb so that the size of the insert fragment is restricted within a certain range by the size of the vector. The vectors of the present invention have a full length of 15 kb or less, preferably 12-14 kb because they are intended to clone a DNA fragment of up to about 25-40 kb, preferably 30-40 kb.

[0025] 1) An origin of replication (oriV) of an IncP plasmid: The oriV is functional in both E.coli and Agrobacterium. The nucleotide sequence of the oriV of the present invention is not specifically limited so far as it has the function of oriV, i.e., the function of an origin of replication of an IncP plasmid.

[0026] The oriV has molecular biological properties described in detail in Pansegrau et al. (1994) J Mol Biol 239: 623-663, and it is defined as nucleotides 12200-12750 of the sequence of Genbank/EMBL Accession Number L27758 (full length 60099 bp). This corresponds to nucleotides 3451-4002 of SEQ ID NO: 1 (core sequence of oriV).

[0027] The oriV can be conventionally prepared from an IncP plasmid such as pVK102 (Knauf and Nester 1982 Plasmid 8: 45-54). For example, a 0.9 kb DNA (nucleotides 3345-4247 of SEQ ID NO: 1) amplified by PCR from pVK102 can be used as the oriV.

[0028] Alternatively, a nucleic acid containing a

- 13 -

nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 3451-4002, more preferably 3345-4247 of SEQ ID NO: 1 described above under stringent conditions and having the function of oriV can also be used. Alternatively, a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 3451-4002, more preferably 3345-4247 of SEQ ID NO: 1 described above and having the function of oriV can also be used.

[0029] It will be recognized by those skilled in the art that a shorter region in nucleotides 3451-4002 of SEQ ID NO: 1 may be selected as a sequence having a similar function. We investigated from various viewpoints the reason why the final transformation efficiency was only about 50% ($80\% \times 60\% = 48\%$) when the method described in WO2005/040374 was used, and concluded that this might be ascribable to the use of the cloning vector pSB200.

[0030] A replication origin of pSB200 is from ColE1, and plasmids having an origin of replication from ColE1 exist in a relatively high copy number, i.e., 30-40 copies per E.coli cell. Tao and Zhang (1998, Nucleic Acids Res 26: 4901-4909) assume that E.coli can stably maintain 1200-1500 kb of foreign DNA per cell. If 30-40 kb of DNA is cloned by pSB200, the total DNA amount including the vector size reaches 1200-2000 kb per cell, which may exceed the range assumed above. Another possible reason is that pSB200 is a plasmid that is not replicated alone in Agrobacterium. Thus, a vector having

- 14 -

an origin of replication (oriV) of an IncP plasmid called pSB1 is preliminarily introduced into Agrobacterium, and a cointegrate between pSB200 and pSB1 is prepared via homologous recombination between DNA sequences contained in both pSB200 and pSB1, thereby introducing pSB200 into Agrobacterium. It is undeniable that some adverse phenomenon could occur during such an operation to result in the failure in the transfer of pSB200.

[0031] If the copy number in E.coli and Agrobacterium is too low, however, the analysis of DNA or the like will be inefficient.

[0032] Based on the foregoing discussion, we prepared and tested vectors containing an origin of replication (oriV) of an IncP plasmid that is functional in both E.coli and Agrobacterium but not any origin of replication of other plasmid groups and existing in 4-5 copies in these bacteria. As a result, we found that the transformation efficiency is improved by using such vectors in plant transformation, specifically e.g., in the method described in WO2005/040374, and thus achieved the present invention.

[0033] 2) The trfA1 gene of an IncP plasmid: The trfA1 gene is important as a transacting replication factor of IncP plasmids and necessary for an oriV to perform its function. The nucleotide sequence of the trfA1 gene of the present invention is not specifically limited so far as it has the function of trfA1, i.e. the function of a transacting replication factor.

[0034] It has molecular biological properties described in

- 15 -

detail in Pansegrau et al. (1994) J Mol Biol 239: 623-663, and it is defined as nucleotides 16521-17669 of the sequence of Genbank/EMBL Accession Number L27758 (full length 60099 bp). This corresponds to nucleotides 6323-7471 of SEQ ID NO: 1 (core sequence of trfA1).

[0035] TrfA1 can be conventionally prepared from an IncP plasmid such as pVK102 (Knauf and Nester 1982 Plasmid 8: 45-54). For example, a 3.2 kb DNA fragment (nucleotides 5341-8507 of SEQ ID NO: 1) amplified by PCR from pVK102 can be used as the trfA1 gene.

[0036] Alternatively, a nucleic acid containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 6323-7471, more preferably 5341-8507 of SEQ ID NO: 1 described above under stringent conditions and having the function of the trfA1 gene can also be used. Alternatively, a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 6323-7471, more preferably 5341-8507 of SEQ ID NO: 1 described above and having the function of the trfA1 gene can also be used.

[0037] It will be recognized by those skilled in the art that a shorter region in nucleotides 6323-7471 of SEQ ID NO: 1 may be selected as a sequence having a similar function.

[0038] 3) An origin of conjugative transfer (oriT) of an IncP plasmid: oriT is an element responsible for conjugation (mating). One of the purposes of the vectors of the present invention is to perform large-scale and high-efficient

- 16 -

transformation. For that purpose, conjugation (mating) between E.coli and Agrobacterium is necessary, and oriT contributes to the conjugation (mating). The sequence of the oriT of the present invention is not specifically limited so far as it has the function of oriT, i.e. the function of an element responsible for conjugation (mating).

[0039] The oriT has molecular biological properties described in detail in Pansegrau et al. (1994) J Mol Biol 239: 623-663, and it is defined as nucleotides 51097-51463 of the sequence of Genbank/EMBL Accession Number L27758 (full length 60099 bp). The oriT can be conventionally prepared from an IncP plasmid such as pVK102 (Knauf and Nester 1982 Plasmid 8: 45-54). For example, a 0.8 kb DNA fragment (nucleotides 1-816 of SEQ ID NO: 1) amplified by PCR from pVK102 can be used as the oriT.

[0040] Alternatively, a nucleic acid containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 1-816 of SEQ ID NO: 1 described above under stringent conditions and having the function of oriT can also be used. Alternatively, a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 1-816 of SEQ ID NO: 1 described above and having the function of oriT can also be used.

[0041] It will be recognized by those skilled in the art that a shorter region in nucleotides 1-816 of SEQ ID NO: 1 may be selected as a sequence having a similar function.

- 17 -

[0042] 4) The *incC1* gene of an IncP plasmid: The *incC1* gene contributes to the stability of IncP plasmids. The nucleotide sequence of the *incC1* gene of the present invention is not specifically limited so far as it has the function of the *incC1* gene contributing the stability of IncP plasmids.

[0043] This gene has molecular biological properties described in detail in Pansegrau et al. (1994) J Mol Biol 239: 623-663, and it is defined as nucleotides 58260-59354 of the sequence of Genbank/EMBL Accession Number L27758 (full length 60099 bp). This corresponds to nucleotides 1179-2273 of SEQ ID NO: 1 (core sequence of the *incC1* gene).

[0044] *IncC1* can be conventionally prepared from an IncP plasmid such as pVK102 (Knauf and Nester 1982 Plasmid 8: 45-54). For example, a 2.1 kb DNA fragment (nucleotides 817-2935 of SEQ ID NO: 1) amplified by PCR from pVK102 can be used as the *incC1* gene.

[0045] Alternatively, a nucleic acid containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 1179-2273, more preferably 817-2935 of SEQ ID NO: 1 described above under stringent conditions and having the function of the *incC1* gene can also be used. Alternatively, a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 1179-2273, more preferably 817-2935 of SEQ ID NO: 1 described above and having the function of the *incC1* gene can also be used.

[0046] It will be recognized by those skilled in the art

- 18 -

that a shorter region in nucleotides 1179-2273 of SEQ ID NO: 1 may be selected as a sequence having a similar function.

[0047] 5) A cos site of lambda phage: The vectors of the present invention contain a cos site of lambda phage to utilize the packaging reaction of cosmid vectors. The nucleotide sequence of the cos site of lambda phage of the present invention is not specifically limited so far as it has the function of a cos site of lambda phage, i.e. the function contributing to the packaging reaction of cosmid vectors.

[0048] The cos site of lambda phage has molecular biological properties described in detail in Sambrook J. and Russell D.W. (2001), and it has the sequence 5'-aggtcgccgccc-3' (SEQ ID NO: 9) (the core sequence of a cos site of lambda phage). The cos can be conventionally prepared from a plasmid such as pSB11 (Komari et al. 1996 Plant J 10:165-174). For example, a 0.4 kb DNA fragment (nucleotides 2936-3344 of SEQ ID NO: 1) amplified by PCR from pSB11 can be used.

[0049] Alternatively, a nucleic acid containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of SEQ ID NO: 9 described above, more preferably the nucleotide sequence of nucleotides 2936-3344 of SEQ ID NO: 1 under stringent conditions and having the function of a cos site of lambda phage can also be used. Alternatively, a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of SEQ ID NO: 9 described above, more preferably the nucleotide sequence of nucleotides 2936-3344 of SEQ ID NO: 1 and having the function

- 19 -

of a cos site of lambda phage can also be used.

[0050] The cos site should be located outside the T-DNA because undesired DNA will be introduced into plants if the cos site is located inside the T-DNA.

[0051] 6) The drug resistance gene expressed in E.coli and a bacterium of the genus Agrobacterium is used as a selectable marker for transformation. This drug resistance gene confers e.g., antibiotic resistance or autotrophy, including, but not limited to, a kanamycin resistance gene, a spectinomycin resistance gene, an ampicillin resistance gene, a tetracycline resistance gene, a gentamycin resistance gene, a hygromycin resistance gene, etc.

[0052] 7), 8) T-DNA right border sequence (RB) and left border sequence (LB) of a bacterium of the genus Agrobacterium are essential for transformation (Zambryski et al. 1980 Science 209: 1385-1391), and a cloning site for a foreign gene is located between them. The nucleotide sequences of the RB and LB of the present invention are not specifically limited so far as they have the function of T-DNA right border sequence (RB) and left border sequence (LB) of a bacterium of the genus Agrobacterium. They can be each conventionally prepared from a plasmid such as pSB11 (Komari et al. 1996 Plant J 10:165-174). For example, nucleotides 13253-13277 and 3479-3503 of SEQ ID NO: 2 can be used, respectively.

[0053] Alternatively, nucleic acids containing nucleotide sequences hybridizing to complementary strands of the nucleotide sequences of nucleotides 13253-13277 and 3479-3503 of SEQ ID NO: 2 described above under stringent conditions and

- 20 -

having the functions of the RB and LB, respectively, can also be used. Alternatively, nucleic acids containing nucleotide sequences having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequences of nucleotides 13253-13277 and 3479-3503 of SEQ ID NO: 2 described above and having the functions of the RB and LB, respectively, can also be used.

[0054] It will be recognized by those skilled in the art that shorter regions in nucleotides 13253-13277 and 3479-3503 of SEQ ID NO: 2 may be selected as sequences having similar functions.

[0055] 9) A selectable marker gene for plant transformation expressed in a plant cell and located between 7) and 8) is included. The selectable marker gene for plant transformation is not specifically limited, and known selectable marker genes can be used. Preferably, it is any one of a hygromycin resistance gene, a phosphinotricin resistance gene, and a kanamycin resistance gene. For use in transformation of monocotyledons, a hygromycin resistance gene or a phosphinotricin resistance gene is preferred.

[0056] 10) Restriction endonuclease recognition site(s) located between 7) and 8) for cloning a foreign gene are included. The restriction endonuclease recognition sites for cloning a foreign gene are not specifically limited, and known restriction endonuclease recognition sites can be used, but the same recognition sites are desirably absent elsewhere on the vectors.

[0057] In the cosmid vector constructs of the present

- 21 -

invention, the order of all of the seven elements consisting of elements 1)-6) and a unit of 7)-10) is not limited. Moreover, the order of 9) and 10) located between 7) and 8) is not limited.

[0058] The cosmid vectors of the present invention preferably satisfy one or more of the following criteria A-G.

[0059] A. The nucleotide sequence of oriV in 1) comprises the following nucleotide sequence:

i) the nucleotide sequence of nucleotides 3451-4002, more preferably 3345-4247 of SEQ ID NO: 1;

ii) a nucleotide sequence containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 3451-4002, more preferably 3345-4247 of SEQ ID NO: 1 under stringent conditions and having the function of oriV; or

iii) a nucleotide sequence containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 3451-4002, more preferably 3345-4247 of SEQ ID NO: 1 and having the function of oriV.

B. The trfA1 gene in 2) comprises the following nucleotide sequence:

i) the nucleotide sequence of nucleotides 6323-7471, more preferably 5341-8507 of SEQ ID NO: 1;

ii) a nucleotide sequence containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 6323-7471, more preferably 5341-8507 of SEQ ID NO: 1 under stringent conditions and

- 22 -

having the function of the trfA1 gene;

iii) a nucleotide sequence containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 6323-7471, more preferably 5341-8507 of SEQ ID NO: 1 and having the function of the trfA1 gene.

C. The oriT in 3) comprises the following nucleotide sequence:

i) the nucleotide sequence of nucleotides 1-816 of SEQ ID NO: 1;

ii) a nucleotide sequence containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 1-816 of SEQ ID NO: 1 under stringent conditions and having the function of oriT;

iii) a nucleotide sequence containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 1-816 of SEQ ID NO: 1 and having the function of oriT.

D. The incC1 gene in 4) comprises the following nucleotide sequence:

i) the nucleotide sequence of nucleotides 1179-2273, more preferably 817-2935 of SEQ ID NO: 1;

ii) a nucleotide sequence containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 1179-2273, more preferably 817-2935 of SEQ ID NO: 1 under stringent conditions and having the function of the incC1 gene;

- 23 -

iii) a nucleotide sequence containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 1179-2273, more preferably 817-2935 of SEQ ID NO: 1 and having the function of the incC1 gene.

E. The cos site of lambda phage in 5) comprises the following nucleotide sequence:

i) the nucleotide sequence of SEQ ID NO: 9, more preferably the nucleotide sequence of nucleotides 2936-3344 of SEQ ID NO: 1;

ii) a nucleotide sequence containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of SEQ ID NO: 9, more preferably the nucleotide sequence of nucleotides 2936-3344 of SEQ ID NO: 1 under stringent conditions and having the function of a cos site of lambda phage;

iii) a nucleotide sequence containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of SEQ ID NO: 9, more preferably the nucleotide sequence of nucleotides 2936-3344 of SEQ ID NO: 1 and having the function of a cos site of lambda phage.

F. The T-DNA right border sequence (RB) of a bacterium of the genus Agrobacterium in 7) comprises the following nucleotide sequence:

i) the nucleotide sequence of nucleotides 13253-13277 of SEQ ID NO: 2;

ii) a nucleotide sequence containing a nucleotide

- 24 -

sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 13253-13277 of SEQ ID NO: 2 under stringent conditions and having the function of RB;

iii) a nucleotide sequence containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 13253-13277 of SEQ ID NO: 2 and having the function of RB.

G. The T-DNA left border sequence (LB) of a bacterium of the genus *Agrobacterium* in 8) comprises the following nucleotide sequence:

i) the nucleotide sequence of nucleotides 3479-3503 of SEQ ID NO: 2;

ii) a nucleotide sequence containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 3479-3503 of SEQ ID NO: 2 under stringent conditions and having the function of LB;

iii) a nucleotide sequence containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 3479-3503 of SEQ ID NO: 2 and having the function of LB.

By satisfying all of the criteria 1)-10) above for the cosmid vectors of the present invention, a vector fulfilling all of the requirements below can be prepared:

- it allows efficient cloning of DNA fragments of about 25-40 kb in size, preferably 30-40 kb;

- it is stably maintained in *E.coli* and *Agrobacterium*

- 25 -

cells ;

- it can be efficiently introduced into Agrobacterium;
- the copy number per cell in E.coli and Agrobacterium is 4-5; and

- it allows efficient transfer of only cloned DNA fragments of interest into plants, preferably monocotyledons.

[0060] However, the development of such a vector was not straightforward even after the requirements above had been defined. This is partially due to the very complex control mechanism of plasmid replication. Specifically, the most suitable vector backbone for the requirements above is a small plasmid (in the order of 12 kb to 15 kb) having an origin of replication (oriV) of an IncP plasmid. However, the backbone 60 kb IncP plasmid has many genes involved in the replication of the plasmid and partitioning during cell division, resulting in a very complex mechanism, though its entire nucleotide sequence has been determined (Pansegrau et al. J. Mol. Biol. 239:623-663, 1994). Thus, it is not easy to prepare a small vector having an origin of replication (oriV) of an IncP plasmid and stably maintained in bacteria. In fact, plasmids of various sizes derived from IncP plasmids have been studied, but small plasmids are generally unstable and widely differ in stability depending on the bacterial species (Schmidhauser and Helinski, J. Bacteriol. 164:446-455, 1985). pE4cos is an example of the plasmid which has lost stability in Agrobacterium by size reduction. The reasons for this have been discussed to a certain extent (Klee et al. 1987 Mol Gen Genet 210: 282-287), but it can be hardly said that

- 26 -

they have been clarified.

[0061] Schmidhauser and Helinski (J. Bacteriol. 164:446-455, 1985) say that "there is no universal set of genetic determinants in plasmid RK2 that accounts for stable maintenance in all gram-negative bacteria", indicating great difficulty in the preparation of a small and stable vector. The plasmid RK2 here (also often designated as pRK2) is one of typical IncP plasmids. The procedure for constructing such a vector often uses the step of cloning elements of a backbone plasmid using another vector. However, DNA fragments involved in the replication of bacterial plasmids or chromosomes are sometimes difficult to clone. If such a problem occurs, a means to solve it must be developed, which contributes to the difficulty in the construction of novel vectors.

[0062] Non-limitative examples of the cosmid vectors (pLC series) of the present invention are as follows.

[0063] i) pLC40 (SEQ ID NO: 2, Figure 6)

A binary cosmid vector having a full length of 13429 bp characterized in that:

[0064] 1) it contains an origin of replication (oriV) of an IncP plasmid, but does not contain any origin of replication of other plasmid groups;

2) it contains the trfA1 gene, 3) oriT, and 4) the incC1 gene of an IncP plasmid;

5) it contains a cos site of lambda phage and the cos site is located outside the T-DNA;

6) it contains the drug resistance gene nptIII (kanamycin resistance gene) expressed in E.coli and a bacterium of the

- 27 -

genus *Agrobacterium*;

7) it contains a T-DNA right border sequence of a bacterium of the genus *Agrobacterium*;

8) it contains a T-DNA left border sequence of a bacterium of the genus *Agrobacterium*;

9) it contains the selectable marker gene for plant transformation *hpt* (hygromycin resistance gene) located between 7) and 8) and expressed in a plant; and

10) it contains restriction endonuclease recognition site(s) located between 7) and 8) for cloning a foreign gene, e.g., an *NspV* site.

[0065] pLC40 was prepared by inserting a region containing the T-DNA region of pSB200PcHm (Figure 1) into p6FRG. It should be noted that p6FRG is a cosmid vector of 8507 bp in full length having the structure shown in Figure 5 (SEQ ID NO: 1 in the Sequence Listing) characterized in that:

[0066] 1) it contains an origin of replication (*oriV*) of an *IncP* plasmid, but does not contain any origin of replication of other plasmid groups;

[0067] 2) it contains the *trfA1* gene, *oriT* and the *incC1* gene of an *IncP* plasmid, and a *cos* site of lambda phage;

[0068] 3) it contains the drug resistance gene *nptIII* (kanamycin resistance gene) expressed in *E.coli* and a bacterium of the genus *Agrobacterium*.

[0069] ii) pLC40GWH (SEQ ID NO: 3, Figure 7)

A binary cosmid vector of 13174 bp in full length. It differs from pLC40 by an insertion of *attB1*, 2 sequences and a deletion of a 317 bp *SspI*-*BalI* region upstream of the RB. It

- 28 -

was prepared by inserting a region containing the T-DNA region of pSB200PcHmGWH (Figure 3) into p6FRG.

[0070] iii) pLC40bar (SEQ ID NO: 4, Figure 8)

A binary cosmid vector of 12884 bp in full length. Principal differences of pLC40bar from pLC40 are in that the selectable marker gene for plant transformation is bar (phosphinotricin resistance gene), and that the orientation of the selectable marker unit (ubiquitin promoter-ubiquitin intron-selectable marker gene for plant transformation) on the T-DNA is opposite. It was prepared by inserting a region containing the T-DNA region of pSB25UNpHm (Figure 2) into p6FRG.

[0071] iv) pLC40GWB (SEQ ID NO: 5, Figure 9)

A binary cosmid vector of 13026 bp in full length. It differs from pLC40 in that the selectable marker gene for plant transformation is bar (phosphinotricin resistance gene) and that attB1, 2 sequences have been inserted. It was prepared by inserting a region containing the T-DNA region of pSB200PcHmGWB (Figure 4) into p6FRG.

[0072]

v) pLC40GWHkorB (SEQ ID NO: 65, Figure 10)

A binary cosmid vector of 14120 bp in full length. It differs from pLC40GWH in that it contains the nucleotide sequence of the korB gene. The korB gene is located near IncC1 described above, and contributes to the stability of IncP plasmids as IncC1 does. The nucleotide sequence of the korB gene of the present invention is not specifically limited so far as it has the function of the korB gene contributing to

- 29 -

the stability of IncP plasmids.

[0073] This sequence has molecular biological properties described in detail in Pansegrau et al. (1994) J Mol Biol 239: 623-663, and it is defined as nucleotides 57187-58263 of the sequence of Genbank/EMBL Accession Number L27758 (full length 60099 bp). This corresponds to nucleotides 6306-7382 of SEQ ID NO: 65.

[0074] The korB can be conventionally prepared from an IncP plasmid such as pVK102 (Knauf and Nester 1982 Plasmid 8: 45-54). For example, a sequence amplified by PCR from pVK102 (nucleotides 6306-7382 of SEQ ID NO: 65) can be used as the korB gene.

[0075] Alternatively, a nucleic acid containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 6306-7382 of SEQ ID NO: 65 under stringent conditions and having the function of the korB gene can be used. Alternatively, a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 6306-7382 of SEQ ID NO: 65 and having the function of the korB gene can also be used.

[0076] vi) pLCleo (SEQ ID NO: 66, Figure 11)

A binary cosmid vector of 14195 bp in full length. It differs from pLC40GWHkorB in that it contains a PspOMI site in the multicloning site, a PI-SceI upstream of it, and an attB3 site upstream of the ubiquitin promoter.

[0077] vii) pLC40GWHvG1 (SEQ ID NO: 7, Figure 13)

A binary cosmid vector of 14222 bp in full length. It

- 30 -

differs from pLC40GWH in that the virG gene has been inserted. It was prepared by inserting the virG gene outside the T-DNA of pPLC40GWH.

[0078] Those skilled in the art can readily derive equivalents to the seven cosmid vectors described above, i.e.,

i) the cosmid vector pLC40 consisting of the nucleotide sequence of SEQ ID NO: 2;

ii) the cosmid vector pLC40GWH consisting of the nucleotide sequence of SEQ ID NO: 3;

iii) the cosmid vector pLC40bar consisting of the nucleotide sequence of SEQ ID NO: 4;

iv) the cosmid vector pLC40GWB consisting of the nucleotide sequence of SEQ ID NO: 5;

v) the cosmid vector pLC40GWHKorB consisting of the nucleotide sequence of SEQ ID NO: 65;

vi) the cosmid vector pLCleo consisting of the nucleotide sequence of SEQ ID NO: 66; and

vii) the cosmid vector pLC40GWHvG1 consisting of the nucleotide sequence of SEQ ID NO: 7;

said equivalents having similar functions to those of these vectors even if the nucleotide sequences are not completely identical. Thus, these "equivalents" are also included as preferred embodiments of the cosmid vectors of the present invention.

[0079] For example, it is thought that even if the nucleotide sequences of the cosmid vectors of the present invention i)-vii) above are modified especially in parts other than the elements related to criteria 1)-10) above (e.g., oriV

- 31 -

in criterion 1), or the trfA1 gene in criterion 2)), they perform similar functions to those of the original vectors as cosmid vectors. Moreover, more than one genes or restriction endonuclease sites having similar functions to those of the drug resistance gene in 6), the selectable marker gene for plant transformation in 9), and the restriction endonuclease recognition site(s) in 10) among criteria 1)-10) are known even if the nucleotide sequences are not completely identical to the nucleotide sequences in the cosmid vectors i)-vii), and those skilled in the art can modify these parts as appropriate.

[0080] Therefore, an "equivalent" to each of the cosmid vectors of the present invention i)-vii) preferably refers to a nucleotide sequence identical to or having an identity of at least 95% or more, 97% or more, 98% or more or 99% or more, more preferably 99.5% or more to the nucleotide sequence of each cosmid vector in the nucleotide sequences of the elements related to criteria 1)-5) and 7)-8) of the cosmid vectors of the present invention, especially the core sequences in these criteria or refers to a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of each cosmid vector under stringent conditions, said equivalent containing a mutation elsewhere in the nucleotide sequence while having similar function and effect to those of each vector. More preferably, it refers to a nucleotide sequence identical to the nucleotide sequence of each cosmid vector in the nucleotide sequences of the elements related to criteria 1)-10) of the cosmid vectors of the present invention, especially the core sequences in these criteria and containing a mutation

- 32 -

elsewhere in the nucleotide sequence while having similar function and effect to those of each vector.

