



(22) Date de dépôt/Filing Date: 2006/06/22
(41) Mise à la disp. pub./Open to Public Insp.: 2007/01/04
(62) Demande originale/Original Application: 2 613 336
(30) Priorité/Priority: 2005/06/24 (US11/166,609)

(51) Cl.Int./Int.Cl. *C12N 15/00* (2006.01),
A01H 1/00 (2006.01), *A01H 5/00* (2006.01),
C12N 15/82 (2006.01), *C12N 15/29* (2006.01)
(71) Demandeur/Applicant:
PIONEER HI-BRED INTERNATIONAL, INC., US
(72) Inventeurs/Inventors:
ALBERTSEN, MARC C., US;
FOX, TIMOTHY W., US;
HERSHEY, HOWARD P., US;
HUFFMAN, GARY A., US;
TRIMNELL, MARY, US;
WU, YONGZHONG, US
(74) Agent: TORYS LLP

(54) Titre : SEQUENCES NUCLEOTIDIQUES MEDIATRICES DE LA FERTILITE MALE ET SON PROCEDE
D'UTILISATION
(54) Title: NUCLEOTIDE SEQUENCE MEDIATING PLANT MALE FERTILITY AND METHOD OF USING SAME

(57) **Abrégé/Abstract:**

Nucleotide sequences mediating male fertility in plants are described, with DNA molecule and amino acid sequences set forth. Promoter sequences and their essential regions are also identified. The nucleotide sequences are useful in mediating male fertility in plants. In one such method, the homozygous recessive condition of male sterility causing alleles is maintained after crossing with a second plant, where the second plant contains a restoring transgene construct having a nucleotide sequence which reverses the homozygous condition. The restoring sequence is linked with a hemizygous sequence encoding a product inhibiting formation or function of male gametes. The maintainer plant produces only viable male gametes which do not contain the restoring transgene construct. Increase of the maintainer plant is also provided by self-fertilization, and selection for seed or plants which contain the construct.



Abstract

Nucleotide sequences mediating male fertility in plants are described, with DNA molecule and amino acid sequences set forth. Promoter sequences and their essential regions are also identified. The nucleotide sequences are useful in mediating male fertility in plants. In one such method, the homozygous recessive condition of male sterility causing alleles is maintained after crossing with a second plant, where the second plant contains a restoring transgene construct having a nucleotide sequence which reverses the homozygous condition. The restoring sequence is linked with a hemizygous sequence encoding a product inhibiting formation or function of male gametes. The maintainer plant produces only viable male gametes which do not contain the restoring transgene construct. Increase of the maintainer plant is also provided by self-fertilization, and selection for seed or plants which contain the construct.

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.

THIS IS VOLUME 1 OF 2

NOTE: For additional volumes please contact the Canadian Patent Office.

NUCLEOTIDE SEQUENCES MEDIATING PLANT MALE FERTILITY AND METHOD OF USING SAME

5 BACKGROUND OF THE INVENTION

Development of hybrid plant breeding has made possible considerable advances in quality and quantity of crops produced. Increased yield and combination of desirable characteristics, such as resistance to disease and insects, heat and drought tolerance, along with variations in plant composition are
10 all possible because of hybridization procedures. These procedures frequently rely heavily on providing for a male parent contributing pollen to a female parent to produce the resulting hybrid.

Field crops are bred through techniques that take advantage of the plant's method of pollination. A plant is self-pollinating if pollen from one flower is
15 transferred to the same or another flower of the same plant. A plant is cross-pollinated if the pollen comes from a flower on a different plant.

In *Brassica*, the plant is normally self-sterile and can only be cross-pollinated. In self-pollinating species, such as soybeans and cotton, the male and female plants are anatomically juxtaposed. During natural pollination, the male
20 reproductive organs of a given flower pollinate the female reproductive organs of the same flower.

Maize plants (*Zea mays* L.) present a unique situation in that they can be bred by both self-pollination and cross-pollination techniques. Maize has male flowers, located on the tassel, and female flowers, located on the ear, on the
25 same plant. It can self or cross pollinate. Natural pollination occurs in maize when wind blows pollen from the tassels to the silks that protrude from the tops of the incipient ears.

A reliable method of controlling fertility in plants would offer the opportunity for improved plant breeding. This is especially true for development of maize
30 hybrids, which relies upon some sort of male sterility system and where a female sterility system would reduce production costs.

The development of maize hybrids requires the development of homozygous inbred lines, the crossing of these lines, and the evaluation of the crosses. Pedigree breeding and recurrent selection are two of the breeding
35 methods used to develop inbred lines from populations. Breeding programs

combine desirable traits from two or more inbred lines or various broad-based sources into breeding pools from which new inbred lines are developed by selfing and selection of desired phenotypes. A hybrid maize variety is the cross of two such inbred lines, each of which may have one or more desirable characteristics
5 lacked by the other or which complement the other. The new inbreds are crossed with other inbred lines and the hybrids from these crosses are evaluated to determine which have commercial potential. The hybrid progeny of the first generation is designated F_1 . In the development of hybrids only the F_1 hybrid plants are sought. The F_1 hybrid is more vigorous than its inbred parents. This
10 hybrid vigor, or heterosis, can be manifested in many ways, including increased vegetative growth and increased yield.

Hybrid maize seed can be produced by a male sterility system incorporating manual detasseling. To produce hybrid seed, the male tassel is removed from the growing female inbred parent, which can be planted in various
15 alternating row patterns with the male inbred parent. Consequently, providing that there is sufficient isolation from sources of foreign maize pollen, the ears of the female inbred will be fertilized only with pollen from the male inbred. The resulting seed is therefore hybrid (F_1) and will form hybrid plants.

Environmental variation in plant development can result in plants tasseling
20 after manual detasseling of the female parent is completed. Or, a detasseler might not completely remove the tassel of a female inbred plant. In any event, the result is that the female plant will successfully shed pollen and some female plants will be self-pollinated. This will result in seed of the female inbred being harvested along with the hybrid seed which is normally produced. Female inbred
25 seed is not as productive as F_1 seed. In addition, the presence of female inbred seed can represent a germplasm security risk for the company producing the hybrid.

Alternatively, the female inbred can be mechanically detasseled by machine. Mechanical detasseling is approximately as reliable as hand
30 detasseling, but is faster and less costly. However, most detasseling machines produce more damage to the plants than hand detasseling. Thus, no form of detasseling is presently entirely satisfactory, and a need continues to exist for alternatives which further reduce production costs and to eliminate self-pollination of the female parent in the production of hybrid seed.

A reliable system of genetic male sterility would provide advantages. The laborious detasseling process can be avoided in some genotypes by using cytoplasmic male-sterile (CMS) inbreds. In the absence of a fertility restorer gene, plants of a CMS inbred are male sterile as a result of factors resulting from the cytoplasmic, as opposed to the nuclear, genome. Thus, this characteristic is inherited exclusively through the female parent in maize plants, since only the female provides cytoplasm to the fertilized seed. CMS plants are fertilized with pollen from another inbred that is not male-sterile. Pollen from the second inbred may or may not contribute genes that make the hybrid plants male-fertile. Usually seed from detasseled normal maize and CMS produced seed of the same hybrid must be blended to insure that adequate pollen loads are available for fertilization when the hybrid plants are grown and to insure cytoplasmic diversity.

There can be other drawbacks to CMS. One is an historically observed association of a specific variant of CMS with susceptibility to certain crop diseases. This problem has discouraged widespread use of that CMS variant in producing hybrid maize and has had a negative impact on the use of CMS in maize in general.

One type of genetic sterility is disclosed in U.S. Patents 4,654,465 and 4,727,219 to Brar, et al. However, this form of genetic male sterility requires maintenance of multiple mutant genes at separate locations within the genome and requires a complex marker system to track the genes and make use of the system convenient. Patterson also described a genic system of chromosomal translocations which can be effective, but which are complicated. (See, U.S. Patents No. 3,861,709 and 3,710,511.)

Many other attempts have been made to improve on these drawbacks. For example, Fabijanski, et al., developed several methods of causing male sterility in plants (see EPO 89/3010153.8 publication no. 329,308 and PCT application PCT/CA90/00037 published as WO 90/08828). One method includes delivering into the plant a gene encoding a cytotoxic substance associated with a male tissue specific promoter. Another involves an antisense system in which a gene critical to fertility is identified and an antisense to the gene inserted in the plant. Mariani, et al. also shows several cytotoxic antisense systems. See EP 89/401, 194. Still other systems use "repressor" genes which inhibit the expression of

another gene critical to male sterility. PCT/GB90/00102, published as WO 90/08829.

5 A still further improvement of this system is one described at U.S. Patent No. 5,478,369 in which a method of imparting controllable male sterility is achieved by silencing a gene native to the plant that is critical for male fertility and replacing the native DNA with the gene critical to male fertility linked to an inducible promoter controlling expression of the gene. The plant is thus constitutively sterile, becoming fertile only when the promoter is induced and its attached male fertility gene is expressed.

10 In a number of circumstances, a male sterility plant trait is expressed by maintenance of a homozygous recessive condition. Difficulties arise in maintaining the homozygous condition, when a transgenic restoration gene must be used for maintenance. For example, a natural mutation in a gene critical to male sterility can impart a male sterility phenotype to plants when this mutant
15 allele is in the homozygous state. This sterility can be restored when the non-mutant form of the gene is introduced into the plant. However, this form of restoration removes the desired homozygous recessive condition, restores full male fertility and prevents maintenance of pure male sterile maternal lines. This issue can be avoided where production of pollen containing the restoration gene is eliminated, thus providing a maintainer plant producing only pollen not
20 containing the restoration gene, and the progeny retains the homozygous condition. An example of one approach is shown in Dellaporta et al., 6,743,968, in which a plant is produced having a hemizygotic construct comprising a gene that produces a product fatal to a cell, linked with a pollen-specific promoter, and the restoration gene. When crossed with the homozygous recessive male sterile
25 plant, the progeny thus retains the homozygous recessive condition.

As noted, an essential aspect of much of the work underway with male sterility systems is the identification of genes impacting male fertility.

30 Such a gene can be used in a variety of systems to control male fertility including those described herein. Previously, a male fertility gene has been identified in *Arabidopsis thaliana* and used to produce a male sterile plant. Aarts, et al., "Transposon Tagging of a Male Sterility Gene in *Arabidopsis*", *Nature*, 363:715-717 (Jun. 24, 1993). U.S. Patent No. 5,478,369 discloses therein one such gene impacting male fertility. In the present invention the inventors provide

novel DNA molecules and the amino acid sequence encoded that are critical to male fertility in plants. These can be used in any of the systems where control of fertility is useful, including those described above.

Thus, one object of the invention is to provide a nucleic acid sequence, the expression of which is critical to male fertility in plants.

Another object of the invention is to provide a DNA molecule encoding an amino acid sequence, the expression of which is critical to male fertility in plants.

Yet another object of the invention is to provide a promoter of such nucleotide sequence and its essential sequences.

A further object of the invention is to provide a method of using such DNA molecules to mediate male fertility in plants.

Further objects of the invention will become apparent in the description and claims that follow.

15 SUMMARY OF THE INVENTION

This invention relates to nucleic acid sequences, and, specifically, DNA molecules and the amino acid encoded by the DNA molecules, which are critical to male fertility. A promoter of the DNA is identified, as well as its essential sequences. It also relates to use of such DNA molecules to mediate fertility in plants.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a locus map of the male fertility gene *Ms26*.

Figure 2A is a Southern blot of the *ms26-m2::Mu8* family hybridized with a *Mu8* probe; Figure 2B is a Southern blot of the *ms26-m2::Mu8* family hybridized with a *PstI* fragment isolated from the *ms26* clone.

Figure 3. is a Northern Blot analysis gel hybridized with a *PstI* fragment isolated from the *Ms26* gene.

Figure 4A-4D is the sequence of *Ms26* (The cDNA is SEQ ID NO: 1, the protein is SEQ ID NOS: 2 and 34)

Figure 5A-5C is the genomic *Ms26* sequence (also referred to as SEQ ID NO: 7).

Figure 6A-6D is a comparison of the genomic *Ms26* sequence (Residues 1051-1450, 1501-2100 and 2201-3326 of SEQ ID NO: 7) with the cDNA of *Ms26* (SEQ ID NO: 1).

Figure 7A is a Northern analysis gel showing expression in various plant tissues and Figure 7B is a gel showing expression stages of microsporogenesis

Figure 8 is the full length promoter of *Ms26* (SEQ ID NO: 5)

Figure 9 is a bar graph showing luciferase activity after deletions of select regions
5 of the *Ms26* promoter.

Figure 10 shows essential regions of the *Ms26* promoter (SEQ ID NO: 6).

Figure 11 is a bar graph showing luciferase activity after substitution by restriction site linker scanning of select small (9-10bp) regions of the *Ms26* essential promoter fragment.

10 Figure 12A and 12B is a comparison of the nucleotide sequence (SEQ ID NO: 3) from the *Ms26* orthologue from a sorghum panicle and *Ms26* maize cDNA (Residues 201-750 of SEQ ID NO: 1), and the sorghum protein sequence (SEQ ID NO: 4) and *Ms26* maize protein (Residues 87-244 of SEQ ID NO: 2).

Figure 13 is a representation of the mapping of the male sterility gene *ms26*.

15 Figure 14 shows a sequence comparison of the region of excision of the *ms26-ref* allele (SEQ ID NO: 8) with wild-type *Ms26* (SEQ ID NO: 9).

Figure 15 shows the transposon sequence within *ms26-ref* (SEQ ID NO: 10).

Figure 16 shows the entire *ms26-ref* sequence (SEQ ID NO: 11).

Figure 17A shows a translated protein sequence alignment between regions of
20 the CYP704B1, a P450 gene (SEQ ID NO: 12) and *Ms26* (SEQ ID NO: 13);

Figure 17B shows the phylogenetic tree analysis of select P450 genes.

Figure 18 demonstrates the heme binding domain frame shift, showing the translated sequence alignment of regions of the *Ms26* cDNA (SEQ ID NOS: 14 and 28-29), the genomic regions of exon 5 in fertile plants (SEQ ID NOS: 15 and
25 30-31) and sterile plants (SEQ ID NOS: 16 and 32-33).

Figure 19 shows the rice *Ms26* cDNA (SEQ ID NO: 17) and protein (SEQ ID NO: 18).

Figure 20 shows alignment of the *Ms26* promoter of corn (Residues 650-1091 of SEQ ID NO: 5), sorghum (SEQ ID NO: 19) and rice (SEQ ID NO: 20).

30 Figure 21 shows alignment of the maize *Ms26* protein (SEQ ID NO: 21); rice *Ms26* protein (SEQ ID NO: 18) and sorghum *Ms26* protein (SEQ ID NO: 22) along with a consensus sequence.

Figure 22 is a plasmid map of PHP 18091, containing *Ms45* fertility gene with a pollen promoter, cytotoxic gene and selectable marker.

Figure 23 is a plasmid map of PHP 24101, containing the *Ms26* fertility gene with a pollen promoter, cytotoxic gene and selectable marker.

Figure 24 shows a sequence of the *Zea mays* α -amylase 1 coding region (SEQ ID NOS: 26 (DNA) and 36 (protein)).

5

DISCLOSURE OF THE INVENTION

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Unless mentioned otherwise, the techniques employed or contemplated therein are standard methodologies well known to one of ordinary skill in the art. The materials, methods and examples are illustrative only and not limiting.

Genetic male sterility results from a mutation, suppression, or other impact to one of the genes critical to a specific step in microsporogenesis, the term applied to the entire process of pollen formulation. These genes can be collectively referred to as male fertility genes (or, alternatively, male sterility genes). There are many steps in the overall pathway where gene function impacts fertility. This seems aptly supported by the frequency of genetic male sterility in maize. New alleles of male sterility mutants are uncovered in materials that range from elite inbreds to unadapted populations.

At Patent No. 5,478,369 there is described a method by which the *Ms45* male sterility gene was tagged and cloned on maize chromosome 9. Previously, there had been described a male sterility gene on chromosome 9, *ms2*, which had never been cloned and sequenced. It is not allelic to the gene referred to in the '369 patent. See Albertsen, M. and Phillips, R.L., "Developmental Cytology of 13 Genetic Male Sterile Loci in Maize" *Canadian Journal of Genetics & Cytology* 23:195-208 (Jan. 1981). The only fertility gene cloned before that had been the *Arabidopsis* gene described at Aarts, et al., *supra*.

Examples of genes that have been discovered subsequently that are critical to male fertility are numerous and include the *Arabidopsis* ABORTED MICROSPORES (AMS) gene, Sorensen et al., *The Plant Journal* (2003) 33(2):413-423; the *Arabidopsis* MS1 gene (Wilson et al., *The Plant Journal* (2001) 39(2):170-181); the NEF1 gene (Ariizumi et al., *The Plant Journal* (2004)

39(2):170-181); *Arabidopsis* AtGPAT1 gene (Zheng et al., *The Plant Cell* (2003) 15:1872-1887); the *Arabidopsis* dde2-2 mutation was shown to be defective in the allene oxide synthase gene (Malek et al., *Planta* (2002)216:187-192); the *Arabidopsis* faceless pollen-1 gene (flp1) (Ariizumi et al, *Plant Mol. Biol.* (2003) 53:107-116); the *Arabidopsis* MALE MEIOCYTE DEATH1 gene (Yang et al., *The Plant Cell* (2003) 15: 1281-1295); the tapetum-specific zinc finger gene, TAZ1 (Kapoor et al., *The Plant Cell* (2002) 14:2353-2367); and the TAPETUM DETERMINANT1 gene (Lan et al, *The Plant Cell* (2003) 15:2792-2804).

The table below lists a number of known male fertility mutants or genes from *Zea mays*.

GENE NAME	ALTERNATE NAME	REFERENCE
ms1 male sterile1	<i>male sterile1, ms1</i>	Singleton, WR and Jones, DF. 1930. J Hered 21:266-268
ms10 male sterile10	<i>male sterile10, ms10</i>	Beadle, GW. 1932. Genetics 17:413-431
ms11 male sterile11	<i>ms11, male sterile11</i>	Beadle, GW. 1932. Genetics 17:413-431
ms12 male sterile12	<i>ms12, male sterile12</i>	Beadle, GW. 1932. Genetics 17:413-431
ms13 male sterile13	<i>ms*-6060, male sterile13, ms13</i>	Beadle, GW. 1932. Genetics 17:413-431
ms14 male sterile14	<i>ms14, male sterile14</i>	Beadle, GW. 1932. Genetics 17:413-431
ms17 male sterile17	<i>ms17, male sterile17</i>	Emerson, RA. 1932. Science 75:566
ms2 male sterile2	<i>male sterile2, ms2</i>	Eyster, WH. 1931. J Hered 22:99-102
ms20 male sterile20	<i>ms20, male sterile20</i>	Eyster, WH. 1934. Genetics of Zea mays. Bibliographia Genetica 11:187-392
ms23 male sterile23	: <i>ms*-6059, ms*-6031, ms*-6027, ms*-6018, ms*-6011, ms35, male sterile23, ms*-Bear7, ms23</i>	West, DP and Albertsen, MC. 1985. MNL 59:87
ms24 male sterile24	<i>ms24, male sterile24</i>	West, DP and Albertsen, MC. 1985. MNL 59:87
ms25 male sterile25	<i>ms*-6065, ms*-6057, ms25, male sterile25, ms*-6022</i>	Loukides, CA; Broadwater, AH; Bedinger, PA. 1995. Am J Bot 82:1017-1023
ms27 male sterile27	<i>ms27, male sterile27</i>	Albertsen, MC. 1996. MNL 70:30-31

ms28 male sterile28	<i>ms28, male sterile28</i>	Golubovskaya, IN. 1979. MNL 53:66-70
ms29 male sterile29	<i>male sterile29, ms*-JH84A, ms29</i>	Trimnell, MR et al. 1998. MNL 72:37-38
ms3 male sterile3	<i>Group 3, ms3, male sterile3</i>	Eyster, WH. 1931. J Hered 22:99-102
ms30 male sterile30	<i>ms30, msx, ms*-6028, ms*-Li89, male sterile30, ms*-Li89</i>	Albertsen, MC et al. 1999. MNL 73:48
ms31 male sterile31	<i>ms*-CG889D, ms31, male sterile31</i>	Trimnell, MR et al. 1998. MNL 72:38
ms32 male sterile32	<i>male sterile32, ms32</i>	Trimnell, MR et al. 1999. MNL 73:48-49
ms33 male sterile33	<i>: ms*-6054, ms*-6024, ms33, ms*-GC89A, ms*-6029, male sterile6019, Group 7, ms*-6038, ms*-Stan1, ms*-6041, ms*-6019, male sterile33</i>	Patterson, EB. 1995. MNL 69:126-128
ms34 male sterile34	<i>Group 1, ms*-6014, ms*-6010, male sterile34, ms34, ms*-6013, ms*-6004, male sterile6004</i>	Patterson, EB. 1995. MNL 69:126-128
ms36 male sterile36	<i>male sterile36, ms*-MS85A, ms36</i>	Trimnell, MR et al. 1999. MNL 73:49-50
ms37 male sterile 37	<i>ms*-SB177, ms37, male sterile 37</i>	Trimnell, MR et al. 1999. MNL 73:48
ms38 male sterile38	<i>ms30, ms38, ms*-WL87A, male sterile38</i>	Albertsen, MC et al. 1996. MNL 70:30
ms43 male sterile43	<i>ms43, male sterile43, ms29</i>	Golubovskaya, IN. 1979. Int Rev Cytol 58:247-290
ms45 male sterile45	<i>Group 6, male sterile45, ms*-6006, ms*-6040, ms*-BS1, ms*-BS2, ms*-BS3, ms45, ms45'-9301</i>	Albertsen, MC; Fox, TW; Trimnell, MR. 1993. Proc Annu Corn Sorghum Ind Res Conf 48:224-233
ms48 male sterile48	<i>male sterile48, ms*-6049, ms48</i>	Trimnell, M et al. 2002. MNL 76:38
ms5 male sterile5	<i>: ms*-6061, ms*-6048, ms*-6062, male sterile5, ms5</i>	Beadle, GW. 1932. Genetics 17:413-431
ms50 male sterile50	<i>ms50, male sterile50, ms*-6055, ms*-6026</i>	Trimnell, M et al. 2002. MNL 76:39
ms7 male sterile7	<i>ms7, male sterile7</i>	Beadle, GW. 1932. Genetics 17:413-431
ms8 male sterile8	<i>male sterile8, ms8</i>	Beadle, GW. 1932. Genetics 17:413-431
ms9 male sterile9	<i>Group 5, male sterile9, ms9</i>	Beadle, GW. 1932. Genetics 17:413-431

Thus the invention includes using the sequences shown herein to impact male fertility in a plant, that is, to control male fertility by manipulation of the genome using the genes of the invention. By way of example, without limitation, any of the methods described *infra* can be used with the sequence of the invention such as introducing a mutant sequence into a plant to cause sterility, causing mutation to the native sequence, introducing an antisense of the sequence into the plant, use of hairpin formations, linking it with other sequences to control its expression, or any one of a myriad of processes available to one skilled in the art to impact male fertility in a plant.

The *Ms26* gene described herein is located on maize chromosome 1 and its dominant allele is critical to male fertility. The locus map is represented at Figure 1. It can be used in the systems described above, and other systems impacting male fertility.

The maize family cosegregating for sterility was named *ms**-SBMu200 and was found to have an approximately 5.5 Kb *EcoRI* fragment that hybridized with a *Mu8* probe (Figure 2A). A genomic clone from the family was isolated which contained a *Mu8* transposon. A probe made from DNA bordering the transposon was found to hybridize to the same ~5.5Kb *EcoRI* fragment (Figure 2B). This probe was used to isolate cDNA clones from a tassel cDNA library. The cDNA is 1906 bp, and the *Mu* insertion occurred in exon 1 of the gene. This probe was also used to map the mutation in an RFLP mapping population. The mutant mapped to the short arm of chromosome 1, near *Ms26*. Allelism crosses between *ms26-ref* and *ms**-SBMu200 showed that these were allelic, indicating that the mutations occurred in the same gene. The *ms**-SBMu200 allele was renamed *ms26-m2::Mu8*. Two additional alleles for the *Ms26* gene were cloned, one containing a Mutator element in the second exon, named *ms26-m3::Mu**, and one containing an unknown transposon in the fifth exon from the *ms26-ref* allele. Figure 5 (discussed further below) represents the genomic nucleotide sequence. Expression patterns, as determined by Northern analysis, show tassel specificity with peak expression at about the quartet to quartet release stages of microsporogenesis.

It will be evident to one skilled in the art that variations, mutations, derivations including fragments smaller than the entire sequence set forth may be used which retain the male sterility controlling properties of the gene. One of

ordinary skill in the art can readily assess the variant or fragment by introduction into plants homozygous for a stable male sterile allele of *Ms26*, followed by observation of the plant's male tissue development.

The sequences of the invention may be isolated from any plant, including,
 5 but not limited to corn (*Zea mays*), canola (*Brassica napus*, *Brassica rapa ssp.*), alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), sunflower (*Helianthus annuus*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), millet (*Panicum spp.*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton
 10 (*Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea spp.*), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus spp.*), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa spp.*), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), oats
 15 (*Avena sativa*), barley (*Hordeum vulgare*), vegetables, ornamentals, and conifers. Preferably, plants include corn, soybean, sunflower, safflower, canola, wheat, barley, rye, alfalfa, rice, cotton and sorghum.

Sequences from other plants may be isolated according to well-known techniques based on their sequence homology to the homologous coding region
 20 of the coding sequences set forth herein. In these techniques, all or part of the known coding sequence is used as a probe which selectively hybridizes to other sequences present in a population of cloned genomic DNA fragments (i.e. genomic libraries) from a chosen organism. Methods are readily available in the art for the hybridization of nucleic acid sequences. An extensive guide to the
 25 hybridization of nucleic acids is found in Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993); and Current Protocols in Molecular Biology, Chapter 2, Ausubel, et al., Eds., Greene Publishing and Wiley-
 30 Interscience, New York (1995).

Thus the invention also includes those nucleotide sequences which selectively hybridize to the *Ms26* nucleotide sequences under stringent conditions. In referring to a sequence that "selectively hybridizes" with *Ms26*, the term includes reference to hybridization, under stringent hybridization conditions, of a

nucleic acid sequence to the specified nucleic acid target sequence to a detectably greater degree than its hybridization to non-target nucleic acid.

The terms "stringent conditions" or "stringent hybridization conditions" includes reference to conditions under which a probe will hybridize to its target
 5 sequence, to a detectably greater degree than to other sequences. Stringent conditions are target-sequence-dependent and will differ depending on the structure of the polynucleotide. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to a probe (homologous probing). Alternatively, stringency
 10 conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, probes of this type are in a range of about 1000 nucleotides in length to about 250 nucleotides in length.

An extensive guide to the hybridization of nucleic acids is found in Tijssen,
 15 Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993); and Current Protocols in Molecular Biology, Chapter 2, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995). See also Sambrook et al.
 20 (1989) Molecular Cloning: A Laboratory Manual (2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

In general, sequences that correspond to the nucleotide sequences of the present invention and hybridize to the nucleotide sequence disclosed herein will be at least 50% homologous, 70% homologous, and even 85%, 86%, 87%, 88%,
 25 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% homologous or more with the disclosed sequence. That is, the sequence similarity between probe and target may range, sharing at least about 50%, about 70%, and even about 85% or more sequence similarity.

Specificity is typically the function of post-hybridization washes, the critical
 30 factors being the ionic strength and temperature of the final wash solution. Generally, stringent wash temperature conditions are selected to be about 5°C to about 2°C lower than the melting point (T_m) for the specific sequence at a defined ionic strength and pH. The melting point, or denaturation, of DNA occurs over a narrow temperature range and represents the disruption of the double helix into its

complementary single strands. The process is described by the temperature of the midpoint of transition, T_m , which is also called the melting temperature. Formulas are available in the art for the determination of melting temperatures.

Preferred hybridization conditions for the nucleotide sequence of the invention include hybridization at 42°C in 50%(w/v) formamide, 6X SSC, 0.5%(w/v) SDS, 100(g/ml salmon sperm DNA. Exemplary low stringency washing conditions include hybridization at 42°C in a solution of 2X SSC, 0.5% (w/v) SDS for 30 minutes and repeating. Exemplary moderate stringency conditions include a wash in 2X SSC, 0.5% (w/v) SDS at 50°C for 30 minutes and repeating. Exemplary high stringency conditions include a wash in 0.1X SSC, 0.1% (w/v) SDS, at 65°C for 30 minutes to one hour and repeating. Sequences that correspond to the promoter of the present invention may be obtained using all the above conditions. For purposes of defining the invention, the high stringency conditions are used.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", and (d) "percentage of sequence identity."