[0081] The degree of mutation is not specifically limited, but the "equivalent" preferably consists of a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of each of cosmid vectors i)-vii) under stringent conditions. The number of nucleotides that can be mutated is more preferably one or more, still more preferably one to a few (e.g., to the extent at which a mutation can be introduced by known site-directed mutagenesis).

[0082] The "equivalent" also preferably consists of a nucleotide sequence having an identity of 95% or more, 97% or more, 98% or more or 99% or more, more preferably 99.5% or more to a nucleotide sequence selected from the nucleotide sequences of cosmid vectors i)-vii).

[0083] The percent identity of two nucleic acid sequences can be determined by visual inspection and mathematical calculation, or more preferably, the comparison is done by comparing sequence information using a computer program. An exemplary, preferred computer program is the Genetics Computer Group (GCG; Madison, Wis.) Wisconsin package version 10.0 program, "GAP" (Devereux et al., 1984, Nucl. Acids Res. 12: 387). This "GAP" program can be used to compare not only two nucleic acid sequences but also two amino acid sequences or a nucleic acid sequence and an amino acid sequence. The preferred default parameters for the "GAP" program include (1) The GCG implementation of a unary comparison matrix (containing a value of 1 for identities and 0 for non-

- 33 -

identities) for nucleotides, and the weighted amino acid comparison matrix of Gribskov and Burgess, Nucl. Acids Res. 14: 6745, 1986 as described by Schwartz and Dayhoff, eds., "Atlas of Polypeptide Sequence and Structure", National Biomedical Research Foundation, pp. 353-358, 1979; or other comparable comparison matrices; (2) a penalty of 30 for each gap and an additional penalty of 1 for each symbol in each gap for amino acid sequences, or penalty of 50 for each gap and an additional penalty of 3 for each symbol in each gap for nucleotide sequences; (3) no penalty for end gaps; and (4) no maximum penalty for long gaps. Other programs used by those skilled in the art of sequence comparison can also be used, such as, for example, the BLASTN program version 2.2.7, available for use via the National Library of Medicine website: <http://www.ncbi.nlm.nih.gov/blast/bl2seq/bls.html>, or the UW-BLAST 2.0 algorithm. Standard default parameter settings for UW-BLAST 2.0 are described at the following Internet site: <http://blast.wustl.edu>. In addition, the BLAST algorithm uses the BLOSUM62 amino acid scoring matrix, and optional parameters that can be used are as follows: (A) inclusion of a filter to mask segments of the query sequence that have low compositional complexity (as determined by the SEG program of Wootton and Federhen (Computers and Chemistry, 1993); also see Wootton and Federhen, 1996, Analysis of compositionally biased regions in sequence databases, Methods Enzymol. 266: 554-71) or segments consisting of short-periodicity internal repeats (as determined by the XNU program of Claverie and States (Computers and Chemistry, 1993)), and

- 34 -

(B) a statistical significance threshold for reporting matches against database sequences, or E-score (the expected probability of matches being found merely by chance, according to the stochastic model of Karlin and Altschul, 1990; if the statistical significance ascribed to a match is greater than this E-score threshold, the match will not be reported.); preferred E-score threshold values are 0.5, or in order of increasing preference, 0.25, 0.1, 0.05, 0.01, 0.001, 0.0001, 1e-5, 1e-10, 1e-15, 1e-20, 1e-25, 1e-30, 1e-40, 1e-50, 1e-75, or 1e-100.

[0084] Plant transformation methods

The present invention also provides a plant transformation method using a cosmid vector of the present invention. Specifically, the plant transformation method of the present invention comprises transforming a plant with a bacterium of the genus *Agrobacterium* harboring a vector containing a nucleic acid fragment of a plant inserted into a cosmid vector of the present invention.

[0085] The type of the nucleic acid fragment inserted into the cosmid vector is not specifically limited, and any fragment can be used, such as a genomic DNA fragment, a cDNA fragment, etc. The nucleic acid fragment is preferably a genomic DNA fragment, more preferably a genomic DNA fragment derived from a plant. The size of the DNA fragment inserted is preferably 1 kb or more, more preferably 10 kb or more, still more preferably 20 kb or more, still more preferably 25-40 kb, still more preferably 30-40 kb.

[0086] The preparation and introduction of the nucleic acid

- 35 -

fragment into the cosmid vector and other operations can be performed by a known method, e.g., the method described in WO2005/040374.

[0087] The source of the nucleic acid fragment are not specifically limited. In the case of plant genomic DNA fragments, preferred examples include plants in which heterosis may occur by cross with recipient plants of genomic DNA fragments. When the recipient plant is Japonica rice, for example, the donor is preferably a wild species of rice *Oryza rufipogon* or Indica rice. When the recipient plant is a specific variety of maize, preferred examples of donor plants include the other varieties of maize and wild species of teosinte. In general, higher heterosis has been observed between more distantly related plants.

[0088] The recipient plant used for transformation may belong to a different species from that of the donor plant of the genomic DNA or a different variety of the same species or the same variety of the same species. Preferred examples of plants include substantially unrestricted wide range of plants, e.g., cereals such as rice, barley, wheat, maize, sorghum, or millet such as an extremely early maturing variety of Italian millet or pearl millet; industrial crops such as sugar cane; pasture grasses such as Sudan grass or rose grass; plants for producing luxury grocery items such as coffee, cocoa, tea and tobacco; vegetables; fruits; ornamental plants such as flowers; weeds such as *Arabidopsis*, etc.

[0089] The cosmid vectors of the present invention were obtained especially to improve the efficiency of

- 36 -

Agrobacterium-mediated transformation among biological transfer methods. Therefore, the plant transformation method is preferably Agrobacterium-mediated. However, other known plant transformation methods are not excluded. For example, known methods include physical transfer methods such as microinjection, electroporation, particle gun, silicon carbide-mediated method and air injection; and chemical transfer methods such as polyethylene glycol-mediated method.

[0090] The type of Agrobacterium strain is not specifically limited so far as it has an antibiotic resistance other than the antibiotic resistance (gene) for the bacterium used for the construction of the vector, and known strains such as LBA4404, A281, EHA105, PC2760, etc. can be used.

[0091] Map-based cloning method

The present invention also provides an efficient map-based cloning method using a cosmid vector of the present invention as described above. The map-based cloning method is characterized in that it comprises the steps of:

[0092] 1) partially or completely digesting BAC clones containing candidate genes responsible for a plant phenotype with a restriction endonuclease;

2) subcloning DNA fragments obtained in step 1) using a cosmid vector to construct a library; and

3) individually transferring clones constituting the library into a plant to evaluate the phenotypes of transformed plants.

[0093] In this map-based cloning method, the DNA fragments obtained in step 1) preferably have a size of, but not limited

- 37 -

to, 25-40 kb. More preferably, the cosmid vector in 2) is a cosmid vector as described in the section "Cosmid vectors" above.

[0094] The "candidate genes" refer to a group of genes including genes likely to be responsible for a plant phenotype. The "plant phenotype" is not specifically limited, but includes various agriculturally useful phenotypes such as high vigor of the whole plant, large sizes of the plant and organs, high yield, high growth speed, disease and insect resistance, resistance to various environmental stresses such as drought, high temperature, low temperature, etc., an increase or decrease of a specific component, an increase or decrease of a specific enzyme activity, dwarfness, etc.

[0095] For example, suppose that candidate genes were found to be contained in DNA fragments carried by more than one BAC clones of 100-200 kb. Then, these cloned DNAs are partially or completely digested with an appropriate restriction endonuclease to prepare overlapping fragments of about 40 kb, which are then subcloned using a transformation vector of the present invention. It is not necessary to investigate in detail the relative positions and the overlapping of the subcloned DNA fragments. According to a statistical calculation, any site on original fragments in 200 kb clones is maintained by randomized 21 subclones with a 99% probability (e.g., see [0043]-[0047] in WO2005/040374).

[0096] Then, each subclone is transferred into a plant to prepare about 10 independent transformants per subclone, and the effect of the gene is analyzed. According to this

- 38 -

operation, the candidate region can be first narrowed down to 40 kb by identifying subclones containing candidate genes, and then the candidate region can be further restricted to a very narrow region by comparing the experimental results between adjacent subclones. Thus, the efficiency of identifying candidate genes greatly improves.

[0097] Transformation method additionally using the virG gene (and the virB gene)

In a preferred embodiment, the plant transformation method of the present invention is characterized in that it uses a bacterium of the genus *Agrobacterium* harboring the following elements:

[0098] 1a) a vector containing a nucleic acid fragment of a plant and the virG gene of a bacterium of the genus *Agrobacterium* inserted into the cosmid vector of the present invention; or

1b) a vector containing a nucleic acid fragment of a plant inserted into the cosmid vector of the present invention, and a plasmid capable of coexisting with an IncP plasmid in a cell of a bacterium of the genus *Agrobacterium* and containing the virG gene of a bacterium of the genus *Agrobacterium*, and

2) a Ti plasmid or Ri plasmid of a bacterium of the genus *Agrobacterium*.

[0099] virG is one of vir genes of *Agrobacterium* that play a role in the transfer of the T-DNA into plants, and it is regarded as a transcription factor of the virB gene or the like (Winans et al. 1986 Proc. Natl. Acad. Sci. USA 83: 8278-

- 39 -

8282). As an example of *virG*, *virGN54D* is a variant in which the amino acid at position 54 of the *virG* protein is changed from asparagine to aspartic acid to increase the expression of the *virB* gene as compared with the wild-type *virG* (Pazour et al. 1992 J.Bac 174:4169-4174). In the present transformation method, the *virG* gene is preferably *virGN54D*.

[0100] In an embodiment of the present invention 1a), the *virG* gene may be further inserted into a cosmid vector of the present invention containing a nucleic acid fragment of a plant. In embodiments where a cosmid vector of the present invention already carries the *virG* gene (e.g., pLC40GWHvG1), the *virG* gene need not be further inserted.

[0101] Alternatively, the *virG* gene may exist in an independent plasmid separate from a cosmid vector of the present invention. In this case, *Agrobacterium* in the method of the present invention harbors a plasmid capable of coexisting with an IncP plasmid in *Agrobacterium* cells and containing the *virG* gene of a bacterium of the genus *Agrobacterium*, in addition to the cosmid vector (embodiment 1b).

[0102] The Ti plasmid or Ri plasmid is not specifically limited, but preferably disarmed by deleting the T-DNA.

[0103] The plasmid containing the *virG* gene of a bacterium of the genus *Agrobacterium* in 1b) may contain an origin of replication of an IncW plasmid. Preferably, it is pVGW having the structure shown in Figure 14. More preferably, it is pVGW2 having the structure shown in Figure 15.

[0104] The plasmid containing the *virG* gene of a bacterium

- 40 -

of the genus *Agrobacterium* in 1b) may further contain the *virB* gene of a bacterium of the genus *Agrobacterium*. Here again, the plasmid may contain an origin of replication of an IncW plasmid. Such a plasmid is preferably pTOK47.

[0105] The *virB* gene of a bacterium of the genus *Agrobacterium* is described in detail in Ward et al. (1988) *J Biol Chem* 263: 5804-5814. For example, it can be conventionally prepared from a plasmid such as pSB1 (Komari et al. 1996 *Plant J* 10: 165-174). The nucleotide sequence of *virB* is defined as, e.g., nucleotides 3416-12851 of the nucleotide sequence of Genbank/EMBL Accession Number: AB027255 (pSB1). As a non-limitative example, a DNA containing a nucleotide sequence hybridizing to this sequence or a complementary strand thereto under stringent conditions can be used as the *virB* gene.

[0106] These transformation methods are more effective for plants normally associated with low efficiency of *Agrobacterium*-mediated transformation, e.g., including, but not limited to, maize and soybean. When the nucleic acid fragment to be transferred is large (e.g., 25-40 kb as a non-limitative example) or has a complex structure (e.g., a highly repeated sequence as a non-limitative example), pVGW described below is preferably used as a plasmid containing the *virG* gene of a bacterium of the genus *Agrobacterium*.

[0107] Plasmid vectors

In plants such as maize, wherein transformation is difficult to occur, the efficiency of *Agrobacterium*-mediated transformation with standard binary vectors containing the T-

- 41 -

DNA is very low except for special cases (Frame et al. 2002 Plant Physiol 129: 13-22). Previous reports show an increase in the efficiency of transient expression in maize by the coexistence of a binary vector with another plasmid containing the virGN54D gene, a variant of the virG gene in Agrobacterium (Hansen et al. 1994 ProNAS 91:7603-7607), and a high efficiency maize transformation system with a binary vector containing virG and virB (Ishida et al. 1996 Nat Biotechnol 14:745-50).

[0108] However, no report has shown that the maize transformation efficiency was increased by the coexistence of a binary vector with a plasmid containing virG or virGN54D in Agrobacterium.

[0109] The cosmid vectors of the present invention (pLC vectors) (IncP plasmids) are also expected to further improve the maize transformation efficiency. Plasmids capable of coexisting with an IncP plasmid include e.g., IncW plasmids (Close et al. 1984 Plasmid 12: 111-118). Previously reported IncW vectors containing virG are large because they contain origins of replication of other plasmids such as pBR322 ori. For example, pTOK47 contains IncW (pSa) ori and pBR322 ori (as well as not only virG but also virB) and it has a full length of about 28 kb (Jin et al. 1987 J Bacteriol 169: 4417-4425). pYW48 contains IncW (pSa) ori and pBR322 ori (as well as not only virG but also virA) and it has a full length of 15.5 kb (Wang et al. 2000 Gene 242: 105-114). Such vectors can also be used in the transformation methods of the present invention. However, these vectors are so long that they may

- 42 -

cause problems in stability in bacteria when they coexist with a pLC vector containing a large fragment, and therefore, small vectors capable of coexisting with a pLC vector and containing virG are desirable.

[0110] As a means to solve these problems, the present invention provides a small plasmid vector capable of further improving the transformation efficiency by the coexistence with the cosmid vectors of the present invention described above.

[0111] The plasmid vector of the present invention satisfies all of the criteria below.

[0112] 1) it contains an origin of replication of an IncW plasmid, but does not contain any origin of replication of other plasmid groups;

2) it contains the repA gene necessary for the replication of an IncW plasmid;

3) it contains a drug resistance gene expressed in E.coli and a bacterium of the genus Agrobacterium; and

4) it contains the virG gene of a bacterium of the genus Agrobacterium.

1) The nucleotide sequence of the origin of replication of an IncW plasmid of the present invention is not specifically limited so far as it has the function as an origin of replication of an IncW plasmid.

[0113] The origin of replication of an IncW plasmid has molecular biological properties described in detail in Okumura and Kado (1992 Mol Gen Genet 235: 55-63), and it is defined as nucleotides 2170-2552 of Genbank/EMBL Accession Number: U30471

- 43 -

(full length 5500 bp). This corresponds to nucleotides 2832-3214 of SEQ ID NO: 8.

[0114] The origin of replication of an IncW plasmid can be conventionally prepared from an IncW plasmid such as pTOK47 (Jin et al. 1987 J Bacteriol 169: 4417-4425). For example, nucleotides 2832-3214 of SEQ ID NO: 8 in a 2.7 kb DNA amplified by PCR from pTOK47 with repA necessary for the replication of an IncW plasmid described below can be used.

[0115] Alternatively, a nucleic acid containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 2832-3214 of SEQ ID NO: 8 described above under stringent conditions and having the function of an origin of replication of an IncW plasmid can also be used. Alternatively, a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 2832-3214 of SEQ ID NO: 8 described above and having the function of an origin of replication of an IncW plasmid can also be used.

[0116] It will be recognized by those skilled in the art that a shorter region in nucleotides 2832-3214 of SEQ ID NO: 8 may be selected as a sequence having a similar function.

[0117] 2) The nucleotide sequence of the repA gene of the present invention is not specifically limited so far as it has the function as the repA gene necessary for the replication of an IncW plasmid.

The repA necessary for the replication of an IncW plasmid has molecular biological properties described in detail in

- 44 -

Okumura and Kado (1992 Mol Gen Genet 235: 55-63), and it is defined as nucleotides 1108-2079 of Genbank/EMBL Accession Number:U30471 (full length 5500 bp). This corresponds to nucleotides 1770-2741 of SEQ ID NO: 8.

[0118] The repA necessary for the replication of an IncW plasmid can be conventionally prepared from an IncW plasmid such as pTOK47 (Jin et al. 1987 J Bacteriol 169: 4417-4425). For example, nucleotides 1770-2741 of SEQ ID NO: 8 in a 2.7 kb DNA amplified by PCR from pTOK47 with an origin of replication of an IncW plasmid described above can be used.

[0119] Alternatively, a nucleic acid containing a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of nucleotides 1770-2741 of SEQ ID NO: 8 described above under stringent conditions and having the function of the repA gene necessary for the replication of an IncW plasmid can also be used. Alternatively, a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably 97%, still more preferably 99% to the nucleotide sequence of nucleotides 1770-2741 of SEQ ID NO: 8 described above and having the function of the repA gene necessary for the replication of an IncW plasmid can also be used.

[0120] It will be recognized by those skilled in the art that a shorter region in nucleotides 1770-2741 of SEQ ID NO: 8 may be selected as a sequence having a similar function.

[0121] 3) The drug resistance gene expressed in E.coli and a bacterium of the genus Agrobacterium is used as a selectable marker for transformation. This drug resistance gene confers

- 45 -

e.g., antibiotic resistance or autotrophy, including, but not limited to, a kanamycin resistance gene, a spectinomycin resistance gene, an ampicillin resistance gene, a tetracycline resistance gene, a gentamycin resistance gene, a hygromycin resistance gene, etc.

[0122] 4) The *virG* gene of a bacterium of the genus *Agrobacterium* and *virGN54D* have molecular biological properties described in detail in Winans et al. (1986) Proc. Natl. Acad. Sci. USA 83: 8278-8282 and Pazour et al. (1992) J. Bacteriol. 174: 4169-4174, Hansen et al. 1994 Proc. Natl. Acad. Sci. USA 91: 7603-7607, respectively. *virG* is a transcription regulator (activator) of other *vir* genes such as *virB* and *virE*. *virG* is activated upon regulation (phosphorylation) by *virA*, whereas *virGN54D* is a variant in a permanently activated state without this regulation. The *virG* gene can be prepared by conventional procedure and *virGN54D* can be prepared by mutagenesis both from a plasmid such as pTOK47 (Jin et al. 1987 J Bacteriol 169: 4417-4425). For example, 1 kb *virG* DNA (nucleotides 4024-5069 of SEQ ID NO: 7) amplified by PCR from pTOK47 and 1 kb *virGN54D* DNA (nucleotides 1-1080 of SEQ ID NO: 8) amplified and prepared by PCR mutagenesis can be used.

[0123] Alternatively, a nucleic acid containing a nucleotide sequence hybridizing to a complementary strand of these nucleotide sequences under stringent conditions and having the function of the *virG* gene of a bacterium of the genus *Agrobacterium* or a nucleic acid containing a nucleotide sequence having an identity of at least 95%, more preferably

- 46 -

97%, still more preferably 99% to these nucleotide sequences and having the function of the virG gene of a bacterium of the genus Agrobacterium can also be used.

[0124] The plasmid vector of the present invention preferably has a full length of 10 kb or less, more preferably 5 kb or less.

[0125] The plasmid vector of the present invention is preferably the pVGW vector having the structure shown in Figure 14. More preferably, it is pVGW2 having the structure shown in Figure 15. pVGW and pVGW2 are vectors satisfying all of criteria 1)-4) above. pVGW shown as SEQ ID NO: 8 has a full length of 4531 bp, and pVGW2 shown as SEQ ID NO: 67 has a full length of 4836 bp, and they are characterized in that:

1) they contain an origin of replication of an IncW plasmid, but do not contain any origin of replication of other plasmid groups;

2) they contain the repA gene necessary for the replication of an IncW plasmid;

3) they contain a gentamycin resistance gene as a drug resistance gene expressed in E.coli and a bacterium of the genus Agrobacterium; and

4) they contain the virGN54D gene of a bacterium of the genus Agrobacterium.

Among these components, the origin of replication of an IncW plasmid and the repA gene necessary for the replication of an IncW plasmid were simultaneously cloned and the gentamycin resistance gene and the virGN54D gene were separately cloned, after which all of the three DNA fragments

- 47 -

(four components) were assembled.

[0126] Those skilled in the art can readily derive equivalents to the two plasmid vectors pVGW, pVGW2 of the present invention described above, said equivalents having similar functions to those of these vectors even if the nucleotide sequences are not completely identical. Thus, these "equivalents" are also included as preferred embodiments of the plasmid vectors of the present invention.

[0127] For example, it is thought that even if the nucleotide sequences of the plasmid vectors of the present invention are modified especially in parts other than the elements related to criteria 1)-4) above (e.g., the origin of replication of an IncW plasmid in criterion 1)), they perform similar functions to those of the original vectors as plasmid vectors. Moreover, more than one genes having similar functions to those of the drug resistance gene in 3) among criteria 1)-4) are known even if the nucleotide sequences are not completely identical to the nucleotide sequences in the plasmid vectors, and those skilled in the art can modify these parts as appropriate.

[0128] Therefore, an "equivalent" to each of the plasmid vectors of the present invention preferably refers to a nucleotide sequence identical to or having an identity of at least 95% or more, 97% or more, 98% or more or 99% or more, more preferably 99.5% or more to the nucleotide sequence of each plasmid vector in the nucleotide sequences of the elements related to criteria 1)-2) and 4) of the plasmid vectors of the present invention or refers to a nucleotide

- 48 -

sequence hybridizing to a complementary strand of the nucleotide sequence of each plasmid vector under stringent conditions, said equivalent containing a mutation elsewhere in the nucleotide sequence while having similar function and effect to those of each vector. More preferably, it refers to a nucleotide sequence identical to the nucleotide sequence of each plasmid vector in the nucleotide sequences of the elements related to criteria 1)-4) of the plasmid vectors of the present invention and containing a mutation elsewhere in the nucleotide sequence while having similar function and effect to those of each vector.

[0129] The degree of mutation is not specifically limited, but the "equivalent" preferably consists of a nucleotide sequence hybridizing to a complementary strand of the nucleotide sequence of each plasmid vector under stringent conditions. The number of nucleotides that can be mutated is more preferably one or more, still more preferably one to a few (e.g., to the extent at which a mutation can be introduced by known site-directed mutagenesis).

[0130] The "equivalent" also preferably consists of a nucleotide sequence having an identity of 95% or more, 97% or more, 98% or more or 99% or more, more preferably 99.5% or more to a nucleotide sequence selected from the nucleotide sequences of the plasmid vectors.

[0131] As used herein, the expression "under stringent conditions" refers to hybridization under conditions of moderate or high stringency. Specifically, conditions of moderate stringency can be readily determined by those having

- 49 -

ordinary skill in the art based on, for example, the length of the DNA. The basic conditions are set forth by Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 3rd Ed., Chapters 6-7, Cold Spring Harbor Laboratory Press, 2001, and include use of a prewashing solution for the nitrocellulose filters containing 5 x SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0), hybridization conditions of 2 x SSC to 6 x SSC with or without about 50% formamide at about 40°C to 50°C (or other similar hybridization solution, such as Stark's solution, in about 50% formamide at about 42°C), and washing conditions of 0.5 to 6 x SSC, 0.1% SDS at about 40°C to 60°C. Preferably, conditions of moderate stringency include hybridization conditions (and washing conditions) of 6 x SSC at about 50°C. Conditions of high stringency can also be readily determined by the skilled artisan based on, for example, the length of the DNA.

[0132] Generally, such conditions include hybridization and/or washing at higher temperatures and/or lower salt concentrations than in the conditions of moderate stringency (e.g., hybridization in 6 x SSC to 0.2 x SSC, preferably 6 x SSC, more preferably 2 x SSC, most preferably 0.2 x SSC at about 65 °C), and are defined to involve hybridization conditions as above and washing in 0.2 x SSC, 0.1% SDS at about 65°C to 68°C. SSPE (1 x SSPE = 0.15 M NaCl, 10 mM NaH₂PO₄, and 1.25 mM EDTA, pH 7.4) can be substituted for SSC (1 x SSC = 0.15 M NaCl and 15 mM sodium citrate) for use as hybridization and washing buffers, and washing is continued for 15 minutes after completion of hybridization.

[0133] Commercially available hybridization kits not using

- 50 -

radioactive substances as probes can also be used.

Specifically, hybridization can be performed by using ECL direct labeling & detection system (from Amersham), etc.

Stringent hybridization conditions include hybridization in the hybridization buffer included in the kit containing 5% (w/v) Blocking reagent and 0.5 M NaCl at 42°C for 4 hours, followed by washing twice in 0.4% SDS, 0.5 x SSC at 55°C for 20 minutes, and once in 2 x SSC at room temperature for 5 minutes.

[0134] pVGW is characterized in that it is small and stable. Specifically, it is effective for improving the transformation efficiency by the coexistence with pLC especially when large fragments are used and/or when maize is used as a host. It is also effective for improving the efficiency of transformation of maize or the like by the coexistence with an ordinary vector other than pLC.