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, or 100 nucleotides in length, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of aligning sequences for comparison are well-known in the art. Thus, the determination of percent sequence identity between any two sequences

can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988) *CABIOS* 4: 11-17; the local alignment algorithm of Smith *et al.* (1981) *Adv. Appl. Math.* 2: 482; the global alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443-453; the search-for-local-alignment-method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci.* 85: 2444-2448; the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87: 2264, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5877.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the GCG Wisconsin Genetics Software Package, Version 10 (available from Accelrys Inc., 9685 Scranton Road, San Diego, California, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins *et al.* (1988) *Gene* 73: 237-244 (1988); Higgins *et al.* (1989) *CABIOS* 5: 151-153; Corpet *et al.* (1988) *Nucleic Acids Res.* 16: 10881-90; Huang *et al.* (1992) *CABIOS* 8: 155-65; and Pearson *et al.* (1994) *Meth. Mol. Biol.* 24: 307-331. The ALIGN program is based on the algorithm of Myers and Miller (1988) *supra*. A PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used with the ALIGN program when comparing amino acid sequences. The BLAST programs of Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403 are based on the algorithm of Karlin and Altschul (1990) *supra*. BLAST nucleotide searches can be performed with the BLASTN program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the invention. BLAST protein searches can be performed with the BLASTX program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to a protein or polypeptide of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25: 3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.* (1997) *supra*. When utilizing BLAST, Gapped

BLAST, PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins) can be used.

Alignment may also be performed manually by inspection.

5 Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 using the following parameters: % identity and % similarity for a nucleotide sequence using GAP Weight of 50 and Length Weight of 3 and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using GAP Weight of 8 and Length Weight
10 of 2; and the BLOSUM62 scoring matrix or any equivalent program thereof. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10.

15 GAP uses the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443-453, to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a
20 gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty
25 values in Version 10 of the GCG Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 200. Thus, for
30 example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity.

The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A
5 similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the GCG Wisconsin Genetics Software Package is BLOSUM62 (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

(c) As used herein, "sequence identity" or "identity" in the context of two
10 nucleic acid or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues
15 are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such
20 conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-
25 conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, "percentage of sequence identity" means the value
30 determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the

number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence
5 identity.

The use of the term "polynucleotide" is not intended to limit the present invention to polynucleotides comprising DNA. Those of ordinary skill in the art will recognize that polynucleotides can comprise ribonucleotides and combinations of ribonucleotides and deoxyribonucleotides. Such deoxyribonucleotides and
10 ribonucleotides include both naturally occurring molecules and synthetic analogues. The polynucleotides of the invention also encompass all forms of sequences including, but not limited to, single-stranded forms, double-stranded forms, hairpins, stem-and-loop structures, and the like.

Identity to the sequence of the present invention would mean a
15 polynucleotide sequence having at least 65% sequence identity, more preferably at least 70% sequence identity, more preferably at least 75% sequence identity, more preferably at least 80% identity, more preferably at least 85% 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity.

20 Promoter regions can be readily identified by one skilled in the art. The putative start codon containing the ATG motif is identified and upstream from the start codon is the presumptive promoter. By "promoter" is intended a regulatory region of DNA usually comprising a TATA box capable of directing RNA polymerase II to initiate RNA synthesis at the appropriate transcription initiation
25 site for a particular coding sequence. A promoter can additionally comprise other recognition sequences generally positioned upstream or 5' to the TATA box, referred to as upstream promoter elements, which influence the transcription initiation rate. It is recognized that having identified the nucleotide sequences for the promoter region disclosed herein, it is within the state of the art to isolate and
30 identify further regulatory elements in the region upstream of the TATA box from the particular promoter region identified herein. Thus the promoter region disclosed herein is generally further defined by comprising upstream regulatory elements such as those responsible for tissue and temporal expression of the coding sequence, enhancers and the like. In the same manner, the promoter

elements which enable expression in the desired tissue such as male tissue can be identified, isolated, and used with other core promoters to confirm male tissue-preferred expression. By core promoter is meant the minimal sequence required to initiate transcription, such as the sequence called the TATA box which is
5 common to promoters in genes encoding proteins. Thus the upstream promoter of *Ms26* can optionally be used in conjunction with its own or core promoters from other sources. The promoter may be native or non-native to the cell in which it is found.

The isolated promoter sequence of the present invention can be modified
10 to provide for a range of expression levels of the heterologous nucleotide sequence. Less than the entire promoter region can be utilized and the ability to drive anther-preferred expression retained. However, it is recognized that expression levels of mRNA can be decreased with deletions of portions of the promoter sequence. Thus, the promoter can be modified to be a weak or strong
15 promoter. Generally, by "weak promoter" is intended a promoter that drives expression of a coding sequence at a low level. By "low level" is intended levels of about 1/10,000 transcripts to about 1/100,000 transcripts to about 1/500,000 transcripts. Conversely, a strong promoter drives expression of a coding sequence at a high level, or at about 1/10 transcripts to about 1/100 transcripts to
20 about 1/1,000 transcripts. Generally, at least about 30 nucleotides of an isolated promoter sequence will be used to drive expression of a nucleotide sequence. It is recognized that to increase transcription levels, enhancers can be utilized in combination with the promoter regions of the invention. Enhancers are nucleotide sequences that act to increase the expression of a promoter region. Enhancers
25 are known in the art and include the SV40 enhancer region, the 35S enhancer element, and the like.

The promoter of the present invention can be isolated from the 5' region of its native coding region of 5' untranslation region (5'UTR). Likewise the terminator can be isolated from the 3' region flanking its respective stop codon. The term
30 "isolated" refers to material such as a nucleic acid or protein which is substantially or essentially free from components which normally accompany or interact with the material as found in its naturally occurring environment or if the material is in its natural environment, the material has been altered by deliberate human intervention to a composition and/or placed at a locus in a cell other than the locus

native to the material. Methods for isolation of promoter regions are well known in the art.

"Functional variants" of the regulatory sequences are also encompassed by the compositions of the present invention. Functional variants include, for example, the native regulatory sequences of the invention having one or more nucleotide substitutions, deletions or insertions. Functional variants of the invention may be created by site-directed mutagenesis, induced mutation, or may occur as allelic variants (polymorphisms).

As used herein, a "functional fragment" of the regulatory sequence is a nucleotide sequence that is a regulatory sequence variant formed by one or more deletions from a larger sequence. For example, the 5' portion of a promoter up to the TATA box near the transcription start site can be deleted without abolishing promoter activity, as described by Opsahl-Sorteberg, H-G. et al., "Identification of a 49-bp fragment of the HvLTP2 promoter directing aleruone cell specific expression" *Gene* 341:49-58 (2004). Such variants should retain promoter activity, particularly the ability to drive expression in male tissues. Activity can be measured by Northern blot analysis, reporter activity measurements when using transcriptional fusions, and the like. See, for example, Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Functional fragments can be obtained by use of restriction enzymes to cleave the naturally occurring regulatory element nucleotide sequences disclosed herein; by synthesizing a nucleotide sequence from the naturally occurring DNA sequence; or can be obtained through the use of PCR technology. See particularly, Mullis et al. (1987) *Methods Enzymol.* 155:335-350, and Erlich, ed. (1989) PCR Technology (Stockton Press, New York).

Sequences which hybridize to the regulatory sequences of the present invention are within the scope of the invention. Sequences that correspond to the promoter sequences of the present invention and hybridize to the promoter sequences disclosed herein will be at least 50% homologous, 70% homologous, and even 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% homologous or more with the disclosed sequence.

Smaller fragments may yet contain the regulatory properties of the promoter so identified and deletion analysis is one method of identifying essential

regions. Deletion analysis can occur from both the 5' and 3' ends of the regulatory region. Fragments can be obtained by site-directed mutagenesis, mutagenesis using the polymerase chain reaction and the like. (See, Directed Mutagenesis: A Practical Approach IRL Press (1991)). The 3' deletions can
5 delineate the essential region and identify the 3' end so that this region may then be operably linked to a core promoter of choice. Once the essential region is identified, transcription of an exogenous gene may be controlled by the essential region plus a core promoter. By core promoter is meant the sequence called the TATA box which is common to promoters in all genes encoding proteins. Thus
10 the upstream promoter of *Ms26* can optionally be used in conjunction with its own or core promoters from other sources. The promoter may be native or non-native to the cell in which it is found.

The core promoter can be any one of known core promoters such as the Cauliflower Mosaic Virus 35S or 19S promoter (U.S. Patent No. 5,352,605),
15 ubiquitin promoter (U.S. Patent No. 5,510,474) the IN2 core promoter (U.S. Patent No. 5,364,780) or a Figwort Mosaic Virus promoter (Gruber, et al. "Vectors for Plant Transformation" Methods in Plant Molecular Biology and Biotechnology) et al. eds, CRC Press pp.89-119 (1993)).

The regulatory region of *Ms26* has been identified as including the 1005 bp
20 region upstream of the putative TATA box. See Figure 8. Further, using the procedures outlined above, it has been determined that an essential region of the promoter includes the -180 bp upstream of the TATA box and specifically, the -176 to -44 region is particularly essential.

Promoter sequences from other plants may be isolated according to well-
25 known techniques based on their sequence homology to the promoter sequence set forth herein. In these techniques, all or part of the known promoter sequence is used as a probe which selectively hybridizes to other sequences present in a population of cloned genomic DNA fragments (i.e. genomic libraries) from a chosen organism. Methods are readily available in the art for the hybridization of
30 nucleic acid sequences.

The entire promoter sequence or portions thereof can be used as a probe capable of specifically hybridizing to corresponding promoter sequences. To achieve specific hybridization under a variety of conditions, such probes include sequences that are unique and are preferably at least about 10 nucleotides in

length, and most preferably at least about 20 nucleotides in length. Such probes can be used to amplify corresponding promoter sequences from a chosen organism by the well-known process of polymerase chain reaction (PCR). This technique can be used to isolate additional promoter sequences from a desired
 5 organism or as a diagnostic assay to determine the presence of the promoter sequence in an organism. Examples include hybridization screening of plated DNA libraries (either plaques or colonies; see e.g. Innis et al., eds., (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press).

Further, a promoter of the present invention can be linked with nucleotide
 10 sequences other than the *Ms26* gene to express other heterologous nucleotide sequences. The nucleotide sequence for the promoter of the invention, as well as fragments and variants thereof, can be provided in expression cassettes along with heterologous nucleotide sequences for expression in the plant of interest, more particularly in the male tissue of the plant. Such an expression cassette is
 15 provided with a plurality of restriction sites for insertion of the nucleotide sequence to be under the transcriptional regulation of the promoter. These expression cassettes are useful in the genetic manipulation of any plant to achieve a desired phenotypic response.

Examples of other nucleotide sequences which can be used as the
 20 exogenous gene of the expression vector with the *Ms26* promoter, or other promoters taught herein or known to those of skill in the art, or other promoters taught herein or known to those of skill in the art complementary nucleotidic units such as antisense molecules (callase antisense RNA, barnase antisense RNA and chalcone synthase antisense RNA, *Ms45* antisense RNA), ribozymes and
 25 external guide sequences, an aptamer or single stranded nucleotides. The exogenous nucleotide sequence can also encode carbohydrate degrading or modifying enzymes, amylases, debranching enzymes and pectinases, such as the alpha amylase gene of Figure 24, auxins, rol B, cytotoxins, diptheria toxin, DAM methylase, avidin, or may be selected from a prokaryotic regulatory system. By
 30 way of example, Mariani, et al., *Nature* Vol. 347; pp. 737; (1990), have shown that expression in the tapetum of either *Aspergillus oryzae* RNase-T1 or an RNase of *Bacillus amyloliquefaciens*, designated "barnase," induced destruction of the tapetal cells, resulting in male infertility. Quaas, et al., *Eur. J. Biochem.* Vol. 173: pp. 617 (1988), describe the chemical synthesis of the RNase-T1, while the nucleotide

sequence of the barnase gene is disclosed in Hartley, *J. Molec. Biol.*; Vol. 202: pp. 913 (1988). The *rolB* gene of *Agrobacterium rhizogenes* codes for an enzyme that interferes with auxin metabolism by catalyzing the release of free indoles from indoxyl- β -glucosides. Estruch, et al., *EMBO J.* Vol. 11: pp. 3125 (1991) and Spena, et al., *Theor. Appl. Genet.*; Vol. 84: pp. 520 (1992), have shown that the anther-specific expression of the *rolB* gene in tobacco resulted in plants having shriveled anthers in which pollen production was severely decreased and the *rolB* gene is an example of a gene that is useful for the control of pollen production. Slightom, et al., *J. Biol. Chem.* Vol. 261: pp. 108 (1985), disclose the nucleotide sequence of the *rolB* gene. DNA molecules encoding the diphtheria toxin gene can be obtained from the American Type Culture Collection (Rockville, MD), ATCC No. 39359 or ATCC No. 67011 and see Fabijanski, et al., E.P. Appl. No. 90902754.2, "Molecular Methods of Hybrid Seed Production" for examples and methods of use. The DAM methylase gene is used to cause sterility in the methods discussed at U.S. Patent No. 5,689,049 and PCT/US95/15229 Cigan, A.M. and Albertsen, M.C., "Reversible Nuclear Genetic System for Male Sterility in Transgenic Plants." Also see discussion of use of the avidin gene to cause sterility at U.S. Patent No. 5,962,769 "Induction of Male Sterility in Plants by Expression of High Levels of Avidin" by Albertsen et al.

The invention includes vectors with the *Ms26* gene. A vector is prepared comprising *Ms26*, a promoter that will drive expression of the gene in the plant and a terminator region. As noted, the promoter in the construct may be the native promoter or a substituted promoter which will provide expression in the plant. The promoter in the construct may be an inducible promoter, so that expression of the sense or antisense molecule in the construct can be controlled by exposure to the inducer. In this regard, any plant-compatible promoter elements can be employed in the construct, influenced by the end result desired. Those can be plant gene promoters, such as, for example, the promoter for the small subunit of ribulose-1,5-bis-phosphate carboxylase, or promoters from the tumor-inducing plasmids from *Agrobacterium tumefaciens*, such as the nopaline synthase and octopine synthase promoters, or viral promoters such as the cauliflower mosaic virus (CaMV) 19S and 35S promoters or the figwort mosaic virus 35S promoter. See Kay et al., (1987) *Science* 236:1299 and European patent application No. 0 342 926; the barley lipid transfer protein promoter, LTP2 (Kalla et al., *Plant J.* (1994) 6(6): 849-60); the ubiquitin promoter (see for example

US patent 5,510,474); the END2 promoter (Linnestad et al. US Patent 6,903,205); and the polygalacturonase PG47 promoter (See Allen and Lonsdale, *Plant J.* (1993) 3:261-271; WO 94/01572; US Patent 5,412,085). See international application WO 91/19806 for a review of illustrative plant promoters suitably employed in the present invention.