EFFECTS OF THE INVENTION

[0135] The vectors (pLC vectors) of the present invention provide the following advantages that could not be achieved by known vectors:

[0136] - they allow efficient cloning of DNA fragments of about 25-40 kb in size, preferably 30-40 kb;

- they are stably maintained in E.coli and Agrobacterium cells;

- they can be efficiently introduced in Agrobacterium;

- the copy number per cell in E.coli and Agrobacterium is 4-5; and

- they allow efficient transfer of only cloned DNA

- 51 -

fragments of interest into plants, preferably monocotyledons (the transformation efficiency of pLC vectors is 90%, in contrast to the transformation efficiency of pSB vectors 60%).

[0137] The combined use of the pLC vectors of the present invention and the pVGW vector allows efficient gene transfer into even plants that are relatively difficult to transform such as maize.

[0138] Candidate gene sites can be narrowed down with little expenditure of labor and time by map-based cloning with the pLC vectors of the present invention.

[0139] The present invention is significantly effective even if mapping information is very limited. For example, suppose that nothing is known except for the presence of candidate genes at an end region of a chromosome. If the entire length of one chromosome is 40 Mb, for example, its end region may be assumed to be 2 Mb. This region can be covered by about 20 BAC clones carrying an insert fragment of 150 kb on average by constructing a library of aligned BACs (BAC contig). Therefore, if 20 subclones are prepared from each BAC, almost all candidate genes can be rapidly identified by preparing a total of 400 fragments and 4000 recombinants. Thus, genes can be identified with little expenditure of labor and time unimaginable from conventional techniques by using the technique of the present invention.

[0140] Embodiments of the present invention

The present invention preferably the following embodiments.

[Embodiment 1]

- 52 -

A cosmid vector having a full length of 15 kb or less characterized in that:

1) it contains an origin of replication (oriV) of an IncP plasmid, but does not contain any origin of replication of other plasmid groups;

2) it contains the trfA1 gene of an IncP plasmid;

3) it contains an origin of conjugative transfer (oriT) of an IncP plasmid;

4) it contains the incC1 gene of an IncP plasmid;

5) it contains a cos site of lambda phage and the cos site is located outside the T-DNA;

6) it contains a drug resistance gene expressed in E.coli and a bacterium of the genus Agrobacterium;

7) it contains a T-DNA right border sequence of a bacterium of the genus Agrobacterium;

8) it contains a T-DNA left border sequence of a bacterium of the genus Agrobacterium;

9) it contains a selectable marker gene for plant transformation located between 7) and 8) and expressed in a plant; and

10) it contains restriction endonuclease recognition site(s) located between 7) and 8) for cloning a foreign gene.
[Embodiment 2]

The cosmid vector of Embodiment 1 wherein the selectable marker gene for plant transformation is selected from the group consisting of a hygromycin resistance gene, a phosphinotricin resistance gene and a kanamycin resistance gene.

- 53 -

[Embodiment 3]

The cosmid vector of Embodiment 1 or 2, which contains the korB gene of an IncP plasmid.

[Embodiment 4]

The cosmid vector of any one of Embodiments 1 to 3 selected from the group consisting of:

the cosmid vector pLC40 consisting of the nucleotide sequence of SEQ ID NO: 2 or an equivalent thereof;

the cosmid vector pLC40GWH consisting of the nucleotide sequence of SEQ ID NO: 3 or an equivalent thereof;

the cosmid vector pLC40bar consisting of the nucleotide sequence of SEQ ID NO: 4 or an equivalent thereof;

the cosmid vector pLC40GWB consisting of the nucleotide sequence of SEQ ID NO: 5 or an equivalent thereof;

the cosmid vector pLC40GWHKorB consisting of the nucleotide sequence of SEQ ID NO: 65 or an equivalent thereof;

the cosmid vector pLCleo consisting of the nucleotide sequence of SEQ ID NO: 66 or an equivalent thereof; and

the cosmid vector pLC40GWHvG1 consisting of the nucleotide sequence of SEQ ID NO: 7 or an equivalent thereof.

[Embodiment 5]

The cosmid vector of Embodiment 4 selected from the group consisting of:

the cosmid vector pLC40 consisting of the nucleotide sequence of SEQ ID NO: 2;

the cosmid vector pLC40GWH consisting of the nucleotide sequence of SEQ ID NO: 3;

the cosmid vector pLC40bar consisting of the nucleotide

- 54 -

sequence of SEQ ID NO: 4;

the cosmid vector pLC40GWB consisting of the nucleotide sequence of SEQ ID NO: 5;

the cosmid vector pLC40GWHKorB consisting of the nucleotide sequence of SEQ ID NO: 65;

the cosmid vector pLCleo consisting of the nucleotide sequence of SEQ ID NO: 66; and

the cosmid vector pLC40GWHvG1 consisting of the nucleotide sequence of SEQ ID NO: 7.

[Embodiment 6]

A method for transforming a plant, comprising transforming the plant with a bacterium of the genus *Agrobacterium* harboring an expression vector containing a nucleic acid fragment of a plant inserted into the cosmid vector of any one of Embodiments 1 to 5.

[Embodiment 7]

The method of Embodiment 6 wherein the nucleic acid fragment inserted has a size of 25-40 kb.

[Embodiment 8]

The method of Embodiment 6 or 7 characterized in that it uses a bacterium of the genus *Agrobacterium* harboring the following elements for transforming the plant:

1a) a vector containing a nucleic acid fragment of a plant and the *virG* gene of a bacterium of the genus *Agrobacterium* inserted into the cosmid vector of any one of Embodiments 1 to 5; or

1b) a vector containing a nucleic acid fragment of a plant inserted into the cosmid vector of any one of

- 55 -

Embodiments 1 to 5, and a plasmid capable of coexisting with an IncP plasmid in a cell of a bacterium of the genus *Agrobacterium* and containing the *virG* gene of a bacterium of the genus *Agrobacterium*, and

2) a Ti plasmid or Ri plasmid of a bacterium of the genus *Agrobacterium*.

[Embodiment 9]

The method of Embodiment 8 wherein the *virG* gene of a bacterium of the genus *Agrobacterium* in 1a) or 1b) is *virGN54D*.

[Embodiment 10]

The method of Embodiment 8 wherein the plasmid containing the *virG* gene of a bacterium of the genus *Agrobacterium* in 1b) contains an origin of replication of an IncW plasmid.

[Embodiment 11]

The method of Embodiment 10 wherein the plasmid containing the *virG* gene of a bacterium of the genus *Agrobacterium* in 1b) is pVGW having the structure shown in Figure 14 or pVGW2 having the structure shown in Figure 15.

[Embodiment 12]

The method of Embodiment 8 wherein the plasmid containing the *virG* gene of a bacterium of the genus *Agrobacterium* in 1b) further contains the *virB* gene of a bacterium of the genus *Agrobacterium*.

[Embodiment 13]

The method of Embodiment 12 wherein the plasmid containing the *virG* gene of a bacterium of the genus *Agrobacterium* in 1b) contains an origin of replication of an

- 56 -

IncW plasmid.

[Embodiment 14]

The method of Embodiment 13 wherein the plasmid containing the virG gene of a bacterium of the genus Agrobacterium in 1b) is pTOK47.

[Embodiment 15]

A map-based cloning method comprising the steps of:

- 1) partially or completely digesting BAC clones containing candidate genes responsible for a plant phenotype with a restriction endonuclease;
- 2) subcloning DNA fragments obtained in step 1) using a cosmid vector to construct a library; and
- 3) individually transferring clones constituting the library into a plant to evaluate the phenotypes of transformed plants.

[Embodiment 16]

The map-based cloning method of Embodiment 15 wherein the DNA fragments obtained in step 1) have a size of 25-40 kb.

[Embodiment 17]

The map-based cloning method of Embodiment 16 wherein the cosmid vector in 2) is the cosmid vector of any one of Embodiments 1 to 5.

[Embodiment 18]

A plasmid vector characterized in that:

- 1) it contains an element necessary for the replication of an IncW plasmid, but does not contain any origin of replication of other plasmid groups;
- 2) it contains the repA gene necessary for the

- 57 -

replication of an IncW plasmid;

3) it contains a drug resistance gene expressed in E.coli and a bacterium of the genus Agrobacterium; and

4) the virG gene of a bacterium of the genus Agrobacterium.

[Embodiment 19]

The plasmid vector of Embodiment 18, which has a full length of 10 kb or less.

[Embodiment 20]

The plasmid vector of Embodiment 19 wherein the virG gene of a bacterium of the genus Agrobacterium is virGN54D.

[Embodiment 21]

The plasmid vector of any one of Embodiments 18 to 20 selected from the group consisting of:

the plasmid vector pVGW consisting of the nucleotide sequence of SEQ ID NO: 8 or an equivalent thereof; and

the plasmid vector pVGW2 consisting of the nucleotide sequence of SEQ ID NO: 67 or an equivalent thereof.

[Embodiment 22]

The plasmid vector pVGW consisting of the nucleotide sequence of SEQ ID NO: 8.

[Embodiment 23]

The plasmid vector pVGW2 consisting of the nucleotide sequence of SEQ ID NO: 67.

[Embodiment 24]

A method for transforming a plant, comprising transforming the plant with a bacterium of the genus Agrobacterium harboring the plasmid vector of any one of

Embodiments 18 to 23.

BRIEF EXPLANATION OF THE DRAWINGS

[0141] [Fig. 1] Figure 1 is a schematic diagram of the vector pSB200PcHm.

[Fig. 2] Figure 2 is a schematic diagram of the vector pSB25UNpHm.

[Fig. 3] Figure 3 is a schematic diagram of the vector pSB200PcHmGWH.

[Fig. 4] Figure 4 is a schematic diagram of the vector pSB200PcHmGWB.

[Fig. 5] Figure 5 is a diagram showing a procedure for constructing the vector pLC40.

[Fig. 6] Figure 6 is a schematic diagram of the vector pLC40.

[Fig. 7] Figure 7 is a schematic diagram of the vector pLC40GWH.

[Fig. 8] Figure 8 is a schematic diagram of the vector pLC40bar.

[Fig. 9] Figure 9 is a schematic diagram of the vector pLC40GWB.

[Fig. 10] Figure 10 is a schematic diagram of the vector pLC40GWHkorB.

[Fig. 11] Figure 11 is a schematic diagram of the vector pLCleo.

[Fig. 12] Figure 12 is a schematic diagram of the vector pLCSBGWBSWa.

[Fig. 13] Figure 13 is a schematic diagram of the vector pLC40GWHvG1.

- 59 -

[Fig. 14] Figure 14 is a schematic diagram of the vector pVGW.

[Fig. 15] Figure 15 is a schematic diagram of the vector pVGW2.

[Fig. 16] Figure 16 shows the results of cloning of a genomic DNA fragment by a pLC vector. An example of a teosinte genomic DNA fragment is shown. M1: marker (1 kb ladder), M2: marker (λ -HindIII); the numbers represent clone numbers, and the arrow indicates the size of the band corresponding to pLC40GWH (13.2 kb). Plasmid DNA of eleven clones from a teosinte library was purified. The DNA was cleaved with the restriction endonucleases HindIII and SacI in the multicloning site at each end of the plasmid insert and separated by agarose gel (0.8%) electrophoresis.

[Fig. 17] Figure 17 shows the results of transformation of a genomic DNA fragment into rice (in the center region of fragment B). M: markers, Yu: Yukihihikari, Ru: *Oryza rufipogon*, Transgenic: transformed rice (two individuals). In the transformed rice, a band derived from *Oryza rufipogon* was detected in addition to a band derived from Yukihihikari.

EXAMPLES

[0142] The following examples further illustrate the present invention but are not intended to limit the technical scope of the invention. Those skilled in the art can readily add modifications/changes to the present invention in the light of the description herein, and those modifications/changes are also included in the technical scope of the present invention.

- 60 -

[0143] Example 1: Construction of pLC series cosmid vectors

In the following procedures, molecular biological experimental methods were performed as described in Sambrook J. and Russell D.W. 2001. Molecular Cloning, A Laboratory Manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA., unless otherwise specified.

[0144] 1) Construction of T-DNA regions

A PacI linker (gttaattaac) (SEQ ID NO: 10) was inserted into the EcoRV site of pSB200 (WO2005/040374) to construct pSB200Pac. The cauliflower mosaic virus 35S promoter in pSB25 (Ishida et al. 1996) was replaced by the ubiquitin promoter of maize (Christensen et al. 1992 Plant Mol Biol 18: 675-689) to construct pSB25U. The adapters HinNspISceRV and HinNspISceFW (Table 1) having recognition sites for the restriction endonuclease NspV and the homing endonuclease I-SceI were annealed. A part of the annealed adapters were phosphorylated with a polynucleotide kinase (PNK, Amersham). The phosphorylated adapters were cloned into the SacI site of pSB200Pac and the HindIII site of pSB25U. The resulting plasmids were designated as pSB200PacHm1 and pSB25UNpHm1, respectively.

[0145] The adapters SpeICEuRV and SpeICEuFW containing a homing endonuclease I-CeuI site (Table 1) were inserted into the SpeI site of pSB200PacHm1 and pSB25UNpHm1. This operation generated the vector pSB200PcHm containing homing endonuclease sites I-SceI and I-CeuI inserted into the SacI and SpeI sites respectively of pSB200Pac (Figure 1), and the vector pSB25UNpHm containing homing endonuclease sites I-SceI (+ NspV

- 61 -

site) and I-CeuI inserted into the HindIII and SpeI sites respectively of pSB25U (Figure 2). Excision by I-SceI and I-CeuI was verified, and the nucleotide sequences were checked by using ABI PRISM Fluorescent Sequencer (Model 310 Genetic Analyzer, from Perkin Elmer) to confirm that a single adapter had been inserted. In these vectors, the I-SceI-selectable marker unit-LB-I-CeuI can be excised.

[0146]

Table 1

Primer Name	Sequence	Length
HinNspISceRV	5'-AgC TTT CgA ATA ggg ATA ACA ggg TAA T-3'	28 mer
HinNspISceFW	5'-AgC TAT TAC CCT gTT ATC CCT ATT CgA A-3'	28 mer
SpeICeuRV	5'-CTA gTA ACT ATA ACg gTC CTA Agg TAg CgA C-3'	31 mer
SpeICeuFW	5'-CTA ggT CgC TAC CTT Agg ACC gTT ATA gTT A-3'	31 mer

SEQ ID NOs: 23-26 in order from the top.

Then, pSB200PcHm was digested with BamHI to remove the hygromycin resistance gene (hpt), and then blunt-ended. This was ligated to the aatR1-ccdB-Cm-aatR fragment (Invitrogen), and transferred into E.coli DB3.1 to select a chloramphenicol-resistant colony, thereby generating the destination vector pDEST3342. Then, the following primers containing an aatB sequence were synthesized (aatB sequences are shown in uppercase) in order to introduce a marker gene into the pDONR/Zeo plasmid (Invitrogen) by the BP reaction.

[0147]

Table 2

Primer Name	Sequence	Length
aatB1-HPT	ggg gAC AAG TTT GTA CAA AAA AGC AGG CTc aat gag ata tga aaa agc c	49mer
HPT-aatB2	ggg gAC CAC TTT GTA CAA GAA AGC TGG GTc tat tcc ttt gcc ctc gga cga g	52mer
aatB1-bar	ggg gAC AAG TTT GTA CAA AAA AGC AGG CTc cat gga ccc aga acg acg c	49mer
bar-aatB2	ggg gAC CAC TTT GTA CAA GAA AGC TGG GTt cct aga cgc gtg aga tca g	49mer

SEQ ID NOs: 27-30 in order from the top.

For the amplification of the Hpt gene, the hpt gene described in Bilang et al. (1991) Gene 100: 247-250 was used as a template DNA along with aatB1-HPT and HPT-aatB2 as primers. For the amplification of the phosphinotricin resistance gene (bar), pSB25 (Ishida et al. 1996) was used as a template DNA along with aatB1-bar and bar-aatB2 as primers (Table 2). In 100 µl of a reaction solution containing 10 ng of each template DNA and 25 pmoles of the primers, 35 cycles of PCR was performed. After the completion of the reaction, the products were recovered by ethanol precipitation and used for the BP reaction (25°C, 6 hrs) according to the protocol attached to BP Clonase Enzyme Mix kit (Invitrogen), and then transferred into E.coli DH5a and E.coli cells harboring plasmids of interest were selected on a low salt LA plate containing the antibiotics Zeocin. The nucleotide sequences of the finally obtained plasmids were confirmed by restriction endonuclease analysis, thereby generating pENT-HPTwt and pENT-bar, respectively.

[0148] The destination vector (pDEST3342) prepared before and the entry vectors (pENT-HPTwt and pENT-bar) were used to

- 63 -

prepare final plasmids of interest by the LR reaction. After the reaction in 20 µl of a reaction solution (containing 300 ng each of the destination vector and entry vectors) at 25°C for 4 hours according to the protocol attached to GATEWAY LR Clonase Enzyme Mix, the reaction products were transferred into E.coli DH5α by electroporation. Plasmid DNAs were prepared from colonies grown on an LA plate containing spectinomycin and candidate clones were selected by restriction fragment patterns. They were confirmed by
10 nucleotide sequence analysis to contain an *aadB* sequence and the sequence of the *HPT* gene or *bar* gene and designated as pSB200PcHmGWH (Figure 3) and pSB200PcHmGWB (Figure 4), respectively.

[0149] 2) Construction of the cosmid vector pLC40

A PCR reaction was performed using Pyrobest DNA Polymerase (Takara) along with OriV3'ClaFW, OriV5'PvNhEc, OriT5'BglRV, OriT3'SpEcFW, Inc5'XbRV, Inc3'BgEcFW, R5'XhoIRV, R3'BmEcFW, 121KIII5'NspV, 121KIII3'SalI, COS5'BmRV and COS3'MunFW designed as PCR primers for amplifying a DNA
20 fragment containing *oriV*, a DNA fragment containing *oriT*, a DNA fragment containing the *incC1* gene, a DNA fragment containing the *trfA1* gene, all of which are derived from the IncP plasmid pVK102 (Knauf and Nester, Plasmid 8: 45-54, 1982), a DNA fragment containing the *nptIII* gene from pBI121 and a DNA fragment containing *cos* from pSB11 (Table 3).

[0150] Each primer contains a restriction endonuclease site for later use. The PCR products other than the *trfA1* gene, i.e., the DNA fragment containing *oriV* from pVK102 (884 bp),

- 64 -

the DNA fragment containing oriT (810 bp), the DNA fragment containing the incC1 gene (2118 bp), the DNA fragment containing the nptIII gene from pBI121 (1087 bp), and the DNA fragment containing cos from pSB11 were each cloned into the vector pCR2.1Topo Blunt (from Invitrogen). As a result, the DNA fragment containing oriV was found to contain two nucleotide substitutions and one nucleotide addition as compared with the corresponding nucleotide sequence in a public database (Genbank accession L27758). These mutations were also found in the template plasmid, showing that they were not introduced by PCR but that the template plasmid had a nucleotide sequence different from the sequence in the public database. The nucleotide sequences of oriT, the incC1 gene, and cos were completely identical to those in the database. However, the trfA1 gene could not be cloned alone. Thus, the construction was pursued by the method described below.

[0151] The plasmid into which the DNA fragment containing oriV had been cloned was digested with the restriction endonucleases EcoRI and ClaI, and a 0.9 kb fragment was purified. Similarly, the DNA fragment containing oriT was digested with EcoRI and BglII, and the DNA fragment containing nptIII was digested with NspV and SalI, and the digests were purified. The PCR product of the DNA fragment containing the trfA1 gene was precipitated with ethanol, and then digested with XhoI and BamHI, and purified. These 4 fragments (oriV, the trfA1 gene, nptIII, oriT) were ligated at a time and cloned together. The nucleotide sequence of the resulting plasmid (designated as pVRKT) was analyzed to reveal a

- 65 -

frameshift mutation in the DNA fragment containing the *trfA1* gene as compared with the corresponding nucleotide sequence in the public database (Genbank accession L27758), but the same mutation was also found in pVK102 used as the template, thereby concluding that the mutation was not introduced by PCR and that pVK102 used as the template contained a nucleotide sequence different from the sequence in the public database.

[0152] The resulting plasmid pVRKT containing the 4 fragments were digested with *EcoRI* and *SpeI*, and the DNA fragment containing the *incC1* gene recovered by digesting the plasmid containing the *incC1* gene with *EcoRI* and *XbaI* was inserted into it. The resulting plasmid was further digested with *EcoRI* and *BglII*, and the DNA fragment containing *cos* recovered by digesting the plasmid containing the DNA fragment containing *cos* with *MunI* and *BamHI* was inserted into it to generate the low-copy vector backbone p6FRG (about 8.5 kb) consisting of the 6 fragments, i.e., the DNA fragment containing *oriV*, the DNA fragment containing the *trfA1* gene, the DNA fragment containing *nptIII*, the DNA fragment containing *oriT*, the DNA fragment containing the *incC1* gene, and the DNA fragment containing *cos* (SEQ ID NO: 1 in the Sequence Listing). The foregoing cloning procedure was summarized in a schematic diagram shown in Figure 5. The T-DNA region (*SspI*-*SpeI* fragment) of pSB200PcHm was inserted into the *PvuII*, *NheI* sites of the p6FRG plasmid to generate the vector pLC40 (Figure 6, SEQ ID NO: 2 in the Sequence Listing).

[0153]

- 66 -

Table 3

Primer Name	Sequence	Target gene	Length
121KIII5'NspV	5'- TCg TTC gAA TCg ATA CTA TgT TAT ACg CCA AC-3'	nptIII	32 mer
121KIII3'Sall	5'- ATC gTC gAC TgC ACg AAT ACC AgC gAC CC-3'		29 mer
COS5'BmRV	5'- ggg ggA TCC TTC CAT TgT TCA TTC CAC ggA C-3'	cos	31 mer
COS3'MunFW	5'- ggg CAA TTg ACA TgA ggT TgC CCC gTA TTC -3'		30 mer
OriV3'ClaFW	5'- gAT ATC gAT AgC gTg gAC TCA Agg CTC TC-3'	oriV	29 mer
OriV5'PvNhEc	5'- AAA gAA TTC gCT AgC CAg CTg gCg CTg CCA TTT TTg ggg Tg-3'		41 mer
R5'XhoIRV	5'- AAA CTC gAg CAg CCg AgA ACA TTg gTT CC-3'	trfA1	29 mer
10 R3'BmEcFW	5'- TAg gAA TTC ggA TCC AAA ACA ACT gTC AAA gCg CAC-3'		36 mer
OriT5'BglRV	5'- CgT AgA TCT ggC gCT Cgg TCT TgC CTT g-3'	oriT	28 mer
OriT3'SpEcFW	5'- TgT gAA TTC ACT AgT gAT ATT CCA CAA AAC AgC Agg g-3'		37 mer
InC5'XbRV	5'- CCg TCT AgA TTC gAg CCA Cgg Tag Cgg C-3'	incC1	28 mer
InC3'BgEcFW	5'- CTT gAA TTC AgA TCT TCT Cgg Cgg CgA TCA CgA C-3'		34 mer

SEQ ID NOs: 31-42 in order from the top.

3) Construction of other pLC series cosmid vectors

pLC40GWH

Of the two BalI sites in the backbone of pSB200PcHmGWH, the one on the left side of the RB is not cleaved because it is methylated. Thus, pSB200PcHmGWH was used in the experiments below after it was once transferred into the E.coli strain GM48 to demethylate that site. pSB200PcHmGWH was treated with BalI and SpeI to excise a region containing the T-DNA, which was cloned into the PvuII, NheI sites of 6FRG described above to generate pLC40GWH (Figure 7, SEQ ID NO: 3 in the Sequence Listing). This differs from pLC40 by insertions of attB1, 2 sequences and a deletion of a 317 bp

- 67 -

SspI-BalI region upstream of the RB.

[0154] pLC40bar, pLC40GWB, pLC40GWBSW

pSB25UNpHm and pSB200PcHmGWB were digested with the restriction endonucleases SpeI and SspI, and fragments containing the T-DNA were recovered. The fragment was cloned into the PvuII, NheI sites of p6FRG to generate pLC40bar (Figure 8, SEQ ID NO: 4 in the Sequence Listing) and pLC40GWB (Figure 9, SEQ ID NO: 5 in the Sequence Listing), respectively. pSB200PcHmGWB was treated with NspV, blunt-ended, and dephosphorylated. A pSwaI linker (Table 4) was inserted into this site (pSB200PcHmGWBSW). This plasmid was digested with the restriction endonucleases SpeI and SspI, and a fragment containing the T-DNA was recovered. The fragment was cloned into the PvuII, NheI sites of p6FRG to generate pLC40GWBSW.

[0155] pLC40:35S-IGUS, pLC40GWB:35S-IGUS

The vector pSB24 (Komari et al. 1996) was treated with the restriction endonucleases HindIII and EcoRI to excise a DNA fragment consisting of 35S promoter-I-GUS gene-NOS terminator. This fragment was further blunt-ended by Klenow treatment, and then a 3.1 kb fragment was purified and recovered. The cosmid vector pLC40 described above was treated with the restriction endonuclease NspV, blunt-ended with Klenow enzyme, and then dephosphorylated and purified. On the other hand, pLC40GWBSW was treated with the restriction endonuclease SwaI, dephosphorylated and then gel-purified. The DNA fragment containing the GUS gene described above was inserted into these vectors to prepare pLC40:35S-IGUS and

- 68 -

pLC40GWB:35S-IGUS, respectively.