The range of available plant compatible promoters includes tissue specific and inducible promoters. An inducible regulatory element is one that is capable of directly or indirectly activating transcription of one or more DNA sequences or genes in response to an inducer. In the absence of an inducer the DNA sequences or genes will not be transcribed. Typically the protein factor that binds specifically to an inducible regulatory element to activate transcription is present in an inactive form which is then directly or indirectly converted to the active form by the inducer. The inducer can be a chemical agent such as a protein, metabolite, growth regulator, herbicide or phenolic compound or a physiological stress imposed directly by heat, cold, salt, or toxic elements or indirectly through the action of a pathogen or disease agent such as a virus. A plant cell containing an inducible regulatory element may be exposed to an inducer by externally applying the inducer to the cell or plant such as by spraying, watering, heating or similar methods.

Any inducible promoter can be used in the instant invention. See Ward et al. *Plant Mol. Biol.* 22: 361-366 (1993). Exemplary inducible promoters include ecdysone receptor promoters, U.S. Patent No. 6,504,082; promoters from the ACE1 system which responds to copper (Mett et al. *PNAS* 90: 4567-4571 (1993)); In2-1 and In2-2 gene from maize which respond to benzenesulfonamide herbicide safeners (U.S. Patent No. 5,364,780; Hershey et al., *Mol. Gen. Genetics* 227: 229-237 (1991) and Gatz et al., *Mol. Gen. Genetics* 243: 32-38 (1994)); the maize GST promoter, which is activated by hydrophobic electrophilic compounds that are used as pre-emergent herbicides; and the tobacco PR-1a promoter, which is activated by salicylic acid. Other chemical-regulated promoters of interest include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter in Schena et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425 and McNellis et al. (1998) *Plant J.* 14(2):247-257) and tetracycline-inducible and tetracycline-repressible promoters (see, for example, Gatz et al. (1991) *Mol. Gen. Genet.* 227:229-237, and U.S. Patent Nos. 5,814,618 and 5,789,156).

Tissue-preferred promoters can be utilized to target enhanced transcription and/or expression within a particular plant tissue. Promoters may express in the tissue of interest, along with expression in other plant tissue, may express strongly in the tissue of interest and to a much lesser degree than other tissue, or may

5 express highly preferably in the tissue of interest. Tissue-preferred promoters include those described in Yamamoto *et al.* (1997) *Plant J.* 12(2): 255-265; Kawamata *et al.* (1997) *Plant Cell Physiol.* 38(7): 792-803; Hansen *et al.* (1997) *Mol. Gen Genet.* 254(3):337-343; Russell *et al.* (1997) *Transgenic Res.* 6(2): 157-168; Rinehart *et al.* (1996) *Plant Physiol.* 112(3): 1331-1341; Van Camp *et al.*

10 (1996) *Plant Physiol.* 112(2): 525-535; Canevascini *et al.* (1996) *Plant Physiol.* 112(2): 513-524; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5): 773-778; Lam (1994) *Results Probl. Cell Differ.* 20: 181-196; Orozco *et al.* (1993) *Plant Mol Biol.* 23(6):1129-1138; Matsuoka *et al.* (1993) *Proc Natl. Acad. Sci. USA* 90(20): 9586-9590; and Guevara-Garcia *et al.* (1993) *Plant J.* 4(3): 495-505. In one

15 embodiment, the promoters are those which preferentially express to the male or female tissue of the plant. The invention does not require that any particular male tissue-preferred promoter be used in the process, and any of the many such promoters known to one skilled in the art may be employed. The native *Ms26* promoter described herein is one example of a useful promoter. Another such

20 promoter is the 5126 promoter, which preferentially directs expression of the gene to which it is linked to male tissue of the plants, as described in U.S. Patents Nos. 5,837,851 and 5,689,051. Other examples include the *Ms45* promoter described at US Patent No. 6,037,523; SF3 promoter described at U.S. Patent No. 6,452,069; the BS92-7 promoter described at WO 02/063021; a SGB6 regulatory

25 element described at U.S. Patent No. 5,470,359; the TA29 promoter (Koltunow *et al.* (1990) "Different temporal and spatial gene expression patterns occur during anther development." *Plant Cell* 2:1201-1224; Goldberg, R. B., Beals, T. P. and Sanders, P. M., (1993) "Anther development: basic principles and practical applications" *Plant Cell* 5:1217-1229; and US Patent No. 6,399,856); the type 2

30 metallothionein-like gene promoter (Charbonnel-Campaa *et al.*, *Gene* (2000) 254:199-208); and the *Brassica Bca9* promoter (Lee *et al.*, *Plant Cell Rep.* (2003) 22:268-273).

Male gamete preferred promoters include the PG47 promoter, *supra* as well as ZM13 promoter (Hamilton *et al.*, *Plant Mol. Biol.* (1998) 38:663-669); actin

depolymerizing factor promoters (such as Zmabp1, Zmabp2; see for example Lopez et al. *Proc. Natl. Acad. Sci. USA* (1996) 93: 7415-7420); the promoter of the maize petctin methylesterase-like gene, ZmC5 (Wakeley et al. *Plant Mol. Biol.* (1998) 37:187-192); the profiling gene promoter Zmpro1 (Kovar et al., *The Plant Cell* (2000) 12:583-598); the sulphated pentapeptide phytosulphokine gene ZmPSK1 (Lorbiecke et al., *Journal of Experimental Botany* (2005) 56(417): 1805-1819); the promoter of the calmodulin binding protein Mpcbp (Reddy et al. *J. Biol. Chem.* (2000) 275(45):35457-70).

Other components of the vector may be included, also depending upon intended use of the gene. Examples include selectable markers, targeting or regulatory sequences, stabilizing or leader sequences, introns etc. General descriptions and examples of plant expression vectors and reporter genes can be found in Gruber, et al., "Vectors for Plant Transformation" in Method in Plant Molecular Biology and Biotechnology, Glick et al eds;CRC Press pp. 89-119 (1993). The selection of an appropriate expression vector will depend upon the host and the method of introducing the expression vector into the host. The expression cassette will also include at the 3' terminus of the heterologous nucleotide sequence of interest, a transcriptional and translational termination region functional in plants. The termination region can be native with the promoter nucleotide sequence of the present invention, can be native with the DNA sequence of interest, or can be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also, Guerineau et al. *Mol. Gen. Genet.* 262:141-144 (1991); Proudfoot, *Cell* 64:671-674 (1991); Sanfacon et al. *Genes Dev.* 5:141-149 (1991); Mogen et al. *Plant Cell* 2:1261-1272 (1990); Munroe et al. *Gene* 91:151-158 (1990); Ballas et al. *Nucleic Acids Res.* 17:7891-7903 (1989); Joshi et al. *Nucleic Acid Res.* 15:9627-9639 (1987).

The expression cassettes can additionally contain 5' leader sequences. Such leader sequences can act to enhance translation. Translation leaders are known in the art and include by way of example, picornavirus leaders, EMCV leader (Encephalomyocarditis 5' noncoding region), Elroy-Stein et al. *Proc. Nat. Acad. Sci. USA* 86:6126-6130 (1989); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus), Allison et al.; MDMV leader (Maize Dwarf Mosaic Virus),

Virology 154:9-20 (1986); human immunoglobulin heavy-chain binding protein (BiP), Macejak et al. *Nature* 353:90-94 (1991); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), Jobling et al. *Nature* 325:622-625 (1987); Tobacco mosaic virus leader (TMV), Gallie et al. (1989) *Molecular Biology of RNA*, pages 237-256; and maize chlorotic mottle virus leader (MCMV) Lommel et al. *Virology* 81:382-385 (1991). See also Della-Cioppa et al. *Plant Physiology* 84:965-968 (1987). The cassette can also contain sequences that enhance translation and/or mRNA stability such as introns.

In those instances where it is desirable to have the expressed product of the heterologous nucleotide sequence directed to a particular organelle, particularly the plastid, amyloplast, or to the endoplasmic reticulum, or secreted at the cell's surface or extracellularly, the expression cassette can further comprise a coding sequence for a transit peptide. Such transit peptides are well known in the art and include, but are not limited to, the transit peptide for the acyl carrier protein, the small subunit of RUBISCO, plant EPSP synthase, *Zea mays* Brittle-1 chloroplast transit peptide (Nelson et al. *Plant physiol* 117(4):1235-1252 (1998); Sullivan et al. *Plant Cell* 3(12):1337-48; Sullivan et al., *Planta* (1995) 196(3):477-84; Sullivan et al., *J. Biol. Chem.* (1992) 267(26):18999-9004) and the like. One skilled in the art will readily appreciate the many options available in expressing a product to a particular organelle. For example, the barley alpha amylase sequence is often used to direct expression to the endoplasmic reticulum (Rogers, *J. Biol. Chem.* 260:3731-3738 (1985)). Use of transit peptides is well known (*e.g.*, see U.S. Patents Nos. 5,717,084; 5,728,925).

In preparing the expression cassette, the various DNA fragments can be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers can be employed to join the DNA fragments or other manipulations can be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction digests, annealing, and resubstitutions, such as transitions and transversions, can be involved.

As noted herein, the present invention provides vectors capable of expressing genes of interest. In general, the vectors should be functional in plant cells. At times, it may be preferable to have vectors that are functional in *E. coli*

(e.g., production of protein for raising antibodies, DNA sequence analysis, construction of inserts, obtaining quantities of nucleic acids). Vectors and procedures for cloning and expression in *E. coli* are discussed in Sambrook et al. (*supra*).

5 The transformation vector comprising the promoter sequence of the present invention operably linked to a heterologous nucleotide sequence in an expression cassette, can also contain at least one additional nucleotide sequence for a gene to be cotransformed into the organism. Alternatively, the additional sequence(s) can be provided on another transformation vector.

10 Reporter genes can be included in the transformation vectors. Examples of suitable reporter genes known in the art can be found in, for example, Jefferson et al. (1991) in *Plant Molecular Biology Manual*, ed. Gelvin et al. (Kluwer Academic Publishers), pp. 1-33; DeWet et al. *Mol. Cell. Biol.* 7:725-737 (1987); Goff et al. *EMBO J.* 9:2517-2522 (1990); Kain et al. *BioTechniques* 19:650-655
15 (1995); and Chiu et al. *Current Biology* 6:325-330 (1996).

 Selectable reporter genes for selection of transformed cells or tissues can be included in the transformation vectors. These can include genes that confer antibiotic resistance or resistance to herbicides. Examples of suitable selectable marker genes include, but are not limited to, genes encoding resistance to
20 chloramphenicol, Herrera Estrella et al. *EMBO J.* 2:987-992(1983); methotrexate, Herrera Estrella et al. *Nature* 303:209-213(1983); Meijer et al. *Plant Mol. Biol.* 16:807-820 (1991); hygromycin, Waldron et al. *Plant Mol. Biol.* 5:103-108 (1985), Zhijian et al. *Plant Science* 108:219-227 (1995); streptomycin, Jones et al. *Mol. Gen. Genet.* 210:86-91(1987); spectinomycin, Bretagne-Sagnard et al.
25 *Transgenic Res.* 5:131-137 (1996); bleomycin, Hille et al. *Plant Mol. Biol.* 7:171-176 (1990); sulfonamide, Guerineau et al. *Plant Mol. Biol.* 15:127-136(1990); bromoxynil, Stalker et al. *Science* 242:419-423 (1988); glyphosate, Shaw et al. *Science* 233:478-481(1986); and phosphinothricin, DeBlock et al. *EMBO J.* 6:2513-2518 (1987).

30 Scorable or screenable markers may also be employed, where presence of the sequence produces a measurable product. Examples include a β -glucuronidase, or uidA gene (GUS), which encodes an enzyme for which various chromogenic substrates are known (for example, US Patents 5,268,463 and 5,599,670); chloramphenicol acetyl transferase (Jefferson et al. *The EMBO*

Journal vol. 6 No. 13 pp. 3901-3907); and alkaline phosphatase. Other screenable markers include the anthocyanin/flavonoid genes in general (See discussion at Taylor and Briggs, *The Plant Cell* (1990)2:115-127) including, for example, a *R*-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta et al., in Chromosome Structure and Function, Kluwer Academic Publishers, Appels and Gustafson eds., pp. 263-282 (1988)); the genes which control biosynthesis of flavonoid pigments, such as the maize *C1* gene (Kao et al., *Plant Cell* (1996) 8: 1171-1179; Scheffler et al. *Mol. Gen. Genet.* (1994) 242:40-48) and maize *C2* (Wienand et al., *Mol. Gen. Genet.* (1986) 203:202-207); the *B* gene (Chandler et al., *Plant Cell* (1989) 1:1175-1183), the *p1* gene (Grotewold et al, *Proc. Natl. Acad. Sci USA* (1991) 88:4587-4591; Grotewold et al., *Cell* (1994) 76:543-553; Sidorenko et al., *Plant Mol. Biol.* (1999)39:11-19); the *bronze* locus genes (Ralston et al., *Genetics* (1988) 119:185-197; Nash et al., *Plant Cell* (1990) 2(11): 1039-1049), among others. Yet further examples of suitable markers include the cyan fluorescent protein (CYP) gene (Bolte et al. (2004) *J. Cell Science* 117: 943-54 and Kato et al. (2002) *Plant Physiol* 129: 913-42), the yellow fluorescent protein gene (PhiYFP™ from Evrogen; see Bolte et al. (2004) *J. Cell Science* 117: 943-54); a lux gene, which encodes a luciferase, the presence of which may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon counting cameras or multiwell luminometry (Teeri et al. (1989) *EMBO J.* 8:343); a green fluorescent protein (GFP) gene (Sheen et al., *Plant J.* (1995) 8(5):777-84); and DsRed2 where plant cells transformed with the marker gene are red in color, and thus visually selectable (Dietrich et al. (2002) *Biotechniques* 2(2):286-293). Additional examples include a p-lactamase gene (Sutcliffe, *Proc. Nat'l. Acad. Sci. U.S.A.* (1978) 75:3737), which encodes an enzyme for which various chromogenic substrates are known (e.g., PADAC, a chromogenic cephalosporin); a xyleE gene (Zukowsky et al., *Proc. Nat'l. Acad. Sci. U.S.A.* (1983) 80:1101), which encodes a catechol dioxygenase that can convert chromogenic catechols; an α -amylase gene (Ikuta et al., *Biotech.* (1990) 8:241); and a tyrosinase gene (Katz et al., *J. Gen. Microbiol.* (1983) 129:2703), which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone, which in turn condenses to form the easily

detectable compound melanin. Clearly, many such markers are available to one skilled in the art.

The method of transformation/transfection is not critical to the instant invention; various methods of transformation or transfection are currently
 5 available. As newer methods are available to transform crops or other host cells they may be directly applied. Accordingly, a wide variety of methods have been developed to insert a DNA sequence into the genome of a host cell to obtain the transcription or transcript and translation of the sequence to effect phenotypic changes in the organism. Thus, any method which provides for efficient
 10 transformation/transfection may be employed.

Methods for introducing expression vectors into plant tissue available to one skilled in the art are varied and will depend on the plant selected. Procedures for transforming a wide variety of plant species are well known and described throughout the literature. See, for example, Miki et al, "Procedures for
 15 Introducing Foreign DNA into Plants" in Methods in Plant Molecular Biotechnology, *supra*; Klein et al, *Bio/Technology* 10:268 (1992); and Weising et al., *Ann. Rev. Genet.* 22: 421-477 (1988). For example, the DNA construct may be introduced into the genomic DNA of the plant cell using techniques such as microprojectile-mediated delivery, Klein et al., *Nature* 327: 70-73 (1987);
 20 electroporation, Fromm et al., *Proc. Natl. Acad. Sci.* 82: 5824 (1985); polyethylene glycol (PEG) precipitation, Paszkowski et al., *EMBO J.* 3: 2717-2722 (1984); direct gene transfer WO 85/01856 and EP No. 0 275 069; *in vitro* protoplast transformation, U.S. Patent No. 4,684,611; and microinjection of plant cell protoplasts or embryogenic callus, Crossway, *Mol. Gen. Genetics* 202:179-
 25 185 (1985). Co-cultivation of plant tissue with *Agrobacterium tumefaciens* is another option, where the DNA constructs are placed into a binary vector system. See e.g., U.S. Patent No. 5,591,616; Ishida et al., "High Efficiency Transformation of Maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*" *Nature Biotechnology* 14:745-750 (1996). The virulence functions of the *Agrobacterium*
 30 *tumefaciens* host will direct the insertion of the construct into the plant cell DNA when the cell is infected by the bacteria. See, for example Horsch et al., *Science* 233: 496-498 (1984), and Fraley et al., *Proc. Natl. Acad. Sci.* 80: 4803 (1983).

Standard methods for transformation of canola are described at Moloney et al. "High Efficiency Transformation of *Brassica napus* using *Agrobacterium*

Vectors" *Plant Cell Reports* 8:238-242 (1989). Corn transformation is described by Fromm et al, *Bio/Technology* 8:833 (1990) and Gordon-Kamm et al, *supra*. *Agrobacterium* is primarily used in dicots, but certain monocots such as maize can be transformed by *Agrobacterium*. See *supra* and U.S. Patent No. 5,550,318.

5 Rice transformation is described by Hiei et al., "Efficient Transformation of Rice (*Oryza sativa* L.) Mediated by *Agrobacterium* and Sequence Analysis of the Boundaries of the T-DNA" *The Plant Journal* 6(2): 271-282 (1994, Christou et al, *Trends in Biotechnology* 10:239 (1992) and Lee et al, *Proc. Nat'l Acad. Sci. USA* 88:6389 (1991). Wheat can be transformed by techniques similar to those used
10 for transforming corn or rice. Sorghum transformation is described at Casas et al, *supra* and sorghum by Wan et al, *Plant Physiol.* 104:37 (1994). Soybean transformation is described in a number of publications, including U.S. Patent No. 5,015,580.

When referring to "introduction" of the nucleotide sequence into a plant, it is
15 meant that this can occur by direct transformation methods, such as *Agrobacterium* transformation of plant tissue, microprojectile bombardment, electroporation, or any one of many methods known to one skilled in the art; or, it can occur by crossing a plant having the heterologous nucleotide sequence with another plant so that progeny have the nucleotide sequence incorporated into
20 their genomes. Such breeding techniques are well known to one skilled in the art.

The plant breeding methods used herein are well known to one skilled in the art. For a discussion of plant breeding techniques, see Poehlman (1987) Breeding Field Crops. AVI Publication Co., Westport Conn. Many of the plants which would be most preferred in this method are bred through techniques that
25 take advantage of the plant's method of pollination.

Backcrossing methods may be used to introduce a gene into the plants. This technique has been used for decades to introduce traits into a plant. An example of a description of this and other plant breeding methodologies that are well known can be found in references such as Plant Breeding Methodology, edit.
30 Neal Jensen, John Wiley & Sons, Inc. (1988). In a typical backcross protocol, the original variety of interest (recurrent parent) is crossed to a second variety (nonrecurrent parent) that carries the single gene of interest to be transferred. The resulting progeny from this cross are then crossed again to the recurrent parent and the process is repeated until a plant is obtained wherein essentially all of the

desired morphological and physiological characteristics of the recurrent parent are recovered in the converted plant, in addition to the single transferred gene from the nonrecurrent parent.

5 In certain embodiments of the invention, it is desirable to maintain the male sterile homozygous recessive condition of a male sterile plant, when using a transgenic restoration approach, while decreasing the number of plants, plantings and steps needed for maintenance plant with such traits. Homozygosity is a genetic condition existing when identical alleles reside at corresponding loci on homologous chromosomes. Heterozygosity is a genetic condition existing when
10 different alleles reside at corresponding loci on homologous chromosomes. Hemizygosity is a genetic condition existing when there is only one copy of a gene (or set of genes) with no allelic counterpart on the sister chromosome. In an embodiment, the homozygous recessive condition results in conferring on the plant a trait of interest, which can be any trait desired and which results from the
15 recessive genotype, such as increased drought or cold tolerance, early maturity, changed oil or protein content, or any of a multitude of the many traits of interest to plant breeders. In one embodiment, the homozygous recessive condition confers male sterility upon the plant. When the sequence which is the functional complement of the homozygous condition is introduced into the plant (that is, a
20 sequence which, when introduced into and expressed in the plant having the homozygous recessive condition, restores the wild-type condition), fertility is restored by virtue of restoration of the wild-type fertile phenotype.

Maintenance of the homozygous recessive condition is achieved by introducing a restoration transgene construct into a plant that is linked to a
25 sequence which interferes with the function or formation of male gametes of the plant to create a maintainer or donor plant. The restoring transgene, upon introduction into a plant that is homozygous recessive for the genetic trait, restores the genetic function of that trait, with the plant producing only viable pollen containing a copy of the recessive allele but does not contain the
30 restoration transgene. The transgene is kept in the hemizygous state in the maintainer plant. By transgene, it is meant any nucleic acid sequence which is introduced into the genome of a cell by genetic engineering techniques. A transgene may be a native DNA sequence, or a heterologous DNA sequence (i.e., "foreign DNA"). The term native DNA sequence refers to a nucleotide sequence

which is naturally found in the cell but that may have been modified from its original form. The pollen from the maintainer can be used to fertilize plants that are homozygous for the recessive trait, and the progeny will therefore retain their homozygous recessive condition. The maintainer plant containing the restoring transgene construct is propagated by self-fertilization, with the resulting seed used to produce further plants that are homozygous recessive plants and contain the restoring transgene construct.

The maintainer plant serves as a pollen donor to the plant having the homozygous recessive trait. The maintainer is optimally produced from a plant having the homozygous recessive trait and which also has nucleotide sequences introduced therein which would restore the trait created by the homozygous recessive alleles. Further, the restoration sequence is linked to nucleotide sequences which interfere with the function or formation of male gametes. The gene can operate to prevent formation of male gametes or prevent function of the male gametes by any of a variety of well-known modalities and is not limited to a particular methodology. By way of example but not limitation, this can include use of genes which express a product cytotoxic to male gametes (See for example, 5,792,853; 5,689,049; PCT/EP89/00495); inhibit product formation of another gene important to male gamete function or formation (See, U.S. Patent Nos. 5,859,341; 6,297,426); combine with another gene product to produce a substance preventing gene formation or function (See U.S. Patent Nos. 6,162,964; 6,013,859; 6,281,348; 6,399,856; 6,248,935; 6,750,868; 5,792,853); are antisense to or cause co-suppression of a gene critical to male gamete function or formation (See U.S. Patent Nos. 6,184,439; 5,728,926; 6,191,343; 5,728,558; 5,741,684); interfere with expression through use of hairpin formations (Smith et al. (2000) *Nature* 407:319-320; WO 99/53050 and WO 98/53083) or the like. Many nucleotide sequences are known which inhibit pollen formation or function and any sequences which accomplish this function will suffice. A discussion of genes which can impact proper development or function is included at U.S. Patent No. 6,399,856 and includes dominant negative genes such as cytotoxin genes, methylase genes, and growth-inhibiting genes. Dominant negative genes include diphtheria toxin A-chain gene (Czako, M. and An, G. (1991) "Expression of DNA coding for Diphtheria toxin Chain A is toxic to plant cells" *Plant Physiol.* 95 687-692. and Greenfield et al *PNAS* 80:6853 (1983),

Palmiter et al *Cell* 50:435 (1987)); cell cycle division mutants such as CDC in maize (Colasanti, J., Tyers, M. and Sundaresan, V., "Isolation and Characterization of cDNA clones encoding a functional P34 cdc2 homologue from *Zea mays*" *PNAS* 88, 3377-3381 (1991)); the WT gene (Farmer, A. A., Loftus, T. M., Mills, A. A., Sato, K. V., Neill, J., Yang, M., Tron, T., Trumpower, B. L. and Stanbridge, E. G. *Hum. Mol. Genet.* 3, 723-728 (1994)); and P68 (Chen, J. J., Pal, J. K., Petryshyn, R., Kuo, I., Yang, J. M., Throop, M. S., Gehrke, L. and London, I. M. "Eukaryotic translation initiation kinases" *PNAS* 88, 315-319 (1991)).

Further examples of so-called "cytotoxic" genes are discussed *supra* and can include, but are not limited to pectate lyase gene pelE, from *Erwinia chrysanthemi* (Kenn et al *J. Bacteriol* 168:595 (1986)); T-urf13 gene from cms-T maize mitochondrial genomes (Braun et al *Plant Cell* 2:153 (1990); Dewey et al. *PNAS* 84:5374 (1987)); CytA toxin gene from *Bacillus thuringiensis Israeliensis* that causes cell membrane disruption (McLean et al *J. Bacteriol* 169:1017 (1987), U.S. Patent No. 4,918,006); DNases, RNases, (U.S. Patent No. 5,633,441); proteases, or a genes expressing anti-sense RNA. A suitable gene may also encode a protein involved in inhibiting pistil development, pollen stigma interactions, pollen tube growth or fertilization, or a combination thereof. In addition genes that either interfere with the normal accumulation of starch in pollen or affect osmotic balance within pollen may also be suitable.

In an illustrative embodiment, the DAM-methylase gene is used, discussed *supra* and at US Patent Nos. 5,792,852 and 5,689,049, the expression product of which catalyzes methylation of adenine residues in the DNA of the plant. Methylated adenines will affect cell viability and will be found only in the tissues in which the DAM-methylase gene is expressed. In another embodiment, an α -amylase gene can be used with a male tissue-preferred promoter. During the initial germinating period of cereal seeds, the aleurone layer cells will synthesize α -amylase, which participates in hydrolyzing starch to form glucose and maltose, so as to provide the nutrients needed for the growth of the germ (J. C. Rogers and C. Milliman, *J. Biol. Chem.*, 259 (19): 12234-12240, 1984; Rogers, J. C., *J. Biol. Chem.*, 260: 3731-3738, 1985). In an embodiment, the α -amylase gene used can be the *Zea mays* α -amylase-1 gene. Young et al. "Cloning of an α -amylase cDNA from aleurone tissue of germinating maize seed" *Plant Physiol.* 105(2) 759-760 and GenBank accession No. L25805, GI:426481). Sequences encoding α -

amylase are not typically found in pollen cells, and when expression is directed to male tissue, the result is a breakdown of the energy source for the pollen grains, and repression of pollen development.

One skilled in this area readily appreciates the methods described herein are applicable to any other crops which have the potential to outcross. By way of example, but not limitation it can include maize, soybean, sorghum, or any plant with the capacity to outcross.

Ordinarily, to produce more plants having the recessive condition, one might cross the recessive plant with another recessive plant. This may not be desirable for some recessive traits and may be impossible for recessive traits affecting reproductive development. Alternatively, one could cross the homozygous plant with a second plant having the restoration gene, but this requires further crossing to segregate away the restoring gene to once again reach the recessive phenotypic state. Instead, in one process the homozygous recessive condition can be maintained, while crossing it with the maintainer plant. This method can be used with any situation in which is it desired to continue the recessive condition. This results in a cost-effective system that is relatively easy to operate to maintain a population of homozygous recessive plants.

A sporophytic gene is one which operates independently of the gametes. When the homozygous recessive condition is one which produces male sterility by preventing male sporophyte development, the maintainer plant, of necessity, must contain a functional restoring transgene construct capable of complementing the mutation and rendering the homozygous recessive plant able to produce viable pollen. Linking this sporophytic restoration gene with a second functional nucleotide sequence which interferes with the function or formation of the male gametes of the plant results in a maintainer plant that produces pollen containing only the recessive allele of the sporophytic gene at the its native locus due to the action of the second nucleotide sequence in interfering with pollen formation or function. This viable pollen fraction is non-transgenic with regard to the restoring transgene construct.

In a still further embodiment, a marker gene, as discussed *supra*, may be provided in the construct with the restoring transgene. By way of example without limitation, use of a herbicide resistant marker, such as *bar* allows one to eliminate cells not having the restoring transgene. In yet another example, when using a

scorable marker, such as a red fluorescent marker, such as DsRed2, any inadvertent transmission of the transgene can also be detected visually, and such escapes eliminated from progeny. Clearly, many other variations in the restoring construct are available to one skilled in the art.