[0156] pLC40GWHKorB

The cloned region of IncC1 in the pLC vector was extended, and the vector pLC40GWHKorB containing the korB gene was constructed. IncC3'BgEcFw (described above) and IncC/KorB-Xba#1 (Table 4) were designed as primers for amplifying a DNA fragment containing IncC1-KorB of the IncP-based plasmid pVK102. Each primer contains a restriction endonuclease site for later use. A PCR reaction was performed as follows. In 50 µl of a reaction solution containing 500 ng of the pVK102 plasmid DNA, 5 µl of 10 x Pyrobest Buffer II, 4 µl of 2.5 mM each dNTP, 50 pmoles of the primers, and 0.5 µl of Pyrobest DNA Polymerase (from Takara), one cycle of 96°C for 3 minutes, and 10 cycles of 96°C for 1 minute, 55°C for 1 minute, and 72°C for 2 minutes and 30 seconds were performed by using Mastercycler gradient (epENDORF). The resulting amplified PCR product of IncC1-korB (3065 bp) was cloned into the vector pCR2.1Topo Blunt (from Invitrogen). Ligation reactions were performed following the instructions attached to the vector kit. The DNA was transferred into E.coli DH5a by electroporation, and incubated overnight at 37°C on a 2 x YT agar plate containing the antibiotic Zeocin (25 µg/ml). Colony direct PCR was performed to select candidate clones by using grown colonies as templates along with the same primer set as used for the amplification of IncC1-KorB. PCR conditions included one cycle of 96°C for 3 minutes, and 30 cycles of 96°C for 1 minutes, 55°C for 1 minute, and 72°C for 2 minutes and 30 seconds using Wastercycler gradient in a

- 69 -

suspension of the colony in 20 µl of a reaction solution containing 2 µl of 10 x Extaq Buffer, 1.6 µl of 2.5 mM each dNTP, 5 pmoles of the primers, and 0.4 µl of Extaq DNA Polymerase (from Takara). The resulting PCR amplified products of about 3 kb were selected as candidate clones. The nucleotide sequences of these clones were determined by ABI PRISM Fluorescent Sequencer (Model 3100 Genetic Analyzer, from Applied Biosystems). As a result, the nucleotide sequence of IncC1-KorB was completely identical to the sequence in the
10 database.

[0157] Then, the plasmid pVRKT described above was digested with EcoRI and SpeI, and the IncC1-KorB fragment recovered by digesting the plasmid containing IncC1-KorB with EcoRI and XbaI was inserted into it. The resulting plasmid was further digested with EcoRI and BglII, and the cos fragment (MunI-BamHI fragment) described above was inserted into it to generate the plasmid p6FRG2 consisting of the 6 fragments, i.e., oriV, trfA1, nptIII, oriT, IncC1-KorB and cos. The T-DNA region (BamI-SpeI fragment) from pSB3342 (pSB200PcHm)
20 was inserted into the PvuII, NheI sites of the p6FRG2 plasmid to generate the vector pLC40GWHKorB (Figure 10, SEQ ID NO: 65).

[0158] pLC40GWHKorBPI

In order that the cloned large genomic fragment could be excised in its intact form, a recognition site for the homing endonuclease PI-SceI was added upstream of the multicloning site. pLC40GWHKorB was digested with HindIII, and PI-SceI adapters (PI-SceIFw, PI-SceIRv, Table 4) were inserted to generate pLC40GWHKorBPI.

- 70 -

[0159] pLC40GWHKorBPIattB3

In order that the promoter of the selectable marker of pLC40GWH could be changed by any other one by a Gateway system, an attB3 site was added upstream of the ubiquitin promoter. pLC40GWHKorBPI was digested with I-SceI and attB3 adapters (attB3Fw, attB3Rv, Table 4) were inserted to prepare pLC40GWHKorBPIattB3.

[0160] pLCleo

In order that an NotI-digested genomic fragment could be cloned, a recognition site for PspOMI (producing the same sticky end as that of NotI) was formed at the multicloning site and simultaneously the recognition site of ApaI (a neoschizomer of PspOMI) in the ubiquitin intron was abolished. pLC40GWHKorBPIattB3 was digested with ApaI and NheI, and ApaIm-NheI adapters (ApaIm-NheIFw, ApaIm-NheIRv, Table 4) were inserted to prepare pLC40GWHKorBPIattB3ApaIm. This plasmid was digested with HindIII and NspV, and HindIII-PspOMI-NspV adapters (HindIII-PspOMI-NspVFw, HindIII-PspOMI-NspVRv, Table 4) were inserted to finally prepare pLCleo (Figure 11, SEQ ID NO: 66 in the Sequence Listing).

[0161]

Table 4

Primer/Adapter name	Sequence(5'-3')	Length
IncC/KorB-Xba#1	CGG TCT AGA GTG CGC AGC AGC TCG TTA TC	29mer
PI-SceIFw	AGC TAT CTA TGT CGG GTG CGG AGA AAG AGG TAA TGA AAT GGC A	43mer
PI-SceIRv	AGC TTG CCA TTT CAT TAC CTC TTT CTC CGC ACC CGA CAT AGA T	43mer
attB3Fw	CAG GGT AAT CAA CTT TGT ATA ATA AAG TTG ATA A	34mer
attB3Rv	CAA CTT TAT TAT ACA AAG TTG ATT ACC CTG TTA T	34mer
Apalm-NheIFw		
	GGGTAGTTCTACTTCTGTTCATGTTTGTGTTAGATCCGTGTTTGTGTTAGATCCGTGCTG	60mer
Apalm-NheIRv		
	CTAGCGCCGGATCTAACACAAACACGGATCTAACACAAACATGAACAGAAGTAGAACTACCCGGCC	66mer
HindIII-PspOMI-NspVFw	AGC TTG GGC CCT T	13mer
HindIII-PspOMI-NspVRv	AGG GCC CA	8mer

SEQ ID NOs: 68-76 in order from the top.

p6FRGSwKp

p6FRG was treated with PvuII and dephosphorylated. The adapter SwaIKpnIRV, SwaIKpnIFW (Table 5) DNAs having recognition sites for SwaI and KpnI were annealed. A part of this was phosphorylated with PNK (Amersham). This SwaI-KpnI linker was inserted into the PvuII site of p6FRG to generate p6FRGSwKp. The KpnI site was designed for cloning a DNA fragment containing the virB gene and the virG gene derived from the Agrobacterium strain A281 in the next step, and the SwaI site was designed for cloning the T-DNA in the step after next.

[0162]

Table 5

Linker/Adapter name	Sequence	Length
pSwaI linker	5'-cca tt t aa tgg-3'	12 mer
SwaIKpnIRV	5'- cca tt t aa tgg tac cgg-3'	18 mer
SwaIKpnIFW	5'-ccg gta cca tt t aa tgg -3'	18 mer

SEQ ID NOs: 43-45 in order from the top.

p6FRGSVR, p6FRGSVRF

The vector pSB1 (Komari et al. 1996) was digested with KpnI, and a 14.8 kb DNA fragment containing the virB gene and the virG gene was recovered. This fragment was inserted into the KpnI-treated and dephosphorylated vector p6FRGSwKp, thereby generating p6FRGSVR and p6FRGSVF.

[0163] pLCSBGWBSW

pSB200PcHmGWBSW was digested with SpeI and SspI, and a DNA fragment containing the T-DNA region was blunt-ended with Klenow enzyme. This fragment was inserted into the SwaI-digested and dephosphorylated vector p6FRGSVR, thereby generating pLCSBGWBSW (Figure 12, SEQ ID NO: 6 in the Sequence Listing). This vector is a low-copy vector having a full length of about 28 kb, which contains the virB gene and the virG gene derived from the Agrobacterium strain A281 so that it may be used for transformation of maize. It also contains a cos site, which allows easy cloning of about 10-20 kb of DNA by a packaging reaction.

[0164] 4) A pLC vector containing virG

The vector pLC40GWHvG containing the virG gene in the pLC40GWH vector was constructed by the procedure described

- 73 -

below as a means for improving the efficiency of plant transformation with a pLC40 series cosmid vector.

[0165] Preparation of the virG gene

The primers virGProSm and virGTerSm for amplifying the virG gene (including its promoter, the structural gene and the 3' region) were designed and synthesized. These primers, and pTOK47 (Jin et al. 1987 J Bacteriol 169: 4417-4425) as a template DNA, were used to amplify the virG gene by PCR. As a result, the PCR product of about 1 kb was amplified. A part of the product was cloned into the vector pCR2.1Topo (from Invitrogen) in the same manner as described above, and the nucleotide sequence was determined. The DNA sequence of the VirG gene contains an NspV site. This restriction site will be used as a cloning site in a future vector. Thus, this site was removed by PCR mutagenesis. The first adenine in the NspV site (ttcgaa) was changed to guanine (ttcgga) to design and synthesize the primer virGonNspVRV and its complementary sequence virGonNspVFW. PCR was performed with two primer sets, i.e., one consisting of VirGonNspVFW and the primer virGProSpe placed upstream of the virG gene promoter and the other consisting of virGonNspVRV and the primer virGTerSpe placed downstream of the virG gene terminator. The virG gene cloned into pCR2.1Topo was used as a template. As a result, the product of about 400 bp and the product of about 600 bp were amplified by the former and latter sets, respectively. These products were purified and used as templates for the next PCR reaction. A PCR reaction was performed with the purified two PCR products as templates and the previous

- 74 -

primers virGProSpe and virGTerSpe. As a result, the PCR product of about 1 kb was amplified. The PCR product was cloned into the pCR2.1Topo vector, and the nucleotide sequence was determined to confirm the mutation (ttcgaa→ttcgga).

[0166] Similarly, the unmutated virG gene was amplified by PCR with virGProSpe and virGTerSpe, and cloned into pCR2.1Topo, and the nucleotide sequence was determined. The primers used in PCR are summarized in Table 6.

[0167]

Table 6

Designation	Sequence 5'-3'	Length
virGProSm	TCA ATA CCC ggg gTA ACC TCg AAg CgT TTC AC	32mer
virGTerSm	Tgg TgA CCC ggg ACC TAT Cgg AAC CCC TCA C	31mer
virGProSpe	TCA ATA ACT AgT gTA ACC TCg AAg CgT TTC AC	32mer
virGTerSpe	Tgg TgA ACT AgT ACC TAT Cgg AAC CCC TCA C	31mer
virGonNspVRV	CTT gAg ATC gTT Cgg AAT CTg	21mer
virGonNspVFW	CAG ATT CCg AAC gAT CTC AAg	21mer

SEQ ID NOs: 46-51 in order from the top.

pLC40GWHvG1, pLC40GWHvGC1

The vector pLC40GWH was digested with the restriction endonuclease PvuII, and dephosphorylated. An SpeI linker (GACTAGTC, from Takara) was inserted to prepare pLC40GWHSpe. This plasmid was digested with the restriction endonuclease SpeI and dephosphorylated. A fragment of about 1 kb of the mutated virG gene excised with SpeI from the vector was inserted into this plasmid to prepare pLC40GWHvG1 (Figure 13, SEQ ID NO: 7 in the Sequence Listing). Similarly, the

- 75 -

unmutated *virG* gene was inserted into pLC40GWHSpe to prepare pLC40GWHvGC1.

[0168] pLC40GWHvG1:35S-IGUS, pLC40GWHvGC1:35S-IGUS

In the same manner as described above, the vector pSB24 (Komari et al. 1996) was treated with the restriction endonucleases HindIII and EcoRI to excise a DNA fragment containing the GUS gene, which was cloned into a vector having a multicloning site SgfI-HindIII-EcoRI-SgfI. The resulting plasmid was digested with SgfI to recover the DNA fragment containing the GUS gene. At this point, both ends of the DNA fragment containing the GUS gene are SgfI sites. The cosmid vector pLC40GWHvG1 described above was treated with the restriction endonuclease PacI and dephosphorylated. The 3.1 kb SgfI fragment (the DNA fragment containing the GUS gene) was cloned into it to generate pLC40GWHvG1:35S-IGUS. Similarly, 35S-IGUS-NOS was introduced into pLC40GWHvGC1 to prepare pLC40GWHvGC1:35S-IGUS.

[0169] 5) *virG*-containing vectors capable of coexisting with pLC

pVGW

pTOK47 is a large IncW plasmid of about 28 kb containing *virG* and *virB* (Jin et al. 1987 J Bacteriol 169: 4417-4425). Thus, a smaller vector capable of coexisting with a pLC vector and containing the origin of replication IncW ori, the *virG* gene, and a selectable marker gene (designated as pVGW) was designed and constructed.

[0170] The primers pSa5'Ect22 and pSa3'BglII for amplifying a fragment containing IncW ori from pTOK47 (Jin et al. 1987 J

- 76 -

Bacteriol 169: 4417-4425), and the primers Gm5'Bm and Gm3'Xh-2nd for amplifying the gentamycin resistance gene (gentamycin acetyltransferase) from pPH1JI (Hirsch and Beringer 1984 Plasmid 12: 139-141) were designed (Table 7). Each primer contains a restriction endonuclease site for later use. pTOK47 and pPH1JI were used as templates, respectively. Pyrobest DNA Polymerase (from TaKaRa) was used to perform PCR. As a result, a DNA fragment of about 2.7 kb containing IncW ori and a DNA fragment of about 0.7 kb corresponding to the gentamycin resistance gene were amplified.

[0171] On the other hand, the primer virGN54DFW for changing the amino acid residue at position 54 of virG derived from pTOK47 from N to D by PCR mutagenesis (virGN54D, Hansen et al. 1994 Proc. Natl. Acad. Sci. USA 91: 7603-7607), and its complementary sequence virGN54DRV were designed. PCR was performed with two primer sets, i.e., one consisting of virGN54DFW and the primer virGProSal placed on the 5' of the virG gene promoter and the other consisting of virGN54DRV and the primer virGTerPst placed on the 3' of the virG gene terminator (Table 7). The pTOK47 plasmid was used as a template. As a result, the product of about 0.4 kb and the product of about 0.7 kb were amplified by the former and latter sets, respectively. These products were purified and used as templates along with the previous primers virGProSal and virGTerPst to further perform a PCR reaction. As a result, the product (virGN54D) of about 1.1 kb was amplified.

[0172] The PCR products of the fragment containing IncW ori, the gentamycin resistance gene, and virGN54D were cloned

- 77 -

into the pCR-Blunt II-TOPO vector (Invitrogen). The nucleotide sequence was determined and compared with a publicly available sequence (Genbank/EMBL Accession Number: U30471) to reveal a deletion of 6 nucleotides in the fragment containing IncW ori, which was also found in pTOK47 used as a template. However, the nucleotide sequence of the gentamycin resistance gene was completely identical to the sequence in the database. virGN54D was found to contain the mutation at the desired site.

[0173] The plasmid into which the fragment containing IncW ori had been cloned was digested with EcoT22I and BglII, and a 2.7 kb fragment was recovered. Similarly, the gentamycin resistance gene was digested with BamHI and XhoI, and virGN54D was digested with SalI and PstI, and each fragment was purified. These three fragments were ligated together (BglII and BamHI, XhoI and SalI, and PstI and EcoT22I produce the same sticky ends) to generate pVGW (Figure 14, SEQ ID NO: 8 in the Sequence Listing).

[0174]

Table 7

Designation	Sequence	Length
pSa5'EcT22	5'-aaa atg cat ggc atg ttt aac aga atc tg-3'	29mer
pSa3'BglII	5'-ttt aga tct act cgt tcg cgg agc tgg-3'	27mer
Gm5'Bm	5'-aaa gga tcc ttc atg gct tgt tat gac tg-3'	29mer
Gm3'Xh-2 nd	5'-tgc ctc gag aca att tac cga aca act ccg-3'	30mer
virGN54DFW	5'-cga cct aaa tct aga tca aca ac-3'	23mer
viGN54DRV	5'-gtt gtt gat cta gat tta ggt cg-3'	23mer
virGProSal	5'-ttt gtc gac cat agg cga tct cct taa tc-3'	29mer
virGTerPst	5'-aaa ctg cag gtg aag agg gac cta tcg g-3'	28mer

SEQ ID NOs: 52-59 in order from the top.

pVGW2

To further increase the convenience of pVGW, the promoter region of the gentamycin resistance gene was extended and additional cloning sites were added to construct the vector pVGW2. The primers BamSmaGmPro and NheI siteGmRv for amplifying the gentamycin resistance gene of the plasmid pPH1JI, and the primers 'MscI site-virG5' Fw (for these primers, see Table 8) and pSa3'BglII (described above) for amplifying the virG-IncW region of pVGW were designed. Each primer contains a restriction endonuclease site. A PCR reaction was performed as follows. One cycle of 98°C for 30 seconds and 35 cycles of 98°C for 10 seconds, 55°C for 5 seconds, and 72°C for 1 minute were performed using Mastercycler gradient (epENDORF) in 50 µl of a reaction solution containing 1 ng of the template plasmid DNA, 25 µl of 2 x PrimeSTAR Max Premix (from Takara), and 15 pmoles of the

- 79 -

primers. As a result, the PCR products of the gentamycin resistance gene (826 bp) and the virG-IncW region (3840 bp) were amplified. The gentamycin resistance gene was cloned into the vector pCR-Blunt II-TOPO (from Invitrogen), and transferred into E.coli TOP10 (Invitrogen) by electroporation. The cells were incubated on an LB agar plate containing the antibiotics kanamycin (50 µg/ml) at 37°C overnight, and a plasmid was purified from the resulting colony. The nucleotide sequences of these clones (pCR-Gm) were determined by ABI PRISM Fluorescent Sequencer (Model 3100 Genetic Analyzer, from Applied Biosystems) to confirm that no mutation had been introduced by PCR error. The plasmid pCR-Gm was digested with BamHI and PvuII to recover the Gm fragment, which was ligated to the virG-IncW fragment digested with BglII (having a BglII site at one end and a blunt end at the other). The resulting clone was transferred into E.coli TOP10 by electroporation, and selected on an LB agar plate containing the antibiotics gentamycin (30 µg/ml). A plasmid was purified from the resulting colony and confirmed by the sequencer to contain no PCR error, thereby generating pVGW2 (Figure 15, SEQ ID NO: 67 in the Sequence Listing).

[0175]

Table 8

Designation	Sequence	Length
BamSmaGmPro	5'-AAA GGA TCC CGG GTT GAC ATA AGC CTG TTC GGT TCG-3'	36mer
NhelsiteGmRv	5'-AAA GCT AGC AAT TTA CCG AAC AAC TCC GCG G-3'	31mer
MscIsite-virG5'Fw	5'-AAA TGG CCA TAG GCG ATC TCC TTA ATC AAT-3'	30mer

SEQ ID NOs: 77-79 in order from the top.

Example 2: Cloning of large fragments by pLC vectors

The present example describes examples of libraries of *Arabidopsis thaliana* (ecotype: colombia), wild species of rice (*Oryza rufipogon*), Sudan grass (*Sorghum sudanense*), an extremely early maturing variety of Italian millet (*Setaria italica*), teosinte (*Zea diploperennis*), pearl millet (*Pennisetum typhoideum*), Bahia grass (*Paspalum notatum* Flugge) and sugar cane (*Saccharum officinarum*) prepared with pLC40, pLC40GWH, pLCleo, pLC40GWHvG1, pSB200, pSB200PcHmGW, or pSB25U.

[0176] 1) Preparation of genomic DNA

About 5 g of young leaves of each plant at about one month after seeding grown in a greenhouse was ground in a mortar under liquid nitrogen, and then the genomic DNA was purified by the CTAB method. The yield was about 500-600 µg expressed as DNA. The genomic DNA was partially digested with 0.02-0.06 U/µg of TaqI enzyme. After the partial digestion, fractions containing a genomic DNA fragment of 30-45 kb were recovered by 10-40% sucrose density gradient centrifugation.

[0177] 2) Preparation of the vectors

The cosmid vectors pLC40, pLC40GWH, pLCleo, pLC40GWHvG1,

- 81 -

pSB200, pSB200PcHmGW, and pSB25U were completely digested with the restriction endonuclease NspV (TOYOBO) and dephosphorylated, and then purified.

[0178] 3) Cloning by a packaging reaction

The vectors prepared as described above were ligated to the genomic DNA fragments, followed by a packaging reaction using GigaPack III XL Packaging extract at room temperature for 2 hours. After the reaction, the clones were incubated with E.coli GeneHogs (Invitrogen). As a result, libraries of 10 10,000-100,000 cfu (colony-forming-unit) were prepared from all of the combinations of the plant species and vectors (Table 9), as shown in Table 7.

[0179]

Table 9

Plant species	Vector	Library cfu
Arabidopsis thaliana	pLC40	ca 80000
	pSB200PcHmGWH	ca 100000
Oryza rufipogon	pLC40GWH	ca 20000
	pSB200	ca 50000
Extremely early maturing	pLC40GWH	ca 20000
Italian millet	pSB200PcHmGWH	ca 20000
Sugar cane	pLC40GWH	ca 50000
	pLC40GWHvG1	ca 50000
Sudan grass	pLC40GWH	ca 50000
	pSB200PcHmGWH	ca 30000
Pearl millet	pLC40GWH	ca 20000
Teosinte	pLC40GWH	ca 100000
	pSB25UNpHm	ca 20000
Bahia grass	pLCleo	ca 10000

4) Analysis of the cloned genomic DNA fragments

Plasmids were purified from 12-24 clones of each library and cleaved with the restriction endonucleases HindIII and SacI in the multicloning site at each end of the insert, thereby yielding bands corresponding to the vectors (9.2-9.8 kb) in all of the clones analyzed in the case of pSB200, pSB25UNpHm and pSB200PcHmGW as well as bands corresponding to the vectors (13.2-14.2 kb) in all of the clones analyzed in the case of pLC40, pLC40GWH, pLCleo and pLC40GWHvG1. The length of the cloned large fragment is estimated to be in the range of 25 kb-45 kb from the total length of the restriction

- 83 -

fragments of the insert of each clone, with an average of about 40 kb in the case of the pSB vectors and an average of about 35 kb in the case of the pLC vectors. Figure 16 shows an example of teosinte genomic DNA/pLC40GWH.

[0180] Then, the human genome (Human Genomic DNA, Male, from Promega, Catalog No.: G1471) was partially digested with *Taq*I to prepare a 30-40 kb fragment, which was then cloned into the vector pLC40GWH. Plasmid DNA was purified from *E. coli* containing the human genomic fragment from arbitrary 12 clones, and the nucleotide sequences at both ends of the insert were analyzed and searched through a database. Homology searches were performed by BLAST through the database of GenBank at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>). The results showed that 11 of the 12 clones isolated are included in 11 single clones containing the human genomic fragment in the database. Ten clones excluding one containing repeated sequences were analyzed for homology to the human genome sequence in the database, whereby the lengths of the cloned human genomic fragments were estimated to be 28023 bp, 31645 bp, 38265 bp, 39599 bp, 31965 bp, 32631 bp, 34727 bp, 36925 bp, 38794 bp, and 34364 bp. The average length was 34693.8 bp, which agreed well with the value obtained by cloning the plant genomes.

[0181] Then, the nucleotide sequences at both ends of the cloned plant genomic DNA fragments were determined. Homology searches were performed by BLAST on thus obtained sequence data of 300-600 nucleotides through the database of GenBank at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) and the database of

- 84 -

Beijing Genomics Institute

(<http://btn.genomics.org.cn:8080/rice/>). As a result, *Oryza rufipogon* and *Arabidopsis* showed a homology of 87-100% to the genome sequence of rice and *Arabidopsis*, respectively, over the range of at least 100 bp or more. The libraries of the other plant species also showed significant homologies to the sequences of rice, *Arabidopsis*, maize, sorghum, etc.

[0182] Example 3: Transfer into *Agrobacterium* via triparental mating

1) Transfer into *Agrobacterium* via triparental mating and its efficiency

Each vector containing a plant genomic fragment was transferred into *Agrobacterium* via triparental mating as follows.

[0183] i) pLC40 series cosmid vectors

pLC40 series cosmid vectors are resistant to kanamycin (Km) and hygromycin (Hm). GeneHogsTM (Invitrogen) was used as host *E.coli*. pRK2073 (spectinomycin (Sp)-resistant) was used as a helper plasmid for triparental mating. HB101 was used as host *E.coli* for the helper plasmid. The *Agrobacterium* strain LBA4404 (no drug resistance) was used.

[0184] Initially, *E.coli* GeneHogsTM was infected with an appropriate amount of a dilution of a packaging reaction, spread on an LA plate containing Km (50 µg/mL), and incubated at 23°C for 3 days. *E. coli* cells in a colony that appeared were streaked with a toothpick on an LA plate containing Km, and incubated at 28°C for 2 nights. On the other hand, LBA4404 was spread on an AB plate, and incubated at 25°C for

- 85 -

5 days. HB101/pRK2073 was spread on an LA plate containing Sp (50 µg/mL), and incubated at 37°C for 2 nights. The cultures of the three strains, i.e., GeneHogsTM harboring a pLC40 series cosmid vector containing a cloned genomic fragment, LBA4404 and HB101/pRK2073 were mixed on an NA plate and incubated at 28°C overnight. The entire amount of the mixture of the three strains was suspended in 250 µl of sterile water, and 5 µl of the suspension was spread on an AB plate containing Km (50 µg/mL) and Hm (25 µg/mL), and incubated at 28°C for 7 days. The resulting recombinant Agrobacterium was used in plant transformation experiments. This single colony was reincubated on an AB plate containing Km and Hm and a part of the grown colony was spread on an LA with drugs, showing that few E.coli cells have been grown.