5 In an illustrative embodiment, a method of maintaining a homozygous recessive condition of a male sterile plant at a genetic locus is provided, in which is employed a first nucleotide sequence which is a gene critical to male fertility, a second nucleotide sequence which inhibits the function or formation of viable male gametes, an optional third nucleotide sequence which is operably linked to
10 the first sequence and preferentially expresses the sequence in male plant cells, an optional fourth nucleotide sequence operably linked to a fourth nucleotide sequence, the fourth sequence directing expression to male gametes, and an optional fifth nucleotide sequence which is a selectable or scorable marker allowing for selection of plant cells.

15 For example, it is desirable to produce male sterile female plants for use in the hybrid production process which are sterile as a result of being homozygous for a mutation in the *Ms45* gene; a gene, which is critical to male fertility. Such a mutant *Ms45* allele is designated as *ms45* and a plant that is homozygous for *ms45* (represented by the notation *ms45/ms45*) displays the homozygous
20 recessive male sterility phenotype and produces no functional pollen. See, U.S. Patents Nos. 5,478,369; 5,850,014; 6,265,640; and 5,824,524. In both the inbred and hybrid production processes, it is highly desired to maintain this homozygous recessive condition. When sequences encoding the *Ms45* gene are introduced into a plant having the homozygous condition, male fertility results. By the method
25 of the invention, a plant which is *ms45/ms45* homozygous recessive may have introduced into it a functional sporophytic *Ms45* gene, and thus is male fertile. This gene can be linked to a gene which operates to render pollen containing the restoring transgene construct nonfunctional or prevents its formation, or which produces a lethal product in pollen, linked to the promoter directing its expression
30 to the male gametes to produce a plant that only produced pollen containing *ms45* without the restoring transgene construct.

An example is a construct which includes the *Ms45* gene, linked with a 5126 promoter, a male tissue-preferred promoter (See U.S. Patent No. 5,750,868; 5,837,851; and 5,689,051) and further linked to the cytotoxic DAM

methylase gene under control of the polygalacturonase promoter, PG47 promoter (See U.S. patent No. 5,792,853; 5,689,049) in a hemizygotic condition. Therefore the resulting plant produces pollen, but the only viable pollen results from the allele not containing the restoring *Ms45/DAM* methylase construct and thus
5 contains only the *ms45* gene. It can therefore be used as a pollinator to fertilize the homozygous recessive plant (*ms45/ms45*), and progeny produced will continue to be male sterile as a result of maintaining homozygosity for *ms45*. The progeny will also not contain the introduced restoring transgene construct.

In yet another restoring construct example, the *Ms26* gene is linked with a
10 5126 promoter, and further linked to the *Zea mays* α -amylase gene under control of the male tissue-preferred PG47 promoter. The scorable marker used in an embodiment is DS-RED EXPRESS.

A desirable result of the process of the invention is that the plant having the restorer nucleotide sequence may be self-fertilized, that is pollen from the plant
15 transferred to the flower of the same plant to achieve the propagation of restorer plants. (Note that in referring to "self fertilization", it includes the situation where the plant producing the pollen is fertilized with that same the pollen, and the situation where two or more identical inbred plants are planted together and pollen from the identical inbred plant pollinate a different identical inbred plant). The
20 pollen will not have the restoring transgene construct but it will be contained in 50% of the ovules (the female gamete). The seed resulting from the self-fertilization can be planted, and selection made for the seed having the restoring transgene construct. The selection process can occur by any one of many known processes; the most common where the restoration nucleotide sequence is linked
25 to a marker gene. The marker can be scorable or selectable, and allows those plants produced from the seed having the restoration gene to be identified.

In an embodiment of the invention, it is possible to provide that the male gamete-tissue preferred promoter is inducible. Additional control is thus allowed in the process, where so desired, by providing that the plant having the restoration
30 nucleotide sequences is constitutively male sterile. This type of male sterility is set forth in U.S. Patent No. 5,859,341. In order for the plant to become fertile, the inducing substance must be provided, and the plant will become fertile. Again, when combined with the process of the invention as described *supra*, the only pollen produced will not contain the restoration nucleotide sequences.

Further detailed description is provided below by way of instruction and illustration and is not intended to limit the scope of the invention.

EXAMPLE 1

Identification and Cosegregation of *ms26-m2::Mu8*

5 Families of plants from a Mutator (*Mu*) population were identified that segregated for plants that were mostly male sterile, with none or only a few extruded abnormal anthers, none of which had pollen present. Male sterility is expected to result from those instances where a *Mu* element has randomly integrated into a gene responsible for some step in microsporogenesis, disrupting
10 its expression. Plants from a segregating F₂ family in which the male sterile mutation was designated *ms26*-SBMu200*, were grown and classified for male fertility/sterility based on the above criteria. Leaf samples were taken and DNA subsequently isolated on approximately 20 plants per phenotypic classification, that is male fertility vs. male sterility.

15 Southern analysis was performed to confirm association of *Mu* with sterility. Southern analysis is a well known technique to those skilled in the art. This common procedure involves isolating the plant DNA, cutting with restriction endonucleases, fractioning the cut DNA by molecular weight on an agarose gel, and transferring to nylon membranes to fix the separated DNA. These
20 membranes are subsequently hybridized with a probe fragment that was radioactively labeled with P³²P-dCTP, and washed in an SDS solution. Southern, E., "Detection of Specific Sequences Among DNA Fragments by Gel Electrophoresis," J. Mol. Biol. 98:503-317 (1975). Plants from a segregating F₂ *ms26*-SBMu200* family were grown and classified for male fertility/sterility. Leaf
25 samples and subsequent DNA isolation was conducted on approximately 20 plants per phenotypic classification. DNA (~7ug) from 5 fertile and 12 sterile plants was digested with *EcoRI* and electrophoresed through a 0.75% agarose gel. The digested DNA was transferred to nylon membrane via Southern transfer. The membrane was hybridized with an internal fragment from the *Mu8*
30 transposon. Autoradiography of the membrane revealed cosegregation of an approximately 5.6 Kb *EcoRI* fragment with the sterility phenotype as shown in Figure 1. This *EcoRI* band segregated in the fertile plants suggesting a heterozygous wild type condition for the allele

EXAMPLE 2

Library Construction, Screening, and Mapping

The process of genomic library screenings is commonly known among those skilled in the art and is described at Sambrook, J., Fritsch, E.F., Maniatis T., et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Lab Press, Plainview, NY (1989). Libraries were created as follows.

DNA from a sterile plant was digested with *Eco*RI and run on a preparative gel. DNA with a molecular weight between 5.0 and 6.0 Kb was excised from the gel, electroeluted and ethanol precipitated. This DNA was ligated into the Lambda Zap vector (Stratagene™) using the manufacturer's protocol. The ligated DNA was packaged into phage particles using Gigapack Gold (Stratagene™). Approximately 500,000 PFU were plated and lifted onto nitrocellulose membranes. Membranes were hybridized with the *Mu8* probe. A pure clone was obtained after 3 rounds of screening. The insert was excised from the phage as a plasmid and designated SBMu200-3.1. A *Pst*I border fragment from this clone was isolated and used to reprobe the original *Eco*RI cosegregation blot as shown in Figure 2B. The approximately 5.6 kb *Eco*RI fragment is homozygous in all the sterile plants, which confirms that the correct *Mu* fragment was isolated. Three of the fertile plants are heterozygous for the 5.5 kb *Eco*RI band and a 4.3 Kb *Eco*RI band. Two of the fertile plants are homozygous for the 4.3 kb *Eco*RI band, presumably the wild type allele.

The *Pst*I probe was used to map the *ms**-SBMu200 mutation in an RFLP mapping population. The mutant mapped to the short arm of chromosome 1, near the male sterile locus, *Ms26* (Loukides *et al.*, (1995) *Amer. J. Bot* 82, 1017-1023). To test whether *ms**-SBMu200 was an allele of *ms26-ref*, *ms**-SBMu200 and *ms26-ref* were crossed with each other using a known heterozygote as the pollen donor. The testcross progeny segregated male-sterile and wild-type plants in a 1:1 ratio, indicating allelism between *ms**-SBMu200 and *ms26-ref*. The *ms**-SBMu200 allele was designated *ms26-m2::Mu8*. The map location is shown in Figure 13.

EXAMPLE 3Identification and Cloning of Additional *ms26* alleles

An additional *Mu* insertion mutations in *Ms26* was identified by using a polymerase chain reaction (PCR) primer for *Mu* and a gene specific primer for *Ms26* and screening a population of *Mu* F₁ families. Sequence analyses of the PCR products showed that all three *Mu* insertions occurred in the second exon (Figure 1). The F₂ seeds from one of these families were grown and examined for male fertility/sterility. Southern blot analyses of this family confirmed the cosegregation of the *Mu* insertion in *Ms26* with the male-sterile phenotype and the allele was designated *ms26-m3::Mu*.

The *ms26* allele described in Loukides *et al.*, (1995) *Amer. J. Bot* 82, 1017-1023 and designated *ms26-ref* was also investigated. To analyze the mutation in *ms26-ref*, *Ms26* genomic sequences were cloned from *ms26-ref* sterile and fertile plants. *Ms26* was cloned as a ~4.2 kb *EcoRI* fragment and *ms26-ref* cloned as a ~6 kb *HindIII* fragment and an overlapping ~2.3 kb *EcoRI* fragment from the sterile plant. Sequence analysis revealed the presence of a new segment (1,430 bp) in the last exon of the *ms26-ref* allele shown in Figure 1. An 8 bp host site duplication (GCCGGAGC) was found that flanks the inserted element and the element also contains a 15 bp terminal inverted repeat (TIR) (TAGGGGTGAAAACGG; SEQ ID NO: 23). The transposon sequence is shown in Figure 15 (SEQ ID NO: 10). The *ms26-ref* genomic sequence in its entirety is shown in Figure 16, SEQ ID NO: 11. A variant of the *ms26-ref* allele was also found. Sequence analysis of this allele, designated *ms26'-0406*, was found to have lost the 1430 bp segment found in the last exon of the *ms26-ref* allele but left an 8bp footprint at the site of insertion. Plants homozygous for the *ms26'-0406* allele were male sterile. A comparison of the excision allele, *ms26'-0406* (SEQ ID NO: 8) with the region in the wild-type *Ms26* gene (SEQ ID NO: 9) is shown in Figure 14.

EXAMPLE 4Expression Analysis and cDNA Isolation

Northern analysis can be used to detect expression of genes characteristic of anther development at various states of microsporogenesis. Northern analysis is also a commonly used technique known to those skilled in the art and is similar to Southern analysis except that mRNA rather than DNA is isolated and placed on

the gel. The RNA is then hybridized with the labeled probe. Potter, E., et al., "Thyrotropin Releasing Hormone Exerts Rapid Nuclear Effects to Increase Production of the Primary Prolactin in RNA Transcript," *Proc. Nat. Acad. Sci. USA* 78:6662-6666 (1981), Lechelt, et al., "Isolation & Molecular Analysis of the Plows," *Mol.Gen.Genet.* 219:225-234 (1989). The *Pst*I fragment from the SBMu200-3.1 clone was used to probe a Northern blot containing kernel, immature ear, seedling and tassel RNA. A signal was seen only in tassel RNA at approximately the quartet stage of microsporogenesis, as reflected in Figure 3. The transcript is about 2.3 kb in length. The same probe was also used to screen a cDNA library constructed from mRNA isolated from meiotic to late uninucleate staged anthers. One clone, designated *Ms26-8.1*, was isolated from the library.

EXAMPLE 5

Sequence and Expression Analysis

The SBMu200-3.1 genomic clone and the *Ms26-8.1* cDNA clone were sequenced by Loftstrand Labs Limited. Sanger, F., Nicklen, S., Coulson A.R. (1977) "DNA sequencing with chain terminating inhibitors" *Proc. Natl. Acad. Sci. USA* 74:5463-5467. The sequences are set forth in Figure 4 and 5 and the comparison is at Figure 6. The cDNA/genomic comparison reveals five introns are present in the genomic clone. The *Mu8* insertion occurs in exon 1. Testing for codon preference and non-randomness in the third position of each codon was consistent with the major ORF in the cDNA being the likely protein-coding ORF. There is a putative Met start codon at position 1089 in the genomic clone. The cDNA homology with respect to the genomic clone begins at nucleotide 1094. Thus *Ms26-8.1* does not represent a full length clone and lacks 5 bases up to the putative Met start codon. A database search revealed significant homology to P450 enzymes found in yeast, plants and mammals. P450 enzymes have been widely studied and three characteristic protein domains have been elucidated. The *Ms26* protein contains several structural motifs characteristic of eukaryotic P450's, including the heme-binding domain FxxGxRxCxG (domain D; SEQ ID NO: 24), domain A A/GGXD/ETT/S (dioxygen-binding), domain B (steroid-binding), and domain C. The highly conserved heme-binding motif was found in *MS26* as FQAGPRICLG (SEQ ID NO: 25), 51 amino acids away from C-terminus. The dioxygen binding domain AGRDTT (SEQ ID NO: 35) was located between amino acids 320-325. The steroid-binding domain was found as LVY LHACVTETLR

(SEQ ID NO: 27), amino acids 397-409. The most significant homologous sequence detected in Genbank database is a deduced protein sequence from rice (GeneBank accession number 19071651). The second highest homologous sequence is a putative *Arabidopsis* P450 gene (CYP704B1) whose function is also unknown. Figure 17A shows a sequence alignment between CYP704B1 (SEQ ID NO: 12) and *Ms26* (SEQ ID NO: 13). Phylogenetic tree analysis of some P450 genes revealed that *Ms26* is most closely related to P450s involved in fatty acid omega-hydroxylation found in *Arabidopsis thaliana* and *Vicia sativa* (Figure 17B). The translational frame shift caused in the *ms26'-0406* excision mutation is believed to destroy the activity of the heme binding domain, thus resulting in sterility. See the comparison at Figure 18 (*Ms26* cDNA at SEQ ID NO: 14; fertile exon 5 region at SEQ ID NO: 15 and sterile exon 5 region is SEQ ID NO: 16).

Further expression studies were done using the *Ms26* cDNA probe against a northern containing mRNA at discrete stages of microsporogenesis. Figure 7A shows a Northern blot with RNA samples from different tissues including root (1), leaf (2), husk (3), cob (4), ear spikelet (5), silk (6), immature embryo (7) mature embryo (8), and tassel from, fertile plant (9), *ms26-m2::Mu8* sterile plant (10), *ms26-ref* sterile plant (11) and fertile plant (12). A hybridization signal using *Ms26* cDNA was detected only in tassel tissues. Figure 7B shows a Northern blot containing mRNA at discrete stages of microsporogenesis. Hybridization signals using *Ms26* cDNA were detected from meiosis II/ quartet stage (4) to late-uninucleate stage (10), with the maximal signal being observed from early-uninucleate through late-uninucleate stage (10).