[0185] ii) pSB200 series cosmid vectors

pSB200 series cosmid vectors are Sp- and Hm-resistant. GeneHogsTM (Invitrogen) was used as host E.coli. pRK2013 (Km-resistant) was used as a helper plasmid. HB101 was used as host E.coli for the helper plasmid. The Agrobacterium strain LBA4404 harboring pSB1 (tetracycline (Tc) resistance) was used.

[0186] Initially, E.coli GeneHogs was infected with an appropriate amount of a dilution of a packaging reaction, spread on an LA plate containing Sp (50 µg/mL), and incubated at 23°C for 3 days. A colony was picked with a toothpick and streaked on an LA plate containing Sp, and incubated at 28°C for further 2 nights. On the other hand, LBA4404/pSB1 was spread on an AB plate containing Tc (15 µg/mL), and incubated

- 86 -

at 25°C for 5 days. HB101/pRK2013 was spread on an LA plate containing Km (50 µg/mL) and incubated at 37°C for 2 nights. The cultures of the three strains, i.e., GeneHogs™ harboring a pSB200 series cosmid vector containing a cloned genomic fragment, LBA4404/pSB1 and HB101/pRK2013 were mixed on an NA plate and incubated at 28°C overnight. The entire amount of the mixture of the three strains was suspended in 250 µl of sterile water, and 25 µl of the suspension was spread on an AB plate containing Sp (50 µg/mL) and Hm (25 µg/mL), and incubated at 28°C for 7 days. The resulting recombinant Agrobacterium was used in plant transformation experiments.

[0187] iii) pCLD04541

Two genome libraries (the genomes of the rice variety CO39 and Arabidopsis ecotype Colombia, both having an average insert length of 110 kb in host E.coli DH10B) prepared with the vector pCLD04541 provided from Dr. Hongbin Zhang of Texas A&M University were used for triparental mating. The pCLD04541 vector is Km- and Tc-resistant. pRK2073 was used as a helper plasmid, and HB101 was used as host E.coli for the helper plasmid. The Agrobacterium strain LBA4404 was used.

[0188] E.coli harboring each clone of the pCLD04541 libraries was spread on an LA containing Tc (10 µg/mL), and incubated at 28°C for 2 nights. On the other hand, LBA4404 was spread on an AB plate, and incubated at 25°C for 5 days. HB101/pRK2073 was spread on an LA containing Sp (50 µg/mL) and incubated at 37°C for 2 nights. The cultures of the three strains, i.e., DH10B harboring pCLD04541 containing a cloned genomic fragment, LBA4404 and HB101/pRK2073 were mixed on an

- 87 -

NA plate and incubated at 28°C overnight. The entire amount of the mixture of the three strains was suspended in 250 µl of sterile water, and a few microliters of the suspension was spread on an AB plate containing Km (25 µg/mL), and incubated at 28°C for 7 days. The resulting recombinant Agrobacterium was used in plant transformation experiments.

[0189] As described above, genome clones included in the libraries prepared with pLC40 series cosmid vectors, pSB200 series cosmid vectors and the pCLD04541 vector were
10 transferred into Agrobacterium. A summary of these triparental mating systems and the triparental mating efficiencies are shown in Table 10. The triparental mating efficiencies were 97% in pLC series vectors, 79% in pSB series vectors, and 93% in pCLD04541, respectively, showing that the pLC series vectors were the most efficient (Table 10).
[0190]

Table 10

DNA donor plant	Vector	# of clones used for triparental mating (a)	# of clones giving recombinant Agrobacterium (b)	Efficiency(%) b/a
<pLC40 series cosmid vectors>				
Oryza rufipogon	pLC40 GWH	5657	5469	96.7
Arabidopsis thaliana	pLC40	1532	1410	92.0
Sudan grass	pLC40 GWH	2301	2201	95.7
Italian millet	pLC40 GWH	2521	2405	95.4
Teosinte	pLC40 GWH	10739	10593	98.6
Bahia grass	pLCleo	384	383	99.7
Total		23134	22461	97.1
<pSB200 series cosmid vectors >				
Oryza rufipogon	pSB200	10375	7504	72.3
Arabidopsis thaliana	pSB200PcHmGWH	1332	1179	88.5
Sudan grass	pSB200PcHmGWH	2096	2031	96.9
Italian millet	pSB200PcHmGWH	2336	2032	87.0
Total		16139	12746	79.0
<pCLD04541>				
Indica riceCO39	pCLD04541	149	127	85.2
Arabidopsis thaliana	pCLD04541	192	190	99.0
Total		341	317	93.0

2) Stability of genomic DNA

To analyze whether or not the genomic DNA fragment carried on each clone has been transferred to Agrobacterium, Southern hybridization was performed using the entire genomic DNA fragment as a probe. Plasmid DNAs were conventionally

- 89 -

extracted from E.coli and Agrobacterium, and digested with the restriction endonucleases HindIII and SacI. Then, a part of the digests were fractionated by agarose gel electrophoresis, and transferred to the nylon membrane filter HybondN+. Then, a part of the HindIII and SacI digest (precipitated with ethanol and redissolved in TE) of the E.coli-derived plasmid was labeled with an ECL labelling kit (Amersham) and hybridized to this membrane as a probe. Hybridization, washing and signal detection were performed following the instructions attached to the ECL kit. All of four plasmids containing a rufipogon fragment cloned into pLC40GWH showed the transfer of the genomic DNA fragment from E.coli to Agrobacterium.

[0191] Example 4: Transformation of large fragments into rice with pLC vectors

1) Rice transformation and its efficiency

i) Method for rice transformation

Immature embryos of the rice variety Yukihihikari were infected with Agrobacterium. Rice transformation was performed by the method described in the Japanese Patent Application No. 2003-293125, except that all of the aseptically dissected immature embryos were centrifuged as a pretreatment before Agrobacterium inoculation. Specifically, the immature embryos were centrifuged in an eppendorf tube containing 1 ml of sterile water at 20000 x g for 10 minutes (25°C). Hygromycin B was used as a selective drug and added at 50 mg/l each in the selective medium, regeneration medium and rooting medium. In the case of pLC40 series cosmid

- 90 -

vectors and pSB200 series cosmid vectors, one immature embryo was inoculated with one *Agrobacterium* strain (one type of DNA fragment). In the case of pCLD04541, however, two immature embryos were inoculated with one *Agrobacterium* strain (one type of DNA fragment). Paromomycin was used as a selective drug and added at a concentration of 400-800 mg/l in the selective medium, regeneration medium and rooting medium.

[0192] ii) Transfer of plant genomic fragments into rice

The results of transformation are shown in Table 11. In the case of pSB200 series cosmid vectors, hygromycin-resistant individuals were obtained from 59.1%-62.7% of the strain. In contrast, pLC40 series cosmid vectors gave the transformants from 86.6%-95.4% of the strain. In all of the three donor plants of genomic DNA (*Oryza rufipogon*, Sudan grass, and an extremely early maturing variety of Italian millet), the efficiency was 24%-36% higher when pLC40 series cosmid vectors were used. In the case of the pCLD04541 vector, however, the efficiency was as low as 41-53.4%. These results suggested that pLC40 series cosmid vectors allow transfer of genomic DNA fragments into rice more efficiently than pSB200 series cosmid vectors and pLCD04541.

[0193] To evaluate the transformation efficiency of normal size gene expression units, vectors were tested by comparison in the transformation with a DNA fragment containing the GUS gene. When 25 Yukihihikari immature embryos were used for each vector, pSB134 (WO2005/017169) gave an average of 11.7 hygromycin-resistant regenerated individuals per immature embryo while pLC40:35S-IGUS gave an average of 11.5

- 91 -

regenerated individuals.

[0194]

Table 11

Results of the transformation of randomized plant genomic fragments into rice using a pSB or pLC vector

Genome donor plant	Vector	# of genomic fragments used for Agrobacterium-mediated transformation (A)	# of genomic fragments that regenerated hygromycin-resistant individuals (B)	B/A (%)
Oryza rufipogon	pSB200	2246	1327	59.1
Oryza rufipogon	pLC40GWH	2271	2166	95.4
Sudan grass	pSB200PcHmGWH	1997	1252	62.7
Sudan grass	pLC40GWH	1760	1524	86.6
Italian millet	pSB200PcHmGWH	1940	1200	61.9
Italian millet	pLC40GWH	2285	1986	86.9
Bahia grass	pLCleo	18	16	88.9
Indica rice CO39	pCLD04541	156	64	41.0
Arabidopsis thaliana	pCLD04541	189	101	53.4

2) Verification of the transfer of large fragments

i) PCR of flanking regions of fragments

Genomic DNAs were extracted from 11 transformants and young leaves of Yukihihikari by the method described above. PCR was performed on these DNAs with 2 sets of the primers shown in the table below. pSB200-9531F and pSB200-4R are primers

- 92 -

for amplifying a 139 bp region from the RB to the genomic DNA fragment. HPTinRV and HPTinFW are primers for amplifying an internal region of the hygromycin resistance gene (Table 12). Thirty-five cycles of PCR were performed. As a result, the products were amplified with HPTinRV and HPTinFW in all of the 11 transformants, while no PCR product was obtained with either primer set in the control Yukihihikari. When pSB200-9531F and pSB200-4R were used, PCR products were obtained in 10 of the 11 individuals. These results show that the flanking regions of the genomic DNA fragments were transferred into most of the plants transformed with pLC vectors, thus verifying the transfer of the genomic DNA fragments.

[0195]

Table 12

Designation	Sequence	Length
pSB200-9531F	5'- ctg aag gcg gga aac gac aat ctg -3'	24mer
pSB200-4R	5'- gct tgc tga gtg gct cct tca acg -3'	24mer
pSB200-170R	5'- aac tgc act tca aac aag tgt gac-3'	24mer
HPTinRV	5'-tat gtc ctg cgg gta aat ag-3'	20mer
HPTinFW	5'- ttg ttg gag ccg aaa tcc g-3'	19mer

SEQ ID NOs: 60-64 in order from the top.

ii) PCR of both terminal and internal sequences of fragments

For each of three *Oryza rufipogon* fragments (called A, B, and C) used for the transformation into Yukihihikari with the pLC40GWH vector, two individuals of T0 plant were analyzed by PCR to determine whether or not both ends and the center

- 93 -

region of each fragment had been introduced. PCR conditions included a treatment at 94°C for 2 minutes, followed by 35 cycles of thermal denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds and extension at 60°C for 30 seconds, and finally a treatment at 72°C for 2 minutes.

[0196] To detect the RB side of fragment A, PCR (PCR1) was performed with pSB200-9531F and a primer specific to fragment A (5'-ggt aat ttc ttg tga tcg aag gac-3' (SEQ ID NO: 11)). To detect the center region of fragment A, a PCR assay was performed by the CAPS method (Konieczny and Ausubel 1993 Plant Journal 4: 403-410) using nucleotide sequence polymorphisms found between the sequence of Nipponbare AP004667 corresponding to fragment A (identified by database searches) and the sequence of Oryza rufipogon. Specifically, PCR (PCR2) was performed with two primers (5'-ggg att ctt tat gct ggg ttt agg-3' (SEQ ID NO: 12) and 5'-gca agc aat acc tct gtt atg ctg-3' (SEQ ID NO: 13)), and the product was digested with SspI. To detect the HPT side, PCR (PCR3) was performed with pSB200-170R and a primer specific to fragment A (5'-ggt ttc aga tgg cga cct cag ctt tg-3' (SEQ ID NO: 14)).

[0197] Similar marker assays were performed on fragment B and fragment C. Thus, to detect the RB side of fragment B, PCR was performed with pSB200-9531F and a primer specific to fragment B (5'-cag gtg gct tta ttc ctc ctc tca-3' (SEQ ID NO: 15)). To detect the center region of fragment B, a PCR assay was performed by the CAPS method using nucleotide sequence polymorphisms found between the sequence of Nipponbare AP005967 corresponding to fragment B (identified by database

- 94 -

searches) and the sequence of *Oryza rufipogon*. Specifically, PCR was performed with two primers (5'-ccg aaa gtt cgt ggg caa tgc cta-3' (SEQ ID NO: 16) and 5'-gcc atc ctt agc ata tga gtg gca-3' (SEQ ID NO: 17)), and the product was digested with HaeIII. To detect the HPT side of fragment B, PCR was performed with pSB200-170R and a primer specific to fragment B (5'-ggc tat tta cgt ggc atg tta cgt-3' (SEQ ID NO: 18)). To detect the RB side of fragment C, PCR was performed with pSB200-9531F and a primer specific to fragment C (5'-tcg taa gtc tac ttc cct tta cga-3' (SEQ ID NO: 19)). To detect the center region of fragment C, a PCR assay was performed by the CAPS method using nucleotide sequence polymorphisms found between the sequence of Nipponbare AL713907 corresponding to fragment C (identified by database searches) and the sequence of *Oryza rufipogon*. Specifically, PCR was performed with two primers (5'-cca aac cac atc ctt ata gtg tgc-3' (SEQ ID NO: 20) and 5'-cct cat tgc atg cgg tca cta c-3' (SEQ ID NO: 21)), and the product was digested with HaeIII. To detect the HPT side of fragment C, PCR was performed with pSB200-170R and a primer specific to fragment C (5'-gca ggg tat taa tcg atc aac acc-3' (SEQ ID NO: 22)).

[0198] Analytical results of fragment B are shown in Figure 17, and analytical results of fragments A-C are summarized in Table 13. Of the two transformants tested for fragment A, no individual containing the entire large fragment was obtained but an individual containing the center region and one end, or both ends was obtained. However, one of the two individuals tested for fragment B and fragment C was shown to contain the

- 95 -

entire *Oryza rufipogon* fragment, i.e., both ends and the center region. These results verified that plant genomic fragments of 25-40 kb in size can be transferred into plants by pLC vectors.

[0199]

Table 13

Fragment	T0 plant	RB side	center	HPT side
A	1	-	+	+
	2	+	-	+
B	1	+	+	-
	2	+	+	+
C	1	-	-	-
	2	+	+	+

+: An *oryza rufipogon* fragment was detected.

-: An *oryza rufipogon* fragment was not detected.

Example 5: Transformation of maize with pLC40 series cosmid vectors

1) Combination of pLC with pTOK47 or pLC with pVGW

A vector containing a vir gene is required for maize transformation to increase the transformation efficiency (Ishida et al. 1996 Nat Biotechnol 14:745-50) because the efficiency with ordinary binary vectors is very low except for special methods (Frame et al. (2002) Plant Physiol 129: 13-22). pLC40 series cosmid vectors are ordinary binary vectors so that they should be modified by using a vir gene to improve the transformation efficiency. Thus, the vector pTOK47

- 96 -

capable of coexisting with pLC40 series cosmid vectors (IncP plasmids) in bacteria and expressing a vir gene (Jin et al. 1987 J Bacteriol 169: 4417-4425), and a vector newly constructed by the present invention, pVGW were initially used in combination with pLC. pTOK47 is an IncW plasmid carrying a DNA fragment (KpnI 14.8 kb fragment) containing the virB gene and the virG gene derived from the Agrobacterium strain A281, and capable of coexisting with IncP plasmids. pVGW is a plasmid containing a variant virG (virGN54D) and IncW ori.

[0200] pTOK47 (tetracycline-resistant) was transferred into the Agrobacterium LBA4404 or EHA105 (a kind gift from Dr. Stanton Gelvin of Purdue University) via triparental mating. A plasmid was extracted from this Agrobacterium and confirmed by restriction endonuclease analysis to contain pTOK47. Further, pLC40:35S-IGUS or pLC40GWB:35S-IGUS was introduced into the resulting LBA4404/pTOK47 or EHA105/pTOK47 (Tc-resistant) via triparental mating. These Agrobacteria are described as LBA4404/pTOK47/pLC40:35S-IGUS, LBA4404/pTOK47/pLC40GWB:35S-IGUS, EHA105/pTOK47/pLC40:35S-IGUS, and EHA105/pTOK47/pLC40GWB:35S-IGUS. Plasmid DNAs were extracted from the Agrobacteria and analyzed by PCR to confirm the presence of the VirG, RB, hpt or bar, and GUS genes.

[0201] In the same manner, pVGW was transferred into the Agrobacterium LBA4404 by electroporation, and a colony was selected by gentamycin (Gm 50 µg/mL). pLC40:35S-IGUS or pLC40GWB:35S-IGUS was introduced into the resulting LBA4404/pVGW via triparental mating. These Agrobacteria are described as LBA4404/pVGW/pLC40:35S-IGUS and

- 97 -

LBA4404/pVGW/pLC40GWB:35S-IGUS. Agrobacterium colonies (Km- and Gm-resistant) were directly analyzed by PCR to confirm the presence of the VirG, hpt or bar, and GUS genes.

[0202] Moreover, pIG121Hm derived from the IncP plasmid pBI121 (Hiei et al. (1994) Plant J 6: 271-282) was introduced into LB4404/pTOK47 to prepare Agrobacterium LB4404/pTOK47/pIG121Hm, which was used as a control in maize transformation experiments.

[0203] 2) Transformation of maize

10 Maize immature embryos having a size of about 1.2 mm (variety: A188) were aseptically removed from a plant grown in a greenhouse, and immersed in a liquid medium for suspending Agrobacterium (LS-inf, Ishida et al. 1996). After thermal treatment at 46°C for 3 minutes, the immature embryos were washed with the same liquid medium. After centrifugation at 15,000 rpm, 4°C, for 10 minutes, the immature embryos were then immersed in a suspension of each strain at about 1×10^9 cfu/ml in LS-inf medium (containing 100 μ M acetosyringon) and then plated on a coculture medium (LS-AS
20 (Ishida et al. 1996 Nat Biotechnol 14:745-50) containing AgNO₃, CuSO₄). After incubation at 25°C in darkness for 3 days, the immature embryos were partially used for GUS analysis.

[0204] The cocultured immature embryos were plated on a selective medium containing hygromycin or phosphinothricin (Ishida et al. (2003) Plant Biotechnology 20:57-66) and incubated. A callus grown from the immature embryo was excised and plated on a regeneration medium containing hygromycin (Hm) or

- 98 -

phosphinothricin (PPT) (Ishida et al. 1996 Nat Biotechnol 14:745-50), and incubated under illumination. After two weeks, regenerated plants showing resistance to Hm or PPT were investigated.

[0205] Initially, A188 immature embryos were inoculated with various strains and observed for the transient expression of the GUS gene on day 3 of coculture. Immature embryos inoculated with the control LBA4404/pSB134 showed the expression of the GUS gene over a wide range. However, few immature embryos inoculated with LBA4404/pLC40:35S-IGUS showed the expression except for limited ones showing the expression in very small spots. No increase in expression was found when EHA105 was used as a host. Most of immature embryos inoculated with LBA4404/pTOK47/pLC40:35S-IGUS, LBA4404/pLC40GWHvG1:35S-IGUS, LBA4404/pVGW/pLC40:35S-IGUS and LBA4404/pVGW/pLC40GWB:35S-IGUS showed spots representing the expression of the GUS gene to a lesser extent than with LBA4404/pSB134, thus verifying that the gene transfer efficiency is improved by the coexistence with a plasmid containing the virB gene and virG gene derived from the Agrobacterium strain A281, or the coexistence with a plasmid containing virGN54D, or the addition of the virG gene. On the other hand, there is no difference in the expression of the GUS gene between pLC40GWHvG1:35S-IGUS and pLC40GWHvGC1:35S-IGUS, showing that a single nucleotide substitution for removing an NspV recognition site does not influence the virG activity.

[0206] Then, we tried to create transformed plants by

- 99 -

incubating the cocultured immature embryos in a selective medium containing Hm or PPT and a regeneration medium. When EHA105 was used as a host, the pLCSBGWBSW vector gave no PPT-resistant plant. When LBA4404 was used as a host, however, the pLCSBGWBSW vector gave plants showing resistance to PPT at an efficiency comparable to that of the superbinary vector pSB131 (containing the GUS gene and the bar gene in the T-DNA region, Ishida et al. 1996 Nat Biotechnol 14:745-50) using the same strain as a host (Table 14).

10 [0207] LBA4404/pTOK47/pLC40GWB:35S-IGUS was also shown to give PPT-resistant plants at a high efficiency comparable to that of the superbinary vector pSB131. When the hygromycin resistance gene was used as a selectable marker gene, a pLC40 series cosmid vector (pLC40:35S-IGUS) combined with pTOK47 also gave hygromycin-resistant plants (Table 13). pLC40GWHvG1 containing the virG gene also achieved an efficiency comparable to that of the superbinary vector SB134 (containing the GUS gene and the hygromycin resistance gene in the T-DNA region, Hiei and Komari 2006 Plant Cell, Tissue and Organ Culture 85: 271-283) (Table 14).

[0208]

- 100 -

Table 14

Results of transformation of maize

Experiment	Strain	Selective drug	# of immature embryos		Redifferentiation ratio (B/A, %)
			Inoculated (A)	redifferentiated (B)	
1	LBA4404 (pLCSBGWBSW)	PPT	46	10	21.7
	EHA105 (pLCSBGWBSW)	PPT	46	0	0
	LBA4404 (pSB131)	PPT	45	9	20.0
2	LBA4404 (pLC40GWB:35S-IGUS)	PPT	56	0	0
	LBA4404 (pLC40GWB:35S-IGUS /pTOK47)	PPT	57	14	24.6
	LBA4404 (pSB131)	PPT	59	19	32.2
3	LBA4404(pLC40:35S-IGUS)	Hm	43	0	0
	LBA4404(pLC40:35S-IGUS /pTOK47)	Hm	44	2	4.5
	LBA4404(pIG121Hm)	Hm	42	0	0
	LBA4404(pIG121Hm/pTOK47)	Hm	42	0	0
4	LBA4404(pLC40GWHvG1)	Hm	59	5	8.5
	LBA4404(pSB134)	Hm	57	5	8.8

PPT:phosphinothricin,Hm:hygromycin

In order to examine the influence of pVGW on maize transformation, maize was then transformed with LBA4404/pLC40:35S-IGUS, LBA4404/pVGW/pLC40:35S-IGUS, LBA4404/pVGW/pLC40GWB:35S-IGUS, and LBA4404/pSB134, and the regenerated individuals were analyzed for GUS expression. As a result, the proportion of the number of GUS-expressing

- 101 -

individuals in pLC40:35S-IGUS was 0% (0/16), while the proportion of the number of GUS-expressing individuals per inoculated immature embryo in pLC40:35S-IGUS and pLC40GWB:35S-IGUS both combined with pVGW reached 40% (6/15) and 30% (6/20), respectively, which were comparable to 41.2% (7/17) in the superbinary vector pSB134. Thus, the transformation of maize with pLC vectors could be achieved at high efficiency by using pVGW.

[0209] We further tried to transform plant genomic fragments into maize by combining a pLC vector and the pVGW vector. A genomic fragment (30-35 kb) of Sudan grass was randomly cloned into the NspV site of the vector pLC40GWB. The resulting E.coli plasmid was transferred to Agrobacterium harboring pVGW (LBA4404) via triparental mating. In this manner, Agrobacterium harboring both of the plasmids pLC40GWB containing the genomic fragment of Sudan grass and pVGW was prepared, and inoculated into maize immature embryos (variety: A188). Transformed cells were selected to show that 17 of the 27 fragments inoculated gave redifferentiated plants (Table 15). This showed that plant genomic fragments can be efficiently transformed into maize by the combination of pLC and pVGW.

[0210] These results demonstrated that maize transformation can be efficiently achieved by the combination with a plasmid carrying a DNA fragment containing the virB gene and virG gene derived from the Agrobacterium strain A281 such as pTOK47, or the combination with a plasmid containing the virGN54D gene such as pVGW, or the incorporation of the virG gene into a pLC

- 102 -

vector such as pLC40GWHvG1.

[0211]

Table 15

Results of the transformation of randomized plant genomic fragments into maize using a pLC/pVGW vector system

Genome donor plant	Strain	# of genomic fragments used for Agrobacterium-mediated transformation (A)	# of genomic fragments that regenerated PPT-resistant individuals (B)	B/A(%)
Sudan grass	LBA4404(pLC40GWB/pVGW)	27	17	63.0

Example 6: Isolation of a gene of interest from BAC clones using pLC vectors

Komori et al. (2004) (Plant J 37: 315-325) found that a cytoplasmic male sterile strain restores fertility when it is transformed with the PPR791 gene isolated from the rice variety IR24, thus demonstrating that PPR791 is the fertility restorer gene Rf-1. The PPR791 gene was identical with the PPR8-1 gene of the rice variety Milyang 23 that had been previously reported as a candidate for Rf-1 by Kazama and Toriyama (2003) (FEBS Lett 544: 99-102). Thus, the BAC clone OSIMBb0046F08 of Milyang 23 from which the PPR8-1 gene had been derived was obtained from Clemson University, and a model experiment was performed for isolating Rf-1 from the BAC.

[0212] Initially, a plasmid was extracted from OSIMBb0046F08 using High Purity Plasmid Midiprep System (Marligen). The plasmid was partially digested with TaqI and a DNA fragment around 30 kb was recovered by sucrose density gradient centrifugation. This DNA fragment was ligated to the

- 103 -

BstBI-digested and CIP-treated pLC40GWH vector or the BstBI-digested and CIP-treated pSB200 vector using DNA Ligation Kit <Mighty Mix> (Takara Bio Inc.). The resulting construct was transferred into E.coli by electroporation to give colonies of transformants on an LB plate containing an appropriate antibiotic (50 µg/ml kanamycin or spectinomycin). To determine the presence or absence of the Rf-1 gene in the resulting plasmid, direct PCR (see Examples 1, 3)) was performed by using these colonies as templates along with primers designed for the Rf-1 gene (WSF7T7R1 and IR50226R, Table 16) to select Rf-1 positive clones giving an amplified product of about 2 kb from Rf-1 negative clones showing no amplification of the product. The incidence of positive clones in this PCR screening was 5/39 (12.8%) in the pLC40GWH construct and 6/96 (6.3%) in the pSB200 construct. That is, the cloning efficiency of a gene of interest was about twice higher in the pLC vector than pSB.

[0213] One positive clone and two negative clones selected from the pLC40GWH construct, and one positive clone and two negative clones selected from the pSB200 construct were transferred from E.coli to Agrobacterium via triparental mating. The cytoplasmic male sterile strain MS Koshihikari was infected with the resulting Agrobacterium by the method described in Komori et al. (2004). The resulting transformed rice was acclimated and then grown in a greenhouse. During the maturing stage, an average ear was collected from each individual and evaluated for the fertility rate. The results showed that transformants from constructs containing no Rf-1

- 104 -

(pLC-7, pLC-11, pSB-1, pSB-7) were sterile, while constructs containing Rf-1 (pLC-8, pSB-37) gave fertile transformants (Table 17).

[0214] These results demonstrated that a gene of interest can be efficiently identified by preparing a library from DNA of BAC containing the gene of interest using a cosmid vector for plant transformation and transferring it into a plant and then selecting a plant showing an expected phenotype.

[0215]

Table 16

Primer Name	Sequence	Length
WSF7T7R1	5'-AGT GTG TGG CAT GGT GCA TTT CCG-3'	24mer
IR50226R	5'-CTC TAC AGG ATA CAC GGT GTA AGG-3'	24mer

SEQ ID NOs: 80-81 in order from the top.

[0216]

Table 17

Fertility restoration by various constructs

Construct	Presence(+) or absence(-) of Rf-1	# of individuals analyzed	# of individuals fertile
pLC-8	+	6	4
pLC-7	-	9	0
pLC-11	-	8	0
pSB-37	+	9	6
pSB-1	-	9	0
pSB-7	-	9	0

In conclusion, pLC vectors are characterized in that:

- 105 -

1. they allow easy cloning of DNA in the order of 25-40 kb;
2. they are stable in bacteria; and
3. they allow efficient transformation of plants, especially monocotyledons.

pLC vector series are useful for handling medium-size DNA in the field of functional genomics.

- 106 -

CLAIMS

[1] A cosmid vector having a full length of 15 kb or less characterized in that:

1) it contains an origin of replication (oriV) of an IncP plasmid, but does not contain any origin of replication of other plasmid groups;

2) it contains the trfA1 gene of an IncP plasmid;

3) it contains an origin of conjugative transfer (oriT) of an IncP plasmid;

4) it contains the incC1 gene of an IncP plasmid;

5) it contains a cos site of lambda phage and the cos site is located outside the T-DNA;

6) it contains a drug resistance gene expressed in E.coli and a bacterium of the genus Agrobacterium;

7) it contains a T-DNA right border sequence of a bacterium of the genus Agrobacterium;

8) it contains a T-DNA left border sequence of a bacterium of the genus Agrobacterium;

9) it contains a selectable marker gene for plant transformation located between 7) and 8) and expressed in a plant; and

10) it contains restriction endonuclease recognition site(s) located between 7) and 8) for cloning a foreign gene.

[2] The cosmid vector of claim 1 wherein the selectable marker gene for plant transformation is a hygromycin resistance gene, a phosphinotricin resistance gene or a kanamycin resistance gene.

- 107 -

[3] The cosmid vector of claim 1 or 2, which contains the korB gene of an IncP plasmid.

[4] The cosmid vector of any one of claims 1 to 3 which is:

the cosmid vector pLC40 consisting of the nucleotide sequence of SEQ ID NO: 2 or an equivalent thereof having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO: 2;

the cosmid vector pLC40GWH consisting of the nucleotide sequence of SEQ ID NO: 3 or an equivalent thereof having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO: 3;

the cosmid vector pLC40bar consisting of the nucleotide sequence of SEQ ID NO: 4 or an equivalent thereof having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO: 4;

the cosmid vector pLC40GWB consisting of the nucleotide sequence of SEQ ID NO: 5 or an equivalent thereof having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO: 5;

the cosmid vector pLC40GWHKorB consisting of the nucleotide sequence of SEQ ID NO: 65 or an equivalent thereof having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO: 65;

the cosmid vector pLCleo consisting of the nucleotide sequence of SEQ ID NO: 66 or an equivalent thereof having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO: 66; or

- 108 -

the cosmid vector pLC40GWHvG1 consisting of the nucleotide sequence of SEQ ID NO: 7 or an equivalent thereof having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO: 7.

[5] The cosmid vector of claim 4 which is:

the cosmid vector pLC40 consisting of the nucleotide sequence of SEQ ID NO: 2;

the cosmid vector pLC40GWH consisting of the nucleotide sequence of SEQ ID NO: 3;

the cosmid vector pLC40bar consisting of the nucleotide sequence of SEQ ID NO: 4;

the cosmid vector pLC40GWB consisting of the nucleotide sequence of SEQ ID NO: 5;

the cosmid vector pLC40GWHKorB consisting of the nucleotide sequence of SEQ ID NO: 65;

the cosmid vector pLCleo consisting of the nucleotide sequence of SEQ ID NO: 66; or

the cosmid vector pLC40GWHvG1 consisting of the nucleotide sequence of SEQ ID NO: 7.

[6] A method for transforming a plant, comprising transforming the plant with a bacterium of the genus *Agrobacterium* harboring an expression vector containing a nucleic acid fragment of a plant inserted into the cosmid vector of any one of claims 1 to 5.

[7] The method of claim 6 wherein the nucleic acid fragment inserted has a size of 25-40 kb.

- 109 -

[8] The method of claim 6 or 7 characterized in that it uses a bacterium of the genus *Agrobacterium* harboring the following elements for transforming the plant:

1a) a vector containing a nucleic acid fragment of a plant and the *virG* gene of a bacterium of the genus *Agrobacterium* inserted into the cosmid vector of any one of claims 1 to 5; or

1b) a vector containing a nucleic acid fragment of a plant inserted into the cosmid vector of any one of claims 1 to 5, and a plasmid capable of coexisting with an IncP plasmid in a cell of a bacterium of the genus *Agrobacterium* and containing the *virG* gene of a bacterium of the genus *Agrobacterium*, and

2) a Ti plasmid or Ri plasmid of a bacterium of the genus *Agrobacterium*.

[9] The method of claim 8 wherein the *virG* gene of a bacterium of the genus *Agrobacterium* in 1a) or 1b) is *virGN54D*.

[10] The method of claim 8 wherein the plasmid containing the *virG* gene of a bacterium of the genus *Agrobacterium* in 1b) contains an origin of replication of an IncW plasmid.

[11] The method of claim 10 wherein the plasmid containing the *virG* gene of a bacterium of the genus *Agrobacterium* in 1b) is pVGW having the amino acid sequence of SEQ ID NO:8 or pVGW2 having the amino acid sequence of SEQ ID NO:67.

[12] The method of claim 8 wherein the plasmid containing the

- 110 -

virG gene of a bacterium of the genus Agrobacterium in 1b) further contains the virB gene of a bacterium of the genus Agrobacterium.

[13] The method of claim 12 wherein the plasmid containing the virG gene of a bacterium of the genus Agrobacterium in 1b) contains an origin of replication of an IncW plasmid.

[14] The method of claim 13 wherein the plasmid containing the virG gene of a bacterium of the genus Agrobacterium in 1b) is pTOK47.

[15] A map-based cloning method comprising the steps of:

1) partially or completely digesting BAC clones containing candidate genes responsible for a plant phenotype with a restriction endonuclease;

2) subcloning DNA fragments obtained in step 1) using the cosmid vector of any one of claims 1 to 5 to construct a library; and

3) individually transferring clones constituting the library into a plant to evaluate the phenotypes of transformed plants.

[16] The map-based cloning method of claim 15 wherein the DNA fragments obtained in step 1) have a size of 25-40 kb.

[17] A method for transforming a plant, comprising transforming the plant with a bacterium of the genus Agrobacterium harboring the cosmid vector of any one of claims 1 to 5 and a plasmid vector characterized in that:

- 111 -

- 1) said plasmid vector contains an origin of replication of an IncW plasmid, but does not contain any origin of replication of other plasmid groups;
- 2) said plasmid vector contains the repA gene necessary for the replication of an IncW plasmid;
- 3) said plasmid vector contains a drug resistance gene expressed in E.coli and a bacterium of the genus Agrobacterium; and
- 4) said plasmid vector contains the virG gene of a bacterium of the genus Agrobacterium.

[18] The method for transforming of claim 17, wherein the plasmid with the characters of 1) - 4) has a full length of 10 kb or less.

[19] The method for transforming of claim 18 wherein the virG gene of a bacterium of the genus Agrobacterium is virGN54D.

[20] The method for transforming of claim 17, wherein the plasmid with the characters of 1) - 4) is:

the plasmid vector pVGW consisting of the nucleotide sequence of SEQ ID NO: 8 or an equivalent thereof consisting of a nucleotide sequence having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO: 8; or

the plasmid vector pVGW2 consisting of the nucleotide sequence of SEQ ID NO: 67 or an equivalent thereof consisting of a nucleotide sequence having at least 95% identity relative to the full length of the nucleotide sequence of SEQ ID NO:

- 112 -

67.

[21] The method for transforming of claim 20, wherein the plasmid with the characters of 1) - 4) is the plasmid vector pVGW consisting of the nucleotide sequence of SEQ ID NO: 8.

[22] The method for transforming of claim 20, wherein the plasmid with the characters of 1) - 4) is the plasmid vector pVGW2 consisting of the nucleotide sequence of SEQ ID NO: 67.