25 EXAMPLE 6

Identification of Promoter and its Essential Regions

A putative TATA box can be identified by primer extension analysis as described in by Current Protocols in Molecular Biology, Ausubel, F.M. et al. eds; John Wiley and Sons, New York pp.4.8.1 - 4.8.5 (1987).

30 Regulatory regions of anther genes, such as promoters, may be identified in genomic subclones using functional analysis, usually verified by the observation of reporter gene expression in anther tissue and a lower level or absence of reporter gene expression in non-anther tissue. The possibility of the regulatory regions residing "upstream" or 5' ward of the translational start site can be tested

by subcloning a DNA fragment that contains the upstream region into expression vectors for transient expression experiments. It is expected that smaller subgenomic fragments may contain the regions essential for male-tissue preferred expression. For example, the essential regions of the CaMV 19S and
 5 35S promoters have been identified in relatively small fragments derived from larger genomic pieces as described in U.S. Pat. No. 5,352,605.

The selection of an appropriate expression vector with which to test for functional expression will depend upon the host and the method of introducing the expression vector into the host and such methods are well known to one skilled in
 10 the art. For eukaryotes, the regions in the vector include regions that control initiation of transcription and control processing. These regions are operably linked to a reporter gene such as UidA, encoding -glucuronidase (GUS), or luciferase. General descriptions and examples of plant expression vectors and reporter genes can be found in Gruber, et al., "Vectors for Plant Transformation"
 15 in Methods in Plant Molecular Biology and Biotechnology; Glick, et al. eds; CRC Press; pp. 89-119; (1993). GUS expression vectors and GUS gene cassettes are commercially available from Clontech, Palo Alto, CA, while luciferase expression vectors and luciferase gene cassettes are available from Promega Corporation, Madison, WI. Ti plasmids and other *Agrobacterium* vectors are described in
 20 Ishida, Y., et al., Nature Biotechnology; Vol. 14; pp. 745-750; (1996) and in U.S. Pat. No. 5,591,616 "Method for Transforming Monocotyledons" (1994).

Expression vectors containing putative regulatory regions located in genomic fragments can be introduced into intact tissues such as staged anthers, embryos or into callus. Methods of DNA delivery include microprojectile
 25 bombardment, DNA injection, electroporation and *Agrobacterium*-mediated gene transfer (see Gruber, et al., "Vectors for Plant Transformation," in Methods in Plant Molecular Biology and Biotechnology, Glick, et al. eds.; CRC Press; (1993); U.S. Pat. No. 5,591,616; and Ishida, Y., et al., Nature Biotechnology; Vol. 14; pp. 745-750; (1996)). General methods of culturing plant tissues are found in Gruber, et
 30 al., *supra* and Glick, *supra*.

For the transient assay system, staged, isolated anthers are immediately placed onto tassel culture medium (Pareddy, D.R. and J.F. Petelino, *Crop Sci. J.*; Vol. 29; pp. 1564-1566; (1989)) solidified with 0.5% Phytigel (Sigma, St. Louis) or other solidifying media. The expression vector DNA is introduced within 5 hours

preferably by microprojectile-mediated delivery with 1.2 μm particles at 1000 - 1100 Psi. After DNA delivery, the anthers are incubated at 26°C upon the same tassel culture medium for 17 hours and analyzed by preparing a whole tissue homogenate and assaying for GUS or for luciferase activity (see Gruber, et al.,
 5 *supra*).

Upstream of the likely translational start codon of *Ms26*, 1088 bp of DNA was present in the genomic clone *ms26-m2::Mu8*. Translational fusions via an engineered *NcoI* site were generated with reporter genes encoding luciferase and β -glucuronidase to test whether this fragment of DNA had promoter activity in
 10 transient expression assays of bombarded plant tissues. Activity was demonstrated in anthers and not in coleoptiles, roots and calli, suggesting anther-preferred or anther-specific promoter activity.

A reasonable TATA box was observed by inspection, about 83-77 bp upstream of the translational start codon. The genomic clone *ms26-m2::Mu8* thus
 15 includes about 1005 bp upstream of the possible TATA box. For typical plant genes, the start of transcription is 26-36 bp downstream of the TATA box, which would give the *Ms26* mRNA a 5'-nontranslated leader of about 48-58 nt. The total *ms26-m2::Mu8* subgenomic fragment of 1088 bp, including nontranslated leader, start of transcription, TATA box and sequences upstream of the TATA box, was
 20 thus shown to be sufficient for promoter activity. See Figure 8, which is SEQ. ID NO.5. The putative TATA box (TATATCA) is underlined. Thus, the present invention encompasses a DNA molecule having a nucleotide sequence of SEQ ID NO: 5 (or those with sequence identity) and having the function of a male tissue-preferred regulatory region.

25 Deletion analysis can occur from both the 5' and 3' ends of the regulatory region: fragments can be obtained by site-directed mutagenesis, mutagenesis using the polymerase chain reaction, and the like (Directed Mutagenesis: A Practical Approach; IRL Press; (1991)). The 3' end of the male tissue-preferred regulatory region can be delineated by proximity to the putative TATA box or by 3'
 30 deletions if necessary. The essential region may then be operably linked to a core promoter of choice. Once the essential region is identified, transcription of an exogenous gene may be controlled by the male tissue-preferred region of *Ms26* plus a core promoter. The core promoter can be any one of known core promoters such as a Cauliflower Mosaic Virus 35S or 19S promoter (U.S. Pat. No. 5,352,605),

Ubiquitin (U.S. Pat. No. 5,510,474), the IN2 core promoter (U.S. Pat. No. 5,364,780), or a Figwort Mosaic Virus promoter (Gruber, et al., "Vectors for Plant Transformation" in Methods in Plant Molecular Biology and Biotechnology; Glick, et al. eds.; CRC Press; pp. 89-119; (1993)). Preferably, the promoter is the core promoter of a male tissue-preferred gene or the CaMV 35S core promoter. More preferably, the promoter is a promoter of a male tissue-preferred gene and in particular, the *Ms26* core promoter.

Further mutational analysis, for example by linker scanning, a method well known to the art, can identify small segments containing sequences required for anther-preferred expression. These mutations may introduce modifications of functionality such as in the levels of expression, in the timing of expression, or in the tissue of expression. Mutations may also be silent and have no observable effect.

The foregoing procedures were used to identify essential regions of the *Ms26* promoter. After linking the promoter with the luciferase marker gene deletion analysis was performed on the regions of the promoter upstream of the putative TATA box, as represented in Figure 9. The x-axis of the bar graph indicates the number of base pairs immediately upstream of the putative TATA box retained in a series of deletion derivatives starting from the 5' end of the promoter. The y-axis shows the normalized luciferase activity as a percent of full-length promoter activity.

As is evident from the graph, approximately 176 bp immediately upstream of the TATA box was sufficient, when coupled to the core promoter (putative TATA box through start of transcription), plus 5' nontranslated leader, for transient expression in anthers. By contrast, luciferase activity was minimal upon further deletion from the 5' end to 91 bp upstream of the putative TATA box. This 176 bp upstream of the putative TATA box through the nontranslated leader can be considered a minimal promoter, which is further represented at Figure 10. The TATA box is underlined. Deletion within the full-length promoter from -176 through -92 relative to the TATA box reduced activity to about 1% of wild type. Deletion of -39 through -8 did not greatly reduce activity. Therefore the -176 to -44bp region contains an essential region and thus would constitute an upstream enhancer element conferring anther expression on the promoter, which we refer to as an "anther box".

Linker scanning analysis was conducted across the anther box in 9-10 bp increments. The locations of the linker scanning substitutions in this region are shown in Figure 10, and the expression levels of the mutants relative to the wild type sequence are shown in Figure 11. The most drastic effect on transient expression in anthers was observed for mutants LS12 and LS13, in the region 52-71 bp upstream of the putative TATA box. A major effect on transient expression in anthers was also observed for mutants LS06, LS07, LS08 and LS10, within the region 82-131 bp upstream of the putative TATA box. Sequences within the anther box required for wild type levels of transient expression in anthers are thus demonstrated in the -52 to -131 region relative to the putative TATA box, particularly the -52 to -71 region. The essential regions are shown at SEQ ID NO: 6 (Figure 10) and, as compared to the genomic sequence, SEQ ID NO: 7 (Figure 5) are bases 1-1088; 830-962; 830-914; 917-962; 875-954; 935-954; and 875-924.

15 EXAMPLE 7

Ms26 Sorghum, Rice and Maize Comparison

As noted above, *Ms26* is a male fertility gene in maize. When it is mutated, and made homozygous recessive, male sterility will result. An orthologue of *Ms26* was identified in sorghum. The sorghum orthologue of the *Ms26* cDNA was isolated by using the maize *Ms26* gene primers in a polymerase chain reaction with sorghum tassel cDNA as the template. The resultant cDNA fragment was sequenced by methods described *supra* and then compared to the *Ms26* cDNA from maize. Nucleotide sequence comparisons are set forth in Figure 12 and show 90% identity. An orthologue from rice was also identified and the predicted coding sequence (SEQ ID NO: 17) and protein (SEQ ID NO: 18) is set forth in Figure 19. It has one intron less than the maize and sorghum *Ms26*, and the coding sequences are highly conserved.

Identification of the sorghum and rice promoters was accomplished. Figure 20 shows an alignment of the *Ms26* promoter of corn (SEQ ID NO: 5), sorghum (SEQ ID NO: 19) and rice (SEQ ID NO: 20). The last three bases of the corn promoter shown in the figure is the ATG start of translation.

Alignment as reflected in Figure 21 of the maize *Ms26* protein (SEQ ID NO: 2), rice *Ms26* protein (SEQ ID NO: 18) and sorghum *Ms26* protein (SEQ ID NO: 4), and a consensus sequence (SEQ ID NO: 21). The comparison of protein

sequences shows the protein is highly conserved among the orthologues, with the rice protein sharing 92% similarity and 86% identity when compared to the maize orthologue. The predicted tissue specificity in rice and sorghum is further reflected in a comparison of the *Ms26* protein in the sorghum and rice EST
 5 database derived from panicle (flower) libraries. Sorghum sequences producing significant alignments (GenBank accession numbers BI075441.1; BI075273.1; BI246000.1; BI246162.1; BG948686.1; BI099541.1 and BG948366.1, among others) all were sequences from immature panicle of sorghum, and sequences showing significant alignment in rice (GenBank accession numbers C73892.1;
 10 CR290740.1, among others) were also from rice immature panicle.

As is evident from the above, nucleotide sequences which map to the short arm of chromosome 1 of the *Zea mays* genome, at the same site as the *Ms26* gene, *ms26-m2::Mu8* and its alleles, are genes critical to male fertility in plants, that is, are necessary for fertility of a plant, or, when mutated from the sequence
 15 found in a fertile plant, cause sterility in the plant.

EXAMPLE 8

Construction of a plant transformation vector comprising a selectable marker, a male fertility gene *Ms45* and a pollen cytotoxin gene.

A construct designated PHP18091, shown in Figure 22 is made by assembling
 20 following DNA components:

1. The plasmid pSB11 backbone DNA (pSB31 lacking the EcoRI fragment carrying the 35SGUS and 35SBAR genes, Ishida et al., *Nature Biotechnol.* (1996) 14:745-750). This DNA backbone contains T-DNA border sequences
 25 and the replication origin from pBR322.
2. The 35S:PAT gene which encodes the enzyme phosphinothricin acetyltransferase (PAT) from *Streptomyces viridochromagenes* (nucleotides 6-557 from accession number A02774, Strauch et al. 1988, EP 0275957-A) under the transcriptional control of the cauliflower mosaic virus (CaMV) 35S
 30 promoter and terminator (nucleotides 6906-7439, and 7439-7632, respectively from Franck et al. 1980, *Cell* 21: 285-294).
3. The 5126:*Ms45* gene which contains the maize male fertility gene coding region (nucleotides 1392-3343, accession number AF360356, Albertsen et a

Am. J. Bot. . (1993) 80:16) under the control of the maize anther-specific promoter 5126 (nucleotides 985-1490, accession number I75204).

4. The PG47:DAM gene which contains the *E. coli* DNA (Adenosine-N⁶) methyltransferase (DAM) coding region (nucleotides 195-1132, Brooks et al., *Nucleic. Acids Res* (1983) 11: 837-851) driven by the maize pollen-specific promoter PG47 (nucleotides 1-2870, accession number X66692, Allen and Lonsdale, *Plant J.* (1993)3:261-271). The transcription of this gene is terminated by the potato proteinase inhibitor II (PinII) terminator (nucleotides 2-310, An et al., *Plant Cell* (1989) 1: 115-122).
5. A 3.34 kb NcoI DNA fragment containing *Ms45:Ms45* was cloned upstream of the 35S:PAT gene in pUC8, creating PHP6641. A 4.7 kb HindIII/EcoRI DNA fragment containing *Ms45:Ms45*-35S:PAT from PHP6641 was cloned into pSB11, creating PHP10890 (Cigan et al, *Sex. Plant Reprod.* (2001)14: 135-142). The native *Ms45* promoter in PHP10890 was replaced by a 528 bp HindIII/NcoI fragment containing the maize 5126 promoter, creating PHP11943.
6. A 2.87 kb HindIII/NcoI fragment containing PG47 promoter was ligated with a 0.8 kb NcoI/HindIII fragment containing the DAM coding region, PinII terminator and 35S enhancer which was from PHP10404 (Unger, et al., *Transgenic Res.* (2001)10: 409-422), creating a 3.67 kb fragment HindIII fragment containing PG47:DAM gene fusion (with the 35S enhancer). This 3.67 kbp HindIII fragment was then cloned into the HindIII site of PHP11943, creating PHP20005. The 35S enhancer in PHP20005 was removed, creating PHP18071. The PHP18071 was introduced into *Agrobacterium* strain LBA4404 carrying plasmid pSB1 by triparental mating (Ishida et al., *Nature Biotechnol.* (1996) 14:745-750). The co-integrate of PHP18071 and pSB1 was named PHP18091.

EXAMPLE 9

Transformation of corn with the restoring transgene construct of Example 8.