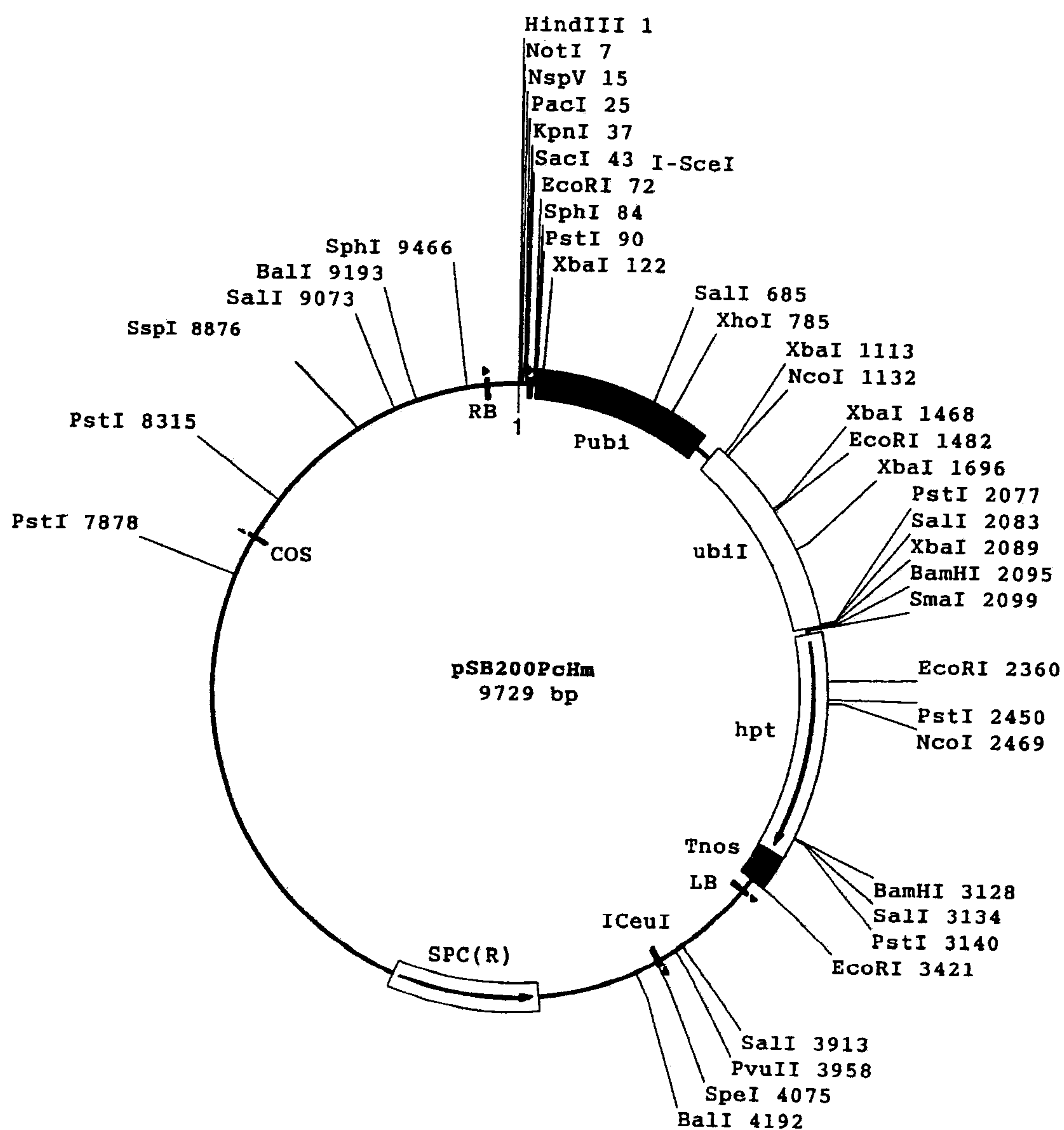
Fig. 1

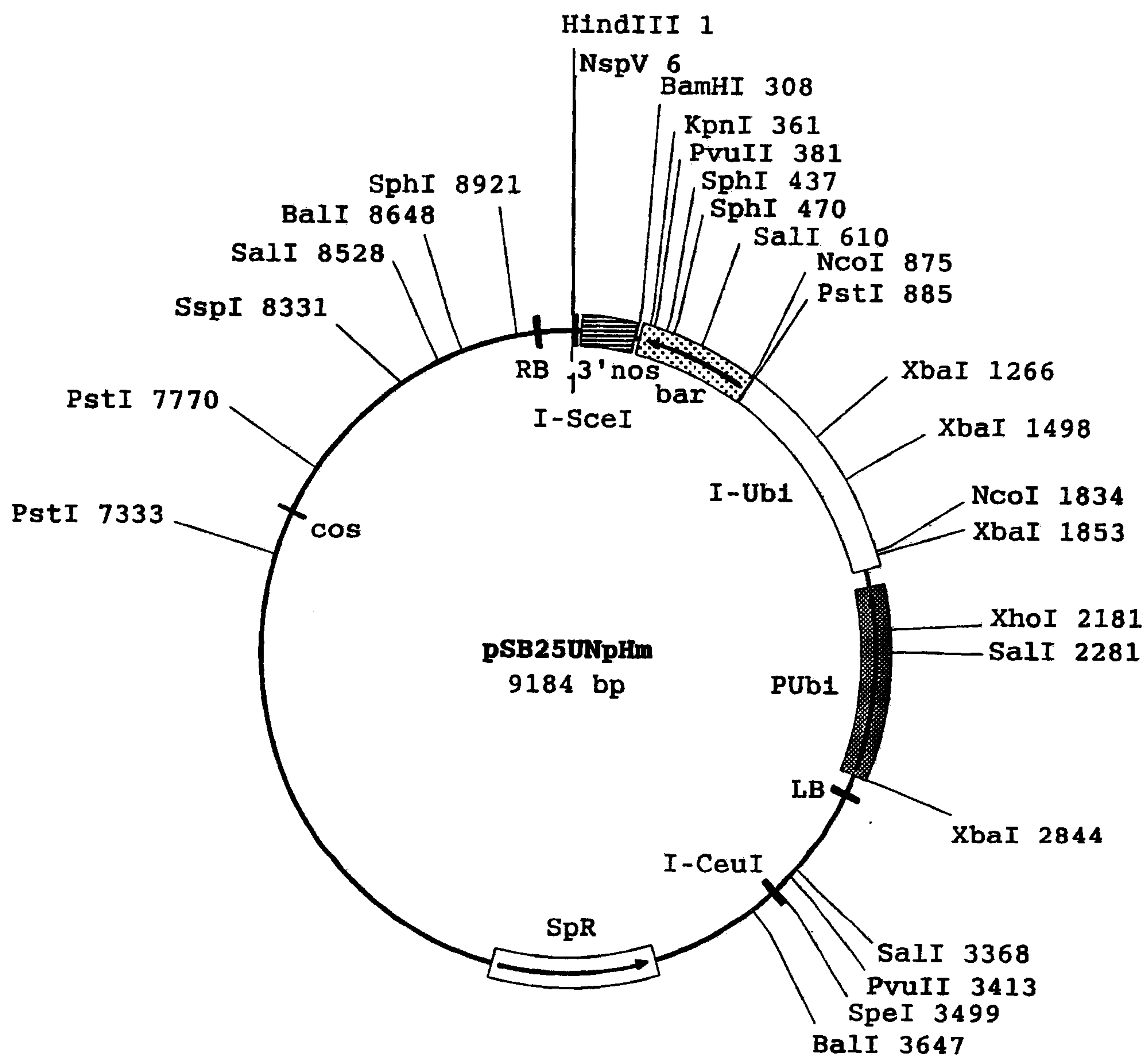
Fig. 2

Fig. 3

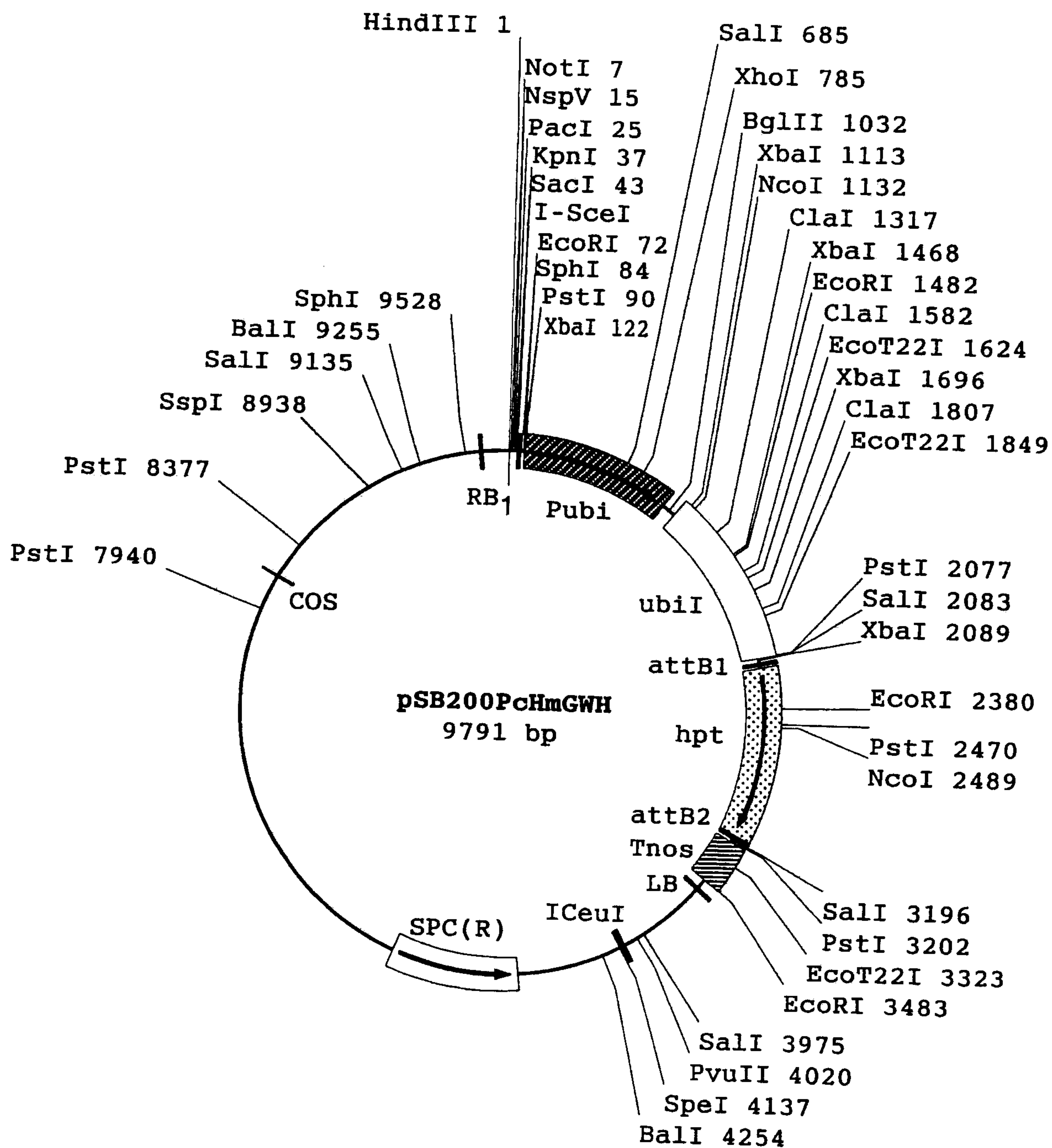


Fig. 4

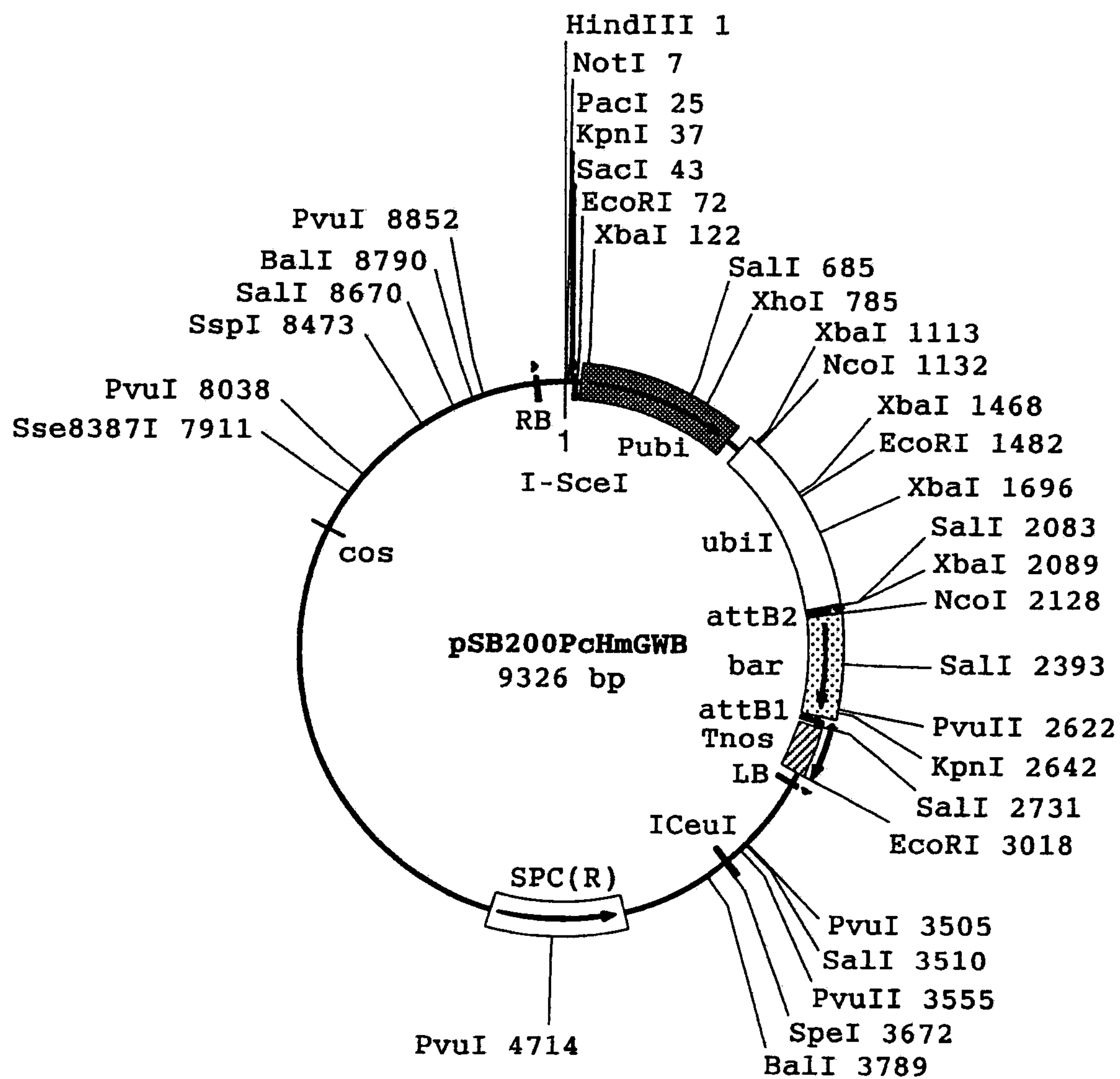
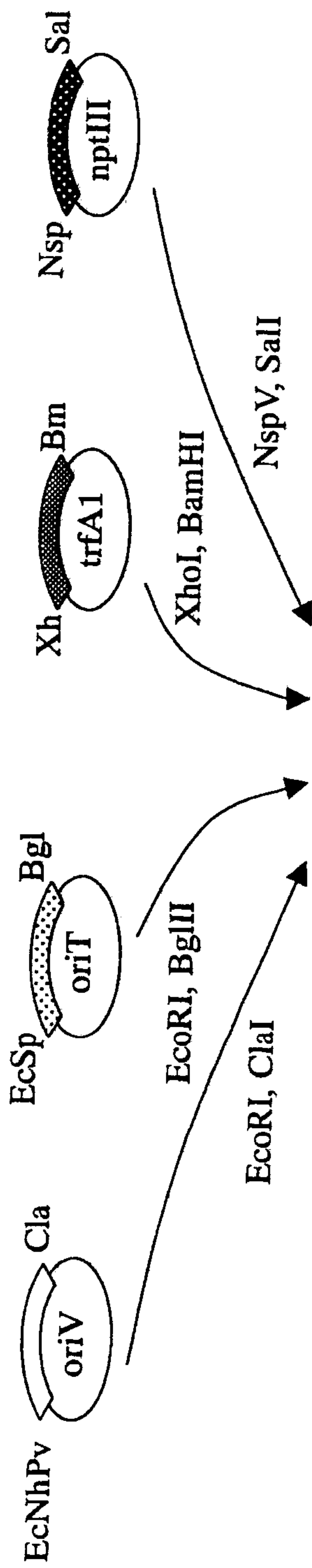
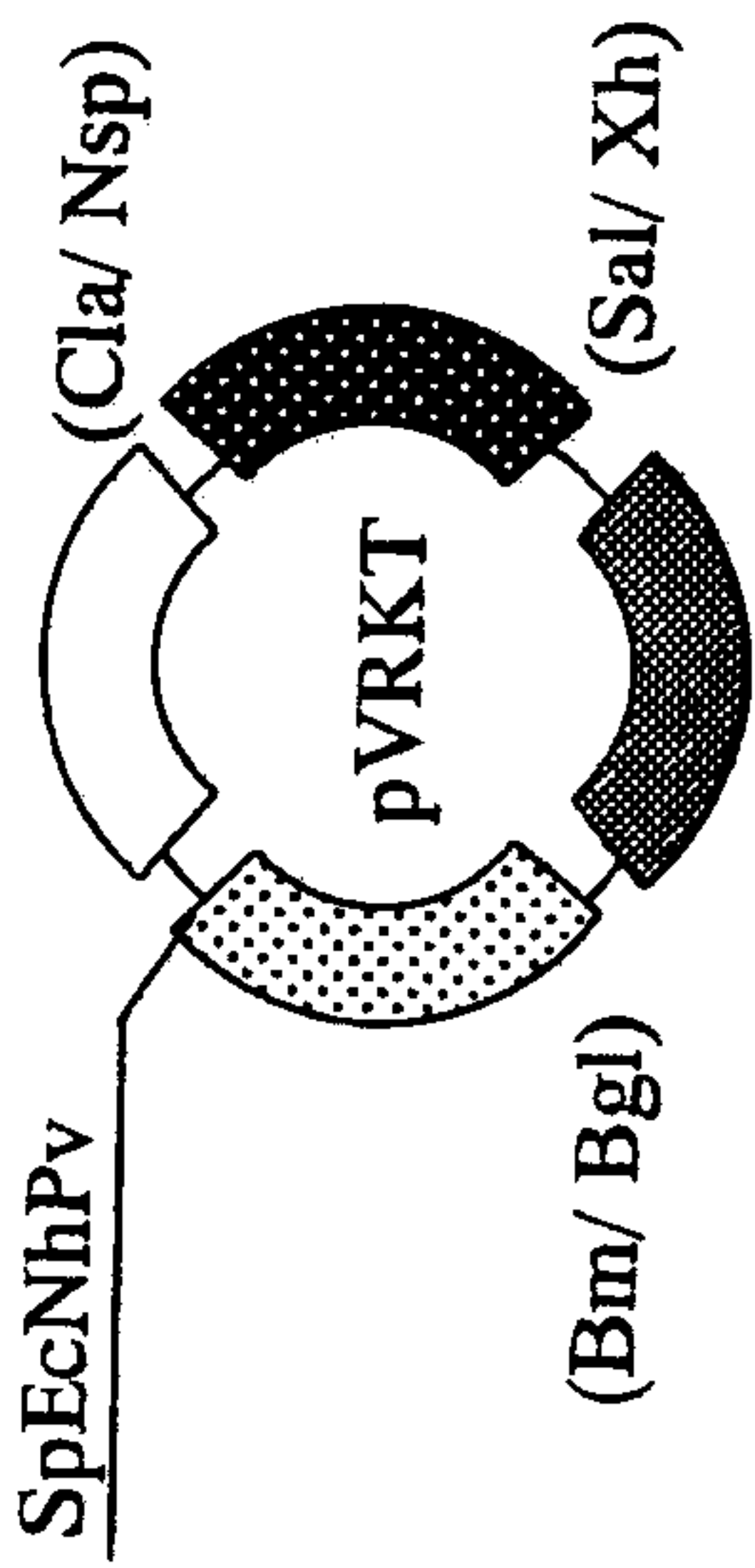


Fig. 5

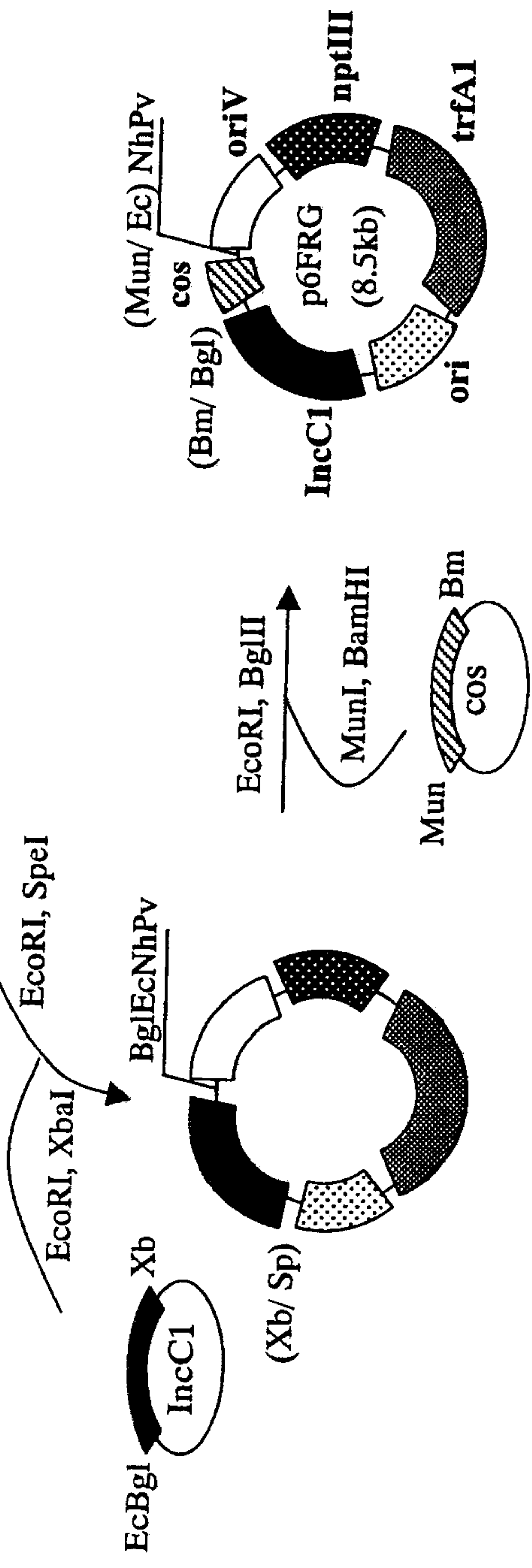
1. PCR & Cloning & Seq Check of each components



2. 4 fragments ligation



3. IncC insertion & cos insertion



4. T-DNA insertion

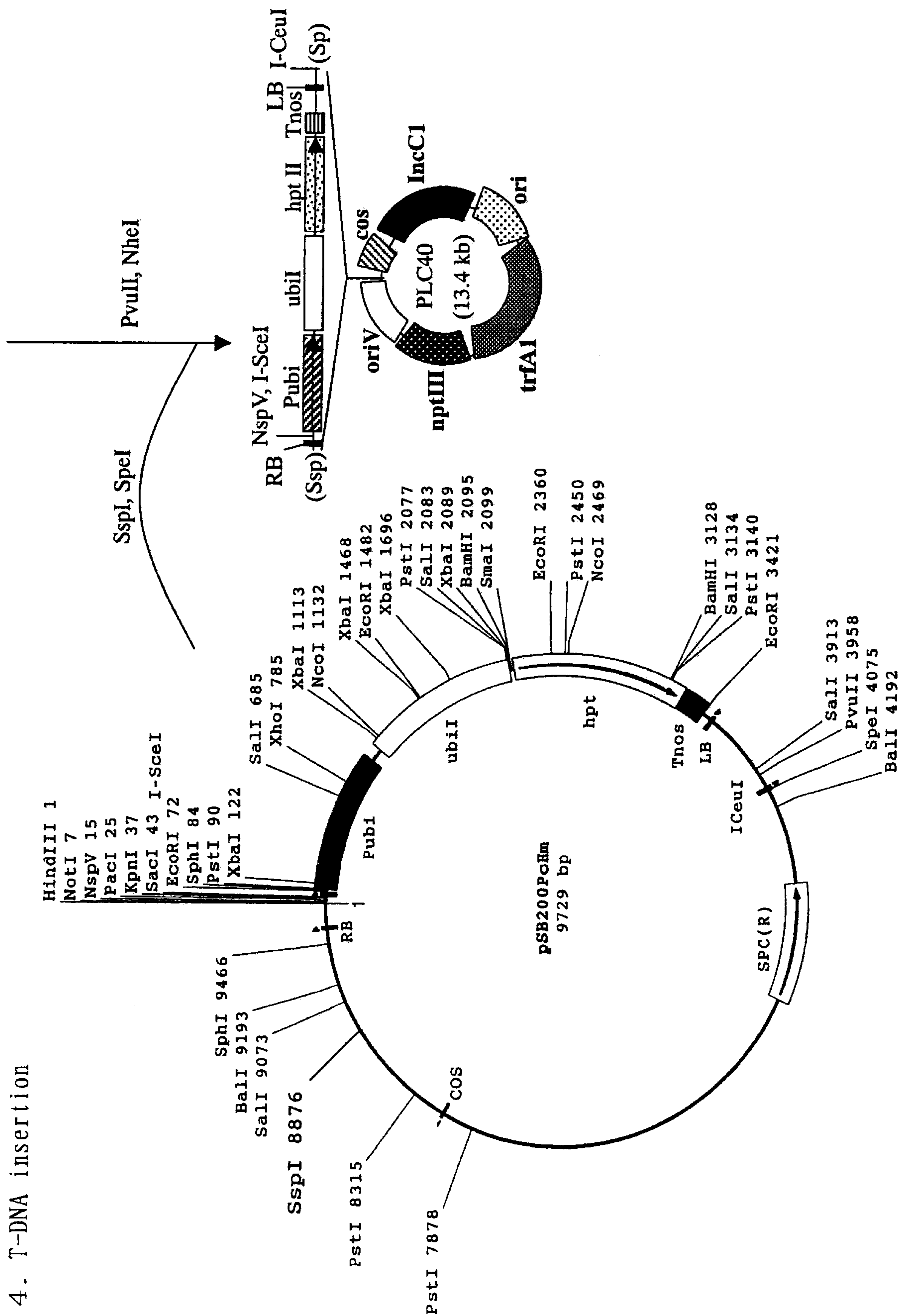


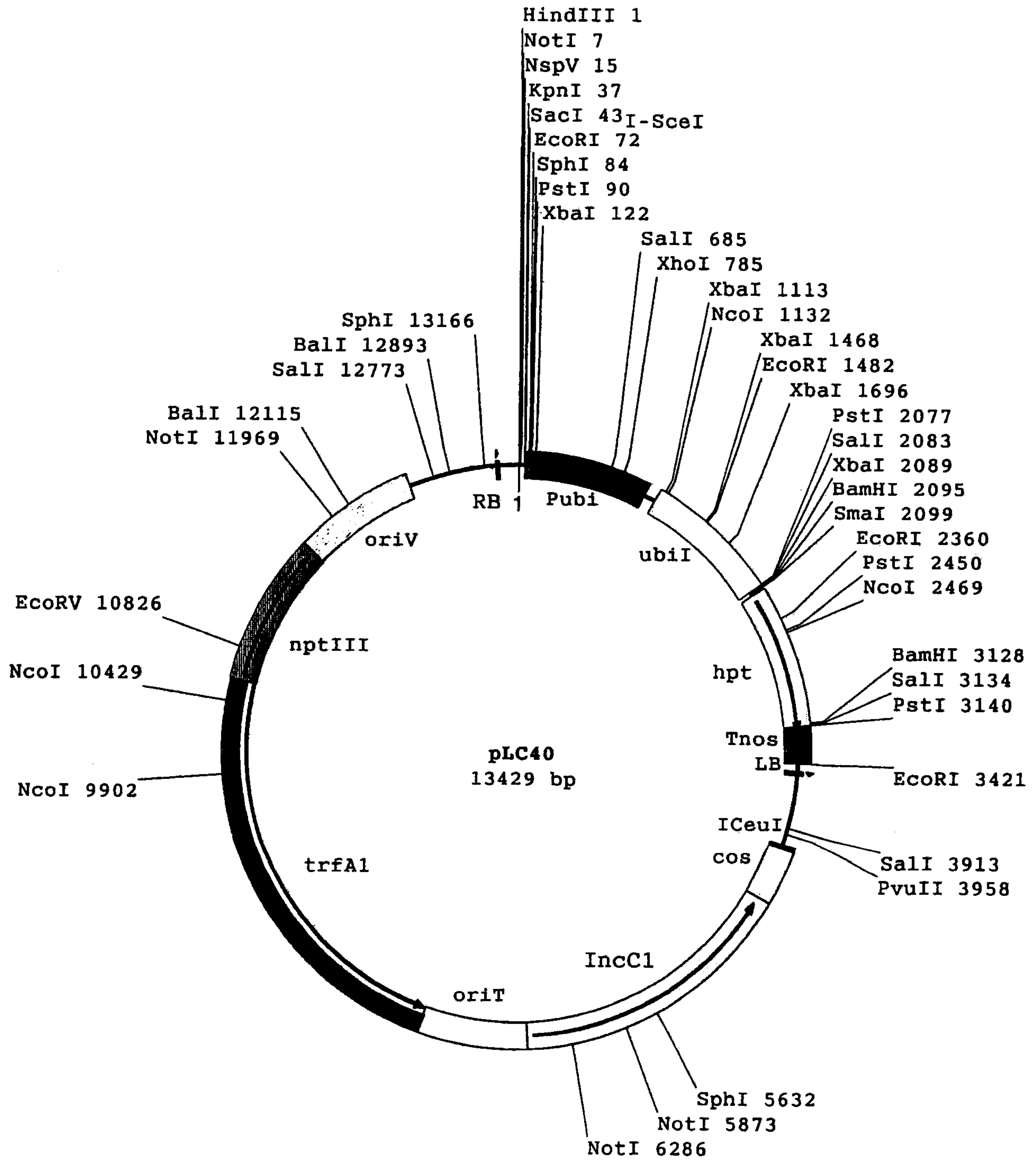
Fig. 6

Fig. 7

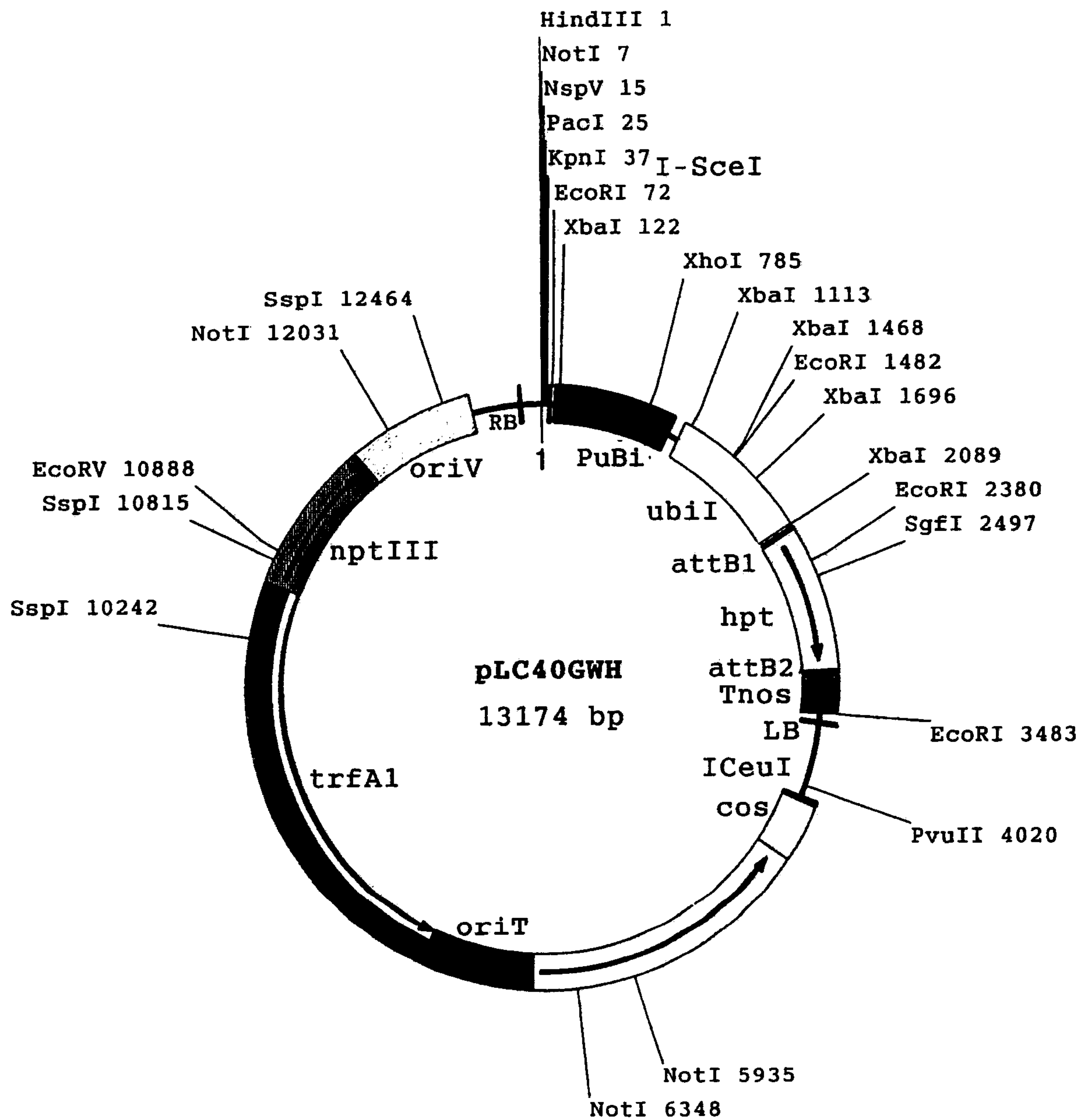


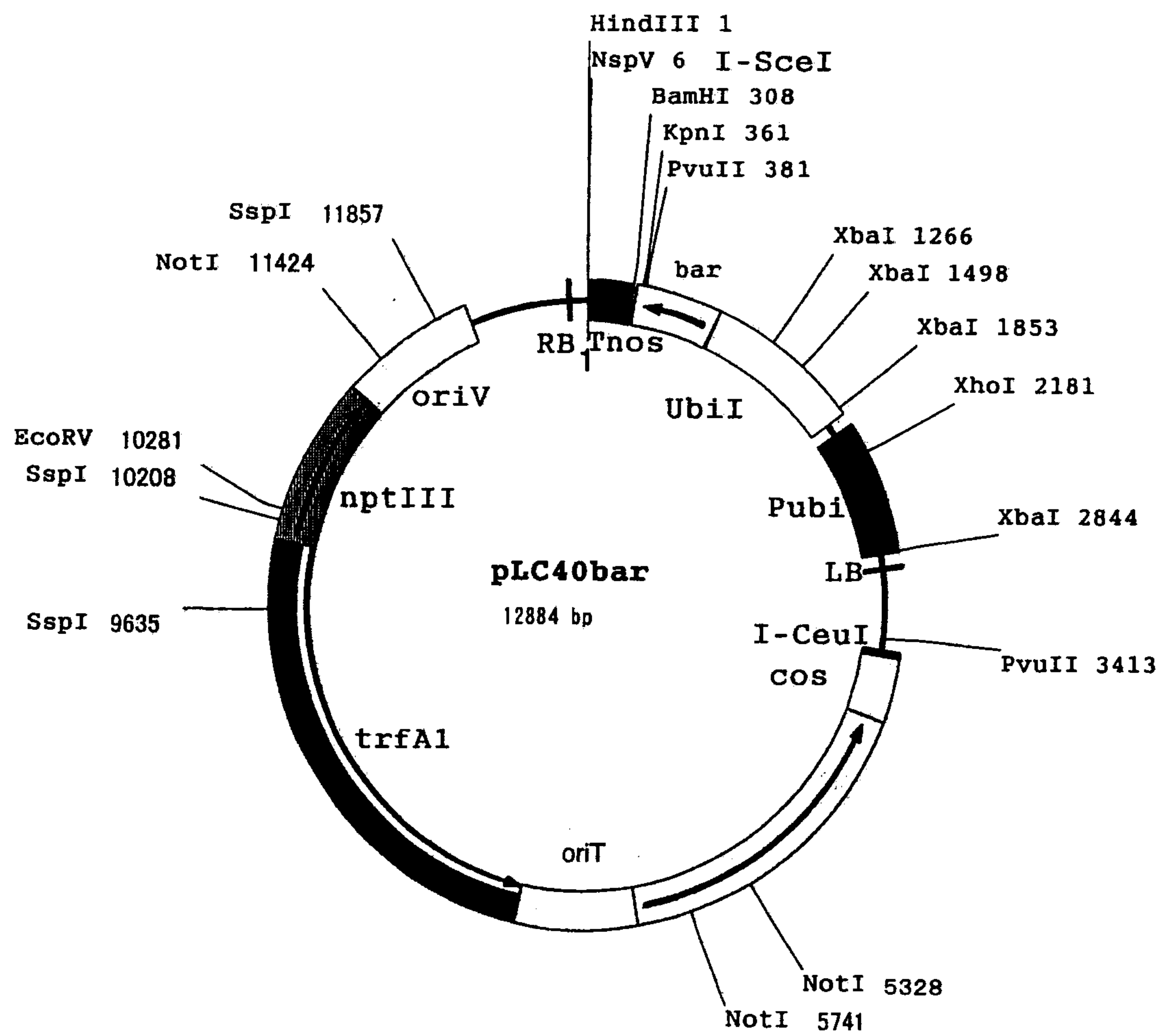
Fig. 8

Fig. 9

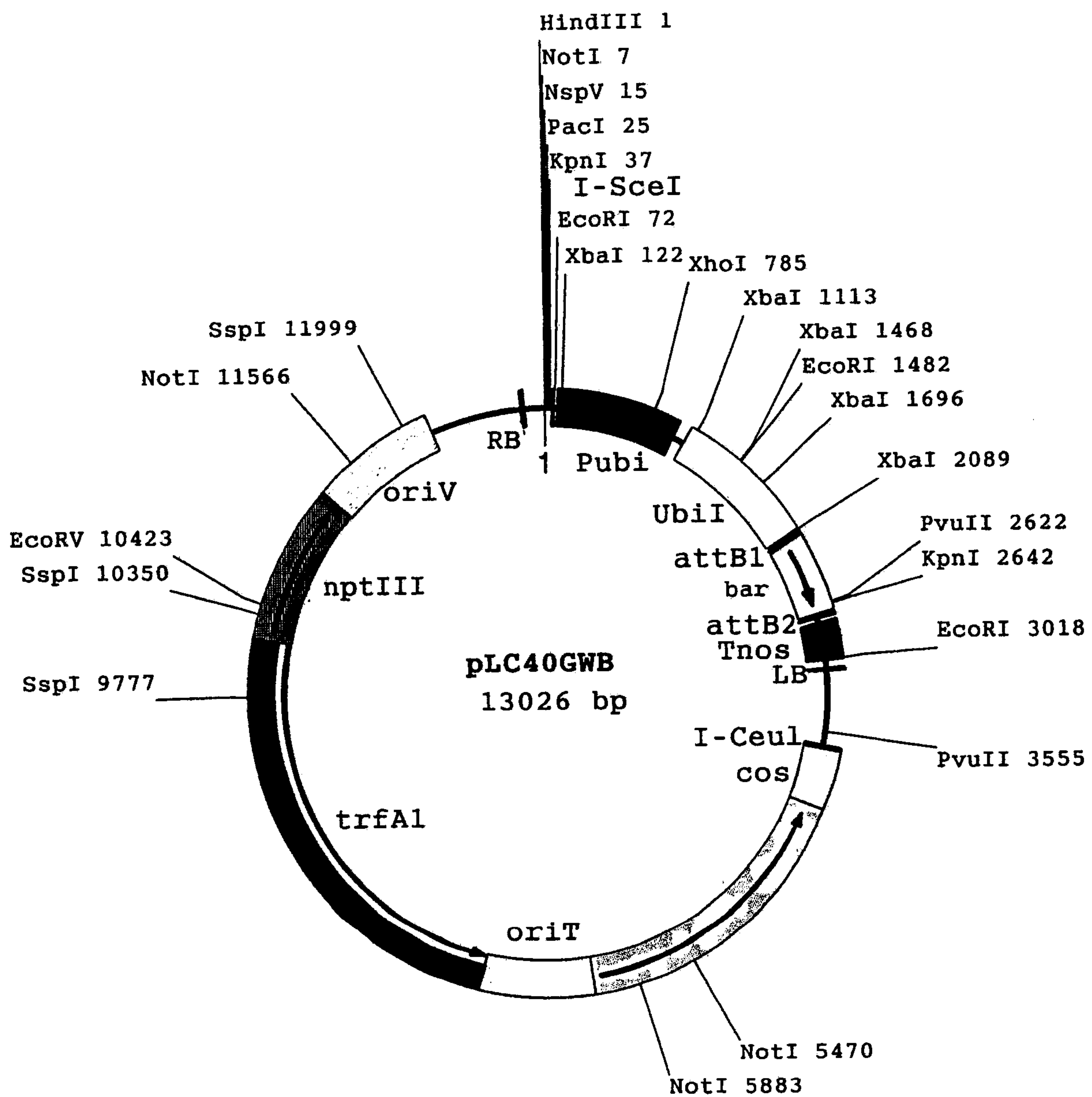


Fig. 10

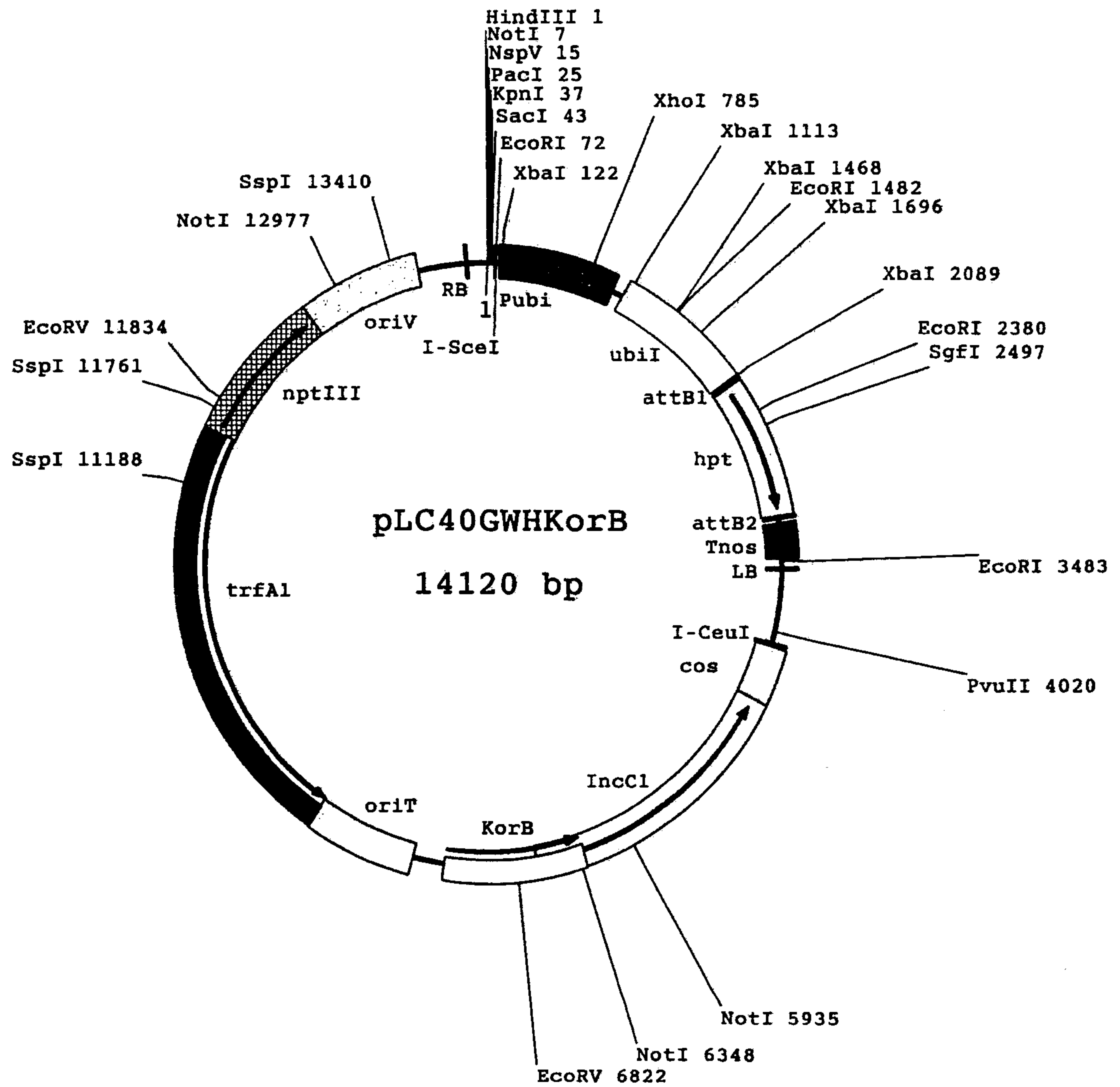


Fig. 11

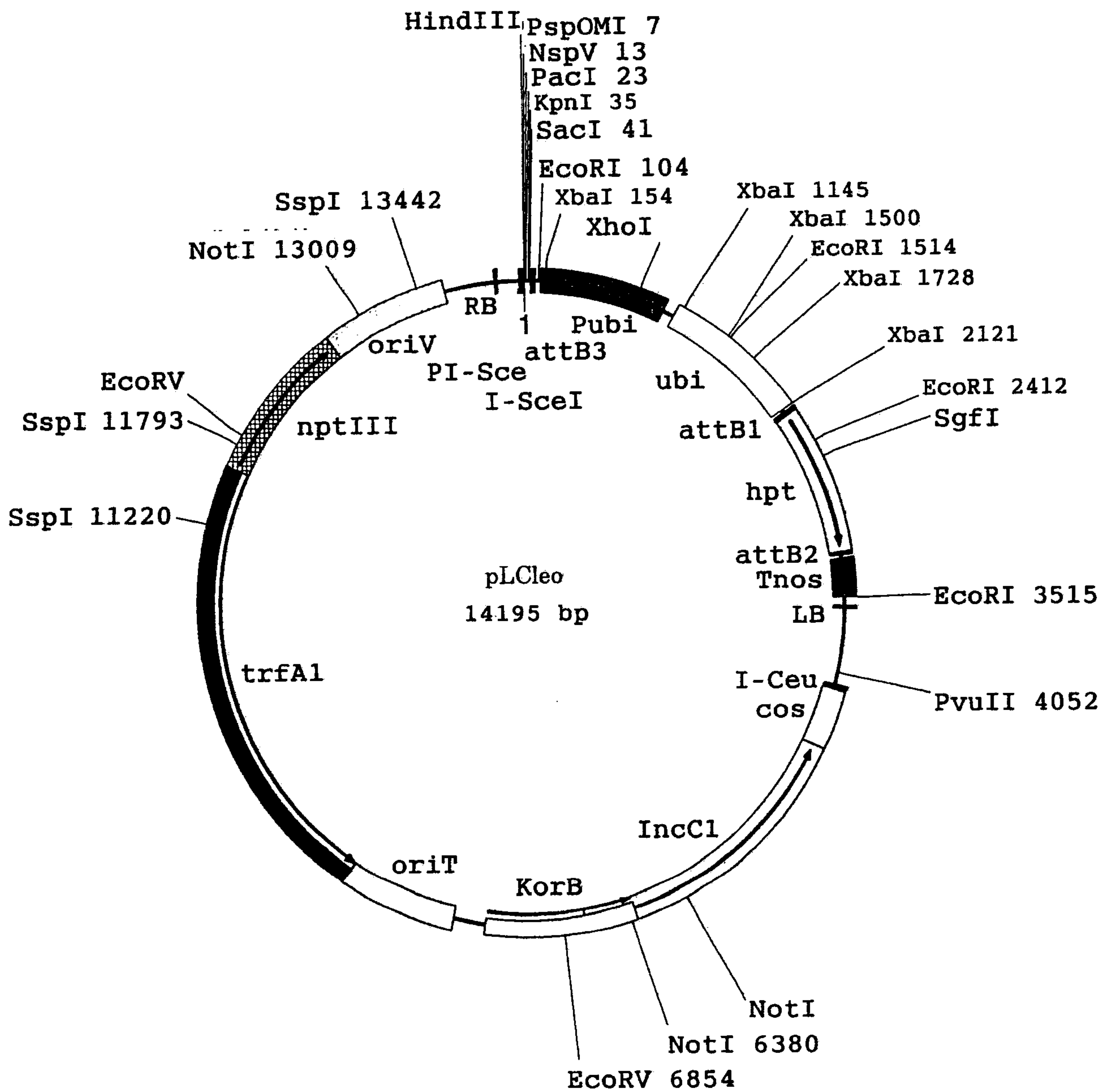


Fig. 12

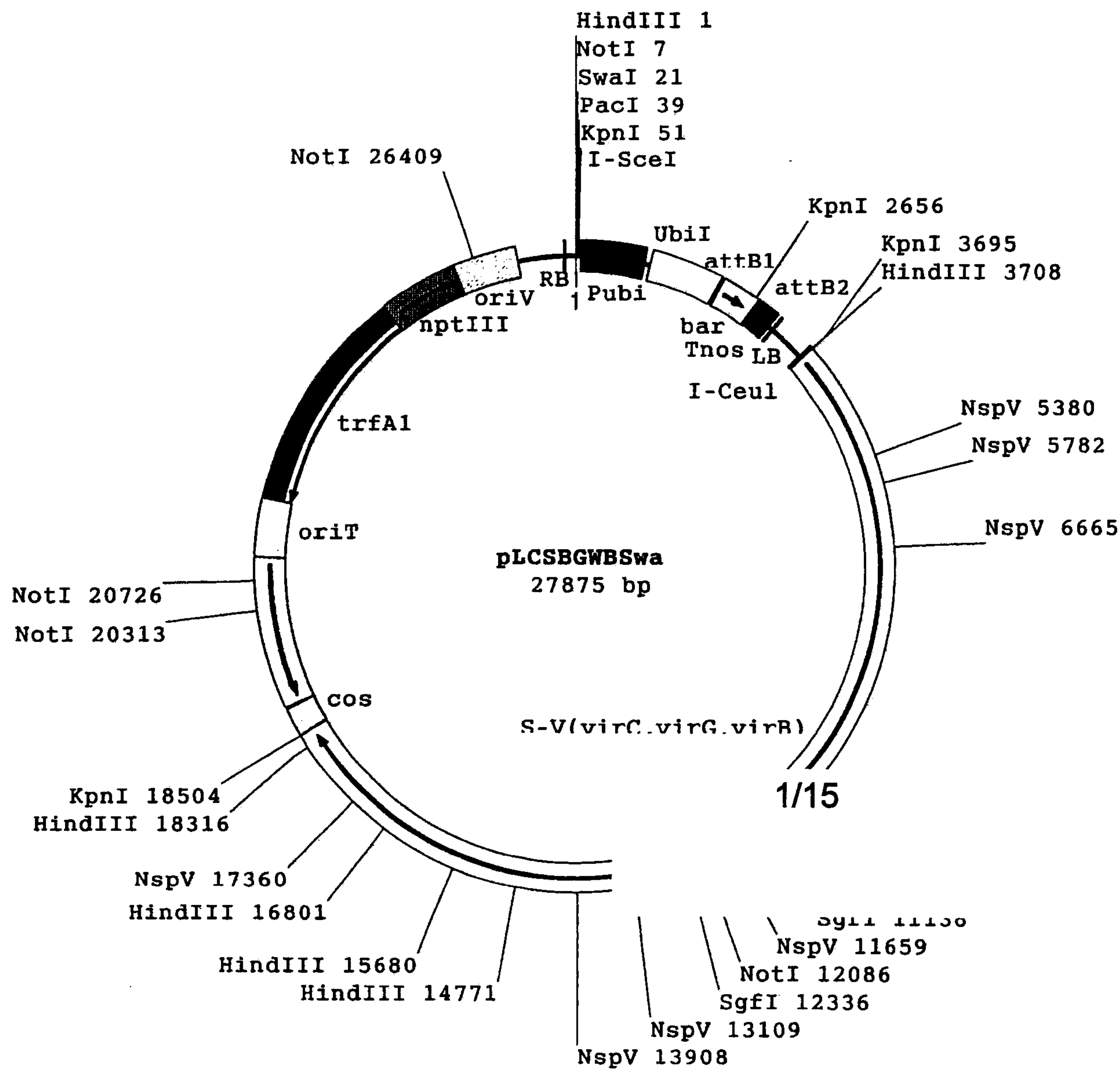
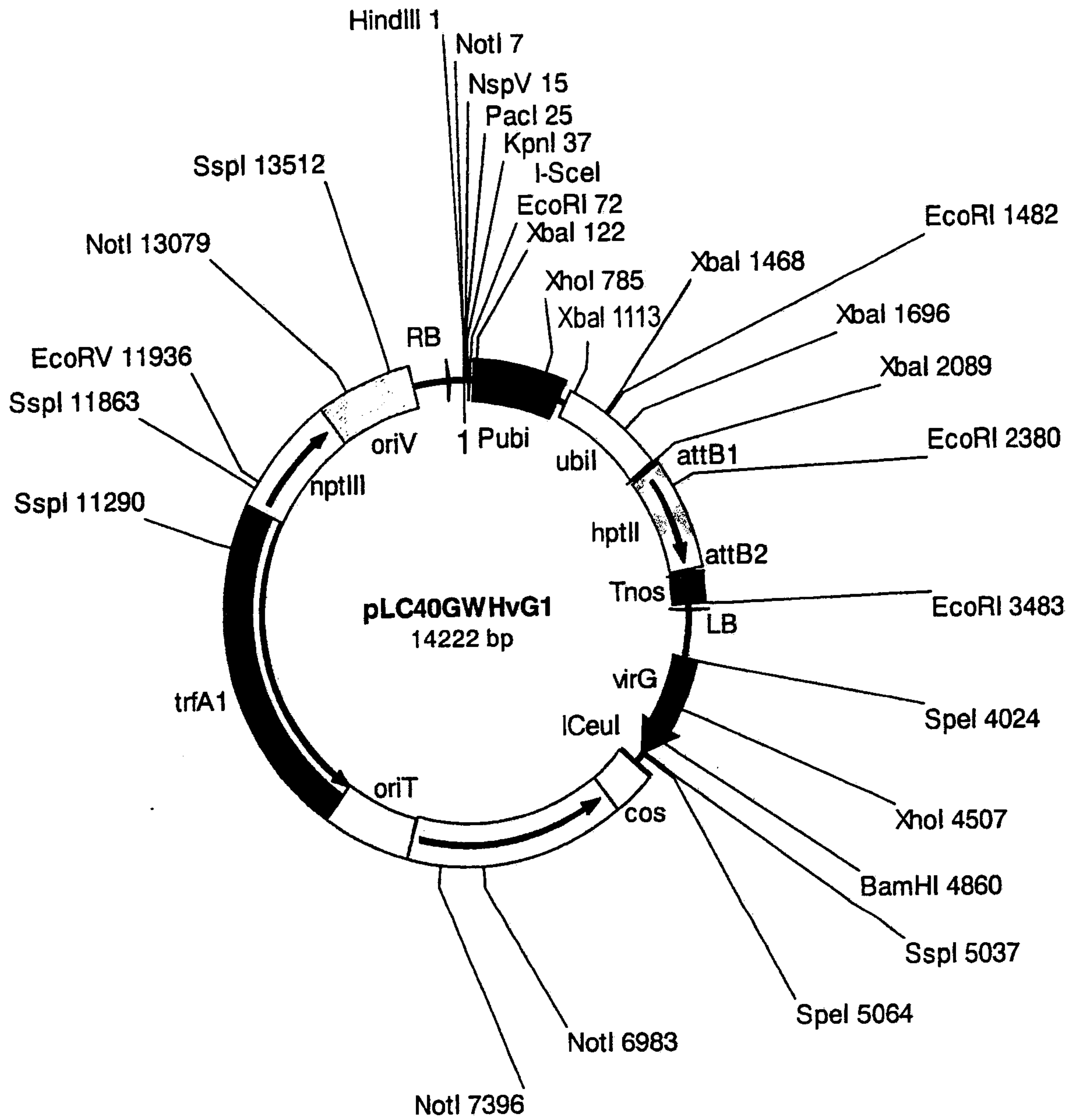
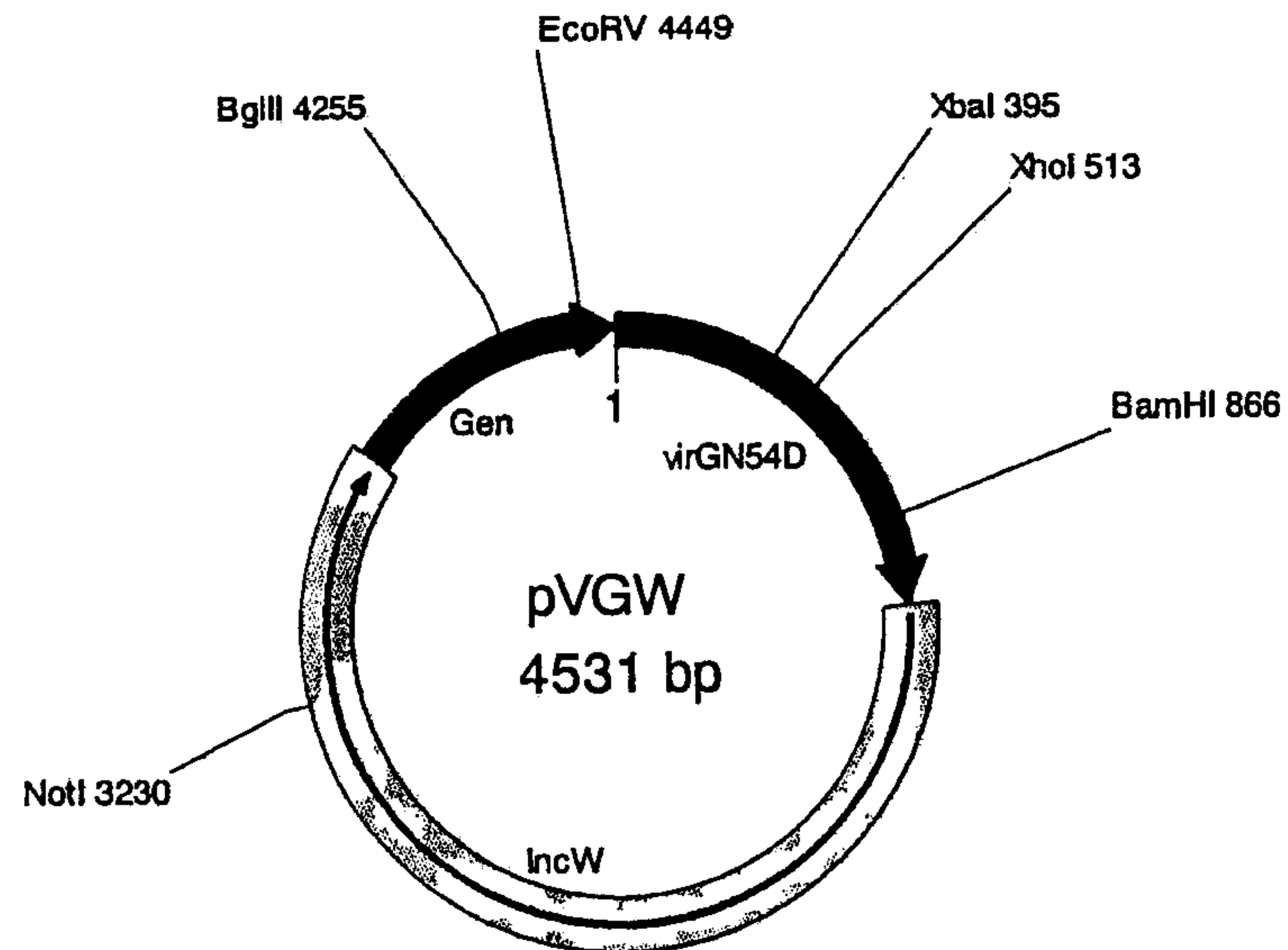
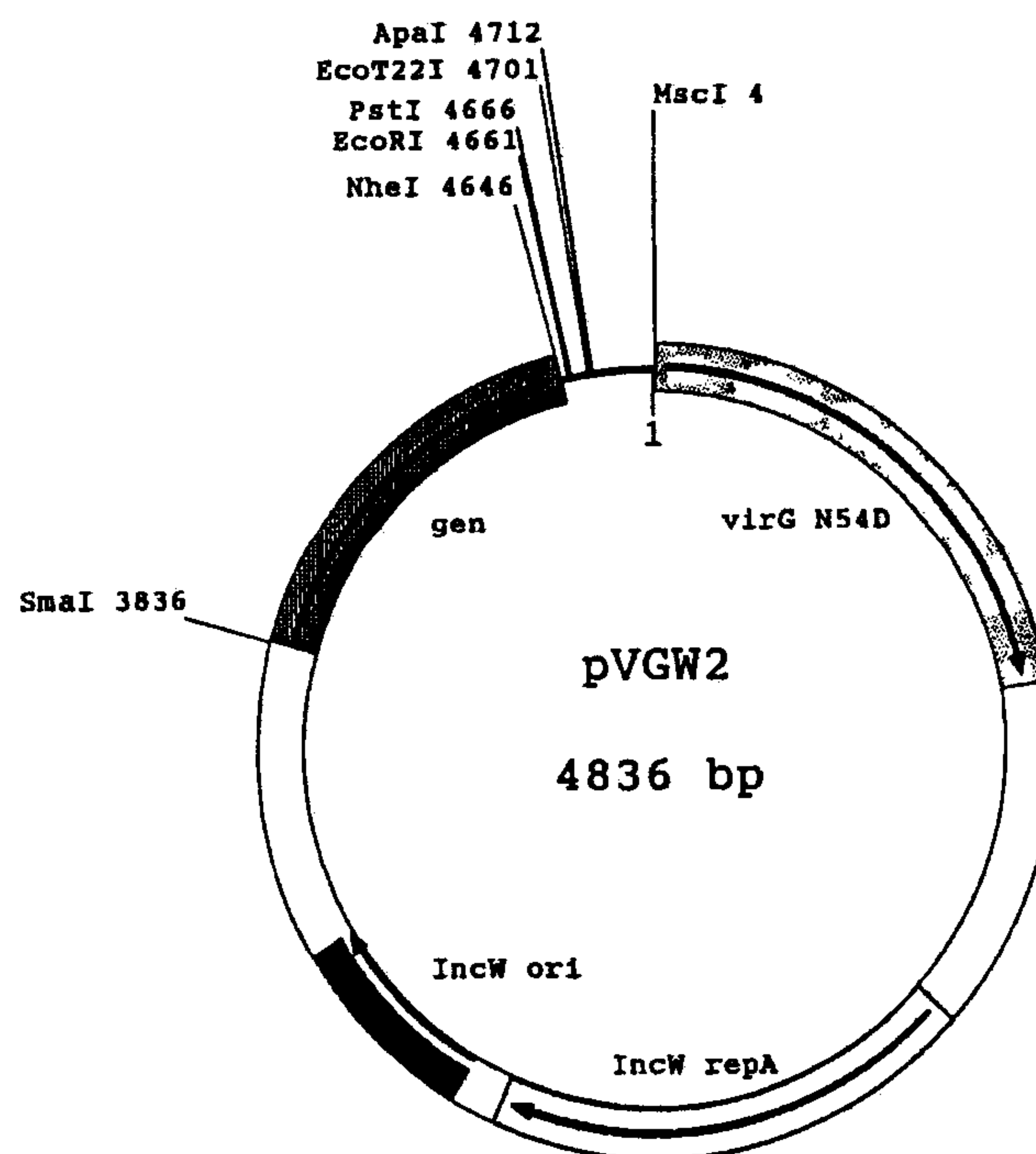


Fig. 13

15/15

**Fig. 14****Fig. 15**