- 30 A male-sterile female which was homozygous for an *ms45* mutant *Ac* excision allele, *ms45*'-9301 (*ms45*) was repeatedly crossed with bulked pollen from maize Hi-type II plants (Armstrong 1994, In: Freeling and Walbot (eds). The Maize Handbook. Springer, New York, pp 663-671) resulting in the introgression of this *ms45* allele in transformation amenable maize germplasm over multiple

generations. The resultant source of material for transformation consisted of embryos segregating (1:1 or 3:1) for *ms45* and allowed for both transformation directly into a homozygous *ms45* background and to test the genetic complementation of the *ms45* mutation in T₀ plants. *Agrobacterum*-mediated transformation was performed according to Zhao et al. 1999, (United States Patent number 5,981,840). Genotyping and molecular analysis (integration and PTU) of transformants were done according Cigan et al., (*Sex. Plant. Reprod.* (2001)14:135-142). Transformants with single-integration and complete PTU were selected for further studies.

10 EXAMPLE 10

Analysis of maize transformants.

Transgenic plants (T₀) from Example 9 were evaluated for the whole plant morphology and analyzed for transgene transmission through both pollen and egg cells. No morphological difference was observed between the transgenic plants and the non-transgenic control plants except for the degree of male fertility. Transformants with single-integration and intact PTU were partial male fertile while non-transgenic control plants were completely male sterile, indicating that the expression of *Ms45* gene complemented the homozygous recessive *ms45* male sterile phenotype. This also demonstrated that the expression of the DAM gene caused partial male sterility by eliminating the pollen grains carrying the transgenes. Without the DAM gene, *Ms45* transgene can completely recover the *ms45* male sterile mutation (Cigan et al., *Sex. Plant. Reprod.* (2001) 14: 135-142). The correct function of DAM gene was further determined by controlled pollinations between T₀ transgenic plants and non-transgenic plants. Pollen grains from T₀ transgenic plants were used to pollinate non-transgenic plants control plants. Immature embryos were harvested from ears of these non-transgenic plants 18 days after pollination and cultured either on MS media or MS media containing 3.0 mg/L of bialaphos (Murashige, T. and Skoog, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant* (1962) 15: 437-439). 100% of the embryos were able to germinate on control medium while none of the embryos were able to germinate on media containing 3 mg/L of bialaphos, indicating that the restoring transgene construct was not transmitted through pollen to progeny.

In addition, pollen from non-transgenic plants was used to pollinate the T_0 transgenic maintainer plants. Immature embryos were harvested from ears of these T_0 transgenic maintainer plants 18 days after pollination and cultured as above control media or media containing 3 mg/L of bialaphos. All embryos were able to germinate on control medium while 50% of the embryos were able to germinate on the medium containing bialaphos, indicating that the restoring transgene construct was transmitted through the ovule to progeny at the expected frequency. The results of embryo rescues are summarized in Tables 1 and 2.

Table 1: Transgene transmission through pollen

Transgenic plants	Pollen to non-transgenic plants					
	Control medium			Medium + 3mg/l bialaphos		
	# embryo cultured	# embryo germinated	%	# embryo cultured	# embryo germinated	%
14089263	40	40	100	60	0	0
14089277	100	100	100	100	0	0
14089839	40	40	100	60	0	0

Table 2: Transgene transmission through egg cells

Transgenic plants	Pollen from non-transgenic plants		
	Medium + 3mg/l bialaphos		
	# embryo cultured	# embryo germinated	%
14089262	20	8	40
14089277	40	22	55
14089839	40	21	53

EXAMPLE 11

Conversion of T_0 plants into different inbred lines and analysis of T_n plants.

T_0 transgenic maintainer plants from Example 9 were converted into different inbred backgrounds through repeated backcross by pollination from inbred lines such as PH09B. To accomplish this, pollen produced by PH09B that is *ms45* heterozygous background were used to pollinate the ears of T_0 maintainer plants that were homozygous for the *ms45* mutant alleles. T_1 seed

harvested from these T₀ plants segregated for both transgenes and *ms45* alleles. T₁ plants that did not contain the restoring transgene construct were eliminated by herbicide selection. T₁ plants containing transgenes were analyzed for *ms45* background and male fertility according to Cigan et al., (*Sex. Plant. Reprod.*, 5 (2001) 14: 135-142). In general, T₁ plants in homozygous *ms45* condition that contained the restoring transgene construct showed partial male fertility like that observed for the T₀ parent plants, while the T₁ plants in homozygous *ms45* condition but containing no transgenes were complete male sterile. This suggested that the *Ms45* transgene continued to function correctly in a different 10 genetic background. Pollen grains from T₁ plants were examined for viability using microscopic and histochemical staining. Pollen grains at different developmental stages were collected and stained with fluorescein diacetate (FDA), 4', 6-diamidino-2-phenylindole (DAPI) and ethidium bromide (EB). About 50% of the pollen grains from the transgenic T₁ plants lost their viability as judged by the 15 absence of fluorescence after staining with FDA after first pollen mitosis, while the pollen grains from non-transgenic control plants showed uniform FDA staining. This was further supported by in vitro pollen germination studies. The germination rate of the pollen grains from the transgenic T₁ plants were about half of that from non-transgenic control plants. Pollen grains from transgenic T₁ plant were also 20 used to pollinate non-transgenic plants to test transgene transmission through pollen. For instance, none of 248 embryos from a non-transgenic plant pollinated by a T₁ plant (20118954) were able to germinate on the medium containing 3 mg/l bialaphos. These experiments confirmed both the correct function of the *Ms45* and DAM transgenes in different genetic backgrounds. The T₁ plants with desired 25 performance were used for the next backcross iteration using pollen from the paternal inbred parent which was heterozygous for the mutant *ms45* allele. This process will be repeated until sixth generation.

EXAMPLE 12

Large scale transmission and maintenance of *ms45* male sterility using the construct of Example 8.

T₁ plants derived from T₀ 14089277 as described in example 9 were used as
 5 males to pollinate either wild type inbred plants or *ms45/ms45* male sterile inbred
 plants. The 10,117 T₂ progeny from the wild type crosses and 6688 T₂ progeny
 from the *ms45/ms45* crosses were evaluated for transgene transmission by
 screening for herbicide resistance. For both types of crosses a total of 16786 T₂
 plants were found to be herbicide sensitive, yielding a non-transmission frequency
 10 of 99.89%. All T₂ plants from the *ms45/ms45* crosses that did not contain the
 transgene, were completely male sterile, indicating that this transgenic line can
 maintain *ms45* sterility.

EXAMPLE 13

Construction of a plant transformation vector comprising a screenable marker, a
 15 male fertility gene *Ms26* and a pollen cytotoxin gene.

A construct designated PHP24101, shown in Figure 23, is made by assembling
 following DNA components:

1. The plasmid pSB11 backbone DNA (pSB31 lacking the EcoRI fragment
 carrying the 35SGUS and 35SBAR genes, Ishida et al., *Nature Biotechnol.*
 20 (1996) 14:745-750). This DNA backbone contains T-DNA border sequences
 and the replication origin from pBR322.
2. The PG47PRO:ZM-AA1 gene which contains alpha-amylase 1 coding region
 from *Zea mays* as set forth in Figure 24. (SEQ ID NO: 26). The transcription
 of this gene is terminated by IN2-1 terminator (U.S. Patent No. 5,364,780).
- 25 3. The *Ms26* (SB200) GENOMIC gene (SEQ ID NO: 7) which contains the maize
 male fertility gene coding region.
4. LTP2:DS-RED2 (ALT1) which contains red florescence coding region (a
 variant of *Discosoma* sp. red fluorescent protein (DsRed), from Clontech
 mutated to remove BstEI site, codon sequence unchanged) driven by LTP2
 30 promoter, *supra*.
5. A 2.143 kb EcoRV/DraI DNA fragment containing LTP2PRO:DS-RED2 (ALT1)
 from PHP21737 was cloned into downstream of the *Ms26* GENOMIC gene in
 SK vector, creating SK-*Ms26* GENOMIC-LTP2PRO:DS-RED2 (ALT1).

6. A 2.143 kb EcoRV/DraI DNA fragment containing LTP2PRO:DS-RED2 (ALT1) from PHP21737 was cloned into downstream of the *Ms45*PRO:*Ms45* GENOMIC gene in SK vector, creating SK-*Ms45*-LTP2PRO:DS-RED2 (ALT1).
7. A 5.429 kb NotI fragment containing 5126PRO:*Ms45* GENOMIC-UBI:MOPAT:PINII in PHP20532 was replaced by A 4.318 kb NotI fragment containing *Ms45*-LTP2PRO:DS-RED2 (ALT1) from SK-*Ms45*-LTP2PRO:DS-RED2 (ALT1), creating PHP22623.
8. A 4.318 kb NotI fragment containing *Ms45*-LTP2PRO:DS-RED2 (ALT1) in PHP22623 was replaced by A 5.960 kb NotI DNA fragment containing *Ms26* GENOMIC-LTP2PRO:DS-RED2(ALT1) from SK-*Ms26* GENOMIC-LTP2PRO:DS-RED2 (ALT1), creating PHP24014. The PHP24014 was introduced into *Agrobacterium* strain LBA4404 carrying plasmid pSB1 by Electrophoresis. Co-integrate of PHPPHP24014 and pSB1 was named PHP24101.

15 EXAMPLE 14

Transformation of corn with the restoring transgene construct of Example13 .

A male-sterile female which was homozygous for a *ms26* mutant excision allele, (*ms26*) was repeatedly crossed with bulked pollen from maize Hi-type II plants (Armstrong 1994, In: Freeling and Walbot (eds). The Maize Handbook. Springer, New York, pp 663-671) resulting in the introgression of this *ms26* allele in transformation amenable maize germplasm over multiple generations. The resultant source of material for transformation consisted of embryos segregating (1:1 or 3:1) for *ms26* and allowed for both transformation directly into a homozygous *ms26* background and to test the genetic complementation of the *ms26* mutation in T₀ plants. *Agrobacterium*-mediated transformation was performed according to Zhao et al. 1999, (United States Patent number 5,981,840). Genotyping and molecular analysis (integration and PTU) of transformants were done according Cigan et al., (*Sex. Plant. Reprod.* 1(2001) 4:135-142). Transformants with single-integration and complete PTU were selected for further studies.

EXAMPLE 15 - Analysis of maize transformants.

Transgenic plants (T₀) from Example 14 were evaluated for the whole plant morphology and analyzed for transgene transmission through pollen. No morphological difference was observed between the transgenic plants and the

non-transgenic control plants except for the degree of male fertility. Transformants with single-integration and intact PTU were partial male fertile while non-transgenic control plants were completely male sterile, indicating that the expression of the *Ms26* gene complemented the homozygous recessive *ms26* male sterile phenotype. This also suggested that the pollen expression of the alpha amylase (AA) gene caused partial male sterility by disrupting the normal function of the pollen grains carrying the transgenes. Staining pollen from transformants with potassium iodide (KI), which stains starch granules, showed that approximately half of the pollen grains contained starch (black grains, non-transgenic) and the other half did not contain starch (gold grains, transgenic). The correct function of AA gene was further determined by controlled pollinations between T₀ transgenic plants and non-transgenic plants. Resultant T₁ kernels were evaluated for a red fluorescence phenotype. If the transgenes were transmitted through the pollen then the T₁ seed would contain red fluorescent kernels due to the expression of RFP in the aleurone layer. For four independent events shown in Table 3, no RFP expression was found in the T₁ seed, whereas seed from the T₀ ears themselves (T₁ seed) contained approximately 50% red fluorescent kernels.

Table 3: Transgene transmission through pollen

Transgenic plants	Pollen to non-transgenic plants		
	Kernel Red Fluorescence		
	# Yellow Kernels	# Red Kernels	%
42772379	338	0	100
42772385	277	0	100
42772400	268	0	100
42772411	598	0	100

Thus it can be seen that the invention achieves at least all of its objectives.

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.

THIS IS VOLUME 1 OF 2

NOTE: For additional volumes please contact the Canadian Patent Office.

What is claimed is:

1. A method of maintaining a homozygous recessive condition of a plant, the method comprising:

(a) providing a first plant comprising homozygous recessive alleles of a gene required for the phenotypic expression of a plant trait;

(b) introducing a construct into a second plant, the second plant comprising homozygous recessive alleles of a gene required for the phenotypic expression of a plant trait, the construct in the hemizygous condition comprising:

(i) a first nucleotide sequence that when expressed in the first plant functionally compliments the homozygous recessive plant trait;

(ii) a second nucleotide sequence, the expression of which inhibits the function or formation of male gametes in the second plant; such that male gametes are produced in the second plant containing the recessive alleles and that do not contain the construct; and

(c) fertilizing the first plant with the male gametes of the second plant to produce progeny which maintain the homozygous recessive condition of the first plant.

2. The method of claim 1, wherein the first nucleotide sequence is operably linked to a third nucleotide sequence directing expression preferentially to male plant cells.

3. The method of claim 1, wherein the second nucleotide sequence is operably linked to a fourth nucleotide sequence, the fourth nucleotide sequence directing expression preferentially to male gametes.

4. The method of claim 1, further comprising a fifth nucleotide sequence encoding a product capable of being used for selection of plant cells having the construct.

5. The method of claim 1, further comprising selecting for said second plant by identifying plants having said construct.

6. The method of claim 1, wherein the trait is male fertility.

7. A method of producing seed from a plant having female and male gametes, the method comprising:

(a) introducing a construct into a first plant, the plant comprising homozygous recessive alleles of a gene required for phenotypic expression of a plant trait, the construct in a hemizygous condition comprising:

(i) a first nucleotide sequence, the sequence that when expressed in a second plant having homozygous recessive alleles of a gene required for male fertility and is male sterile, would restore male fertility;

(ii) a second nucleotide sequence operably linked to the first nucleotide sequence which inhibits the function or formation of male gametes in the plant, such that viable male gametes not having the first nucleotide sequence are produced in the plant;

(b) self fertilizing the plant; and

(c) producing seed which contain the construct.

8. The method of claim 7, wherein the first nucleotide sequence is operably linked to a third nucleotide sequence directing expression preferentially to male plant cells.

9. The method of claim 7, wherein the second nucleotide sequence is operably linked to a fourth nucleotide sequence, the fourth nucleotide sequence directing expression to male gametes.

10. The method of claim 7, further comprising a fifth nucleotide sequence encoding a product capable of being used for selection of plant cells having the construct.

11. The method of claim 7, further comprising identifying plants having said construct.

12. The method of claim 7, wherein the trait is male fertility.

Application number / numéro de demande: 27 55 243

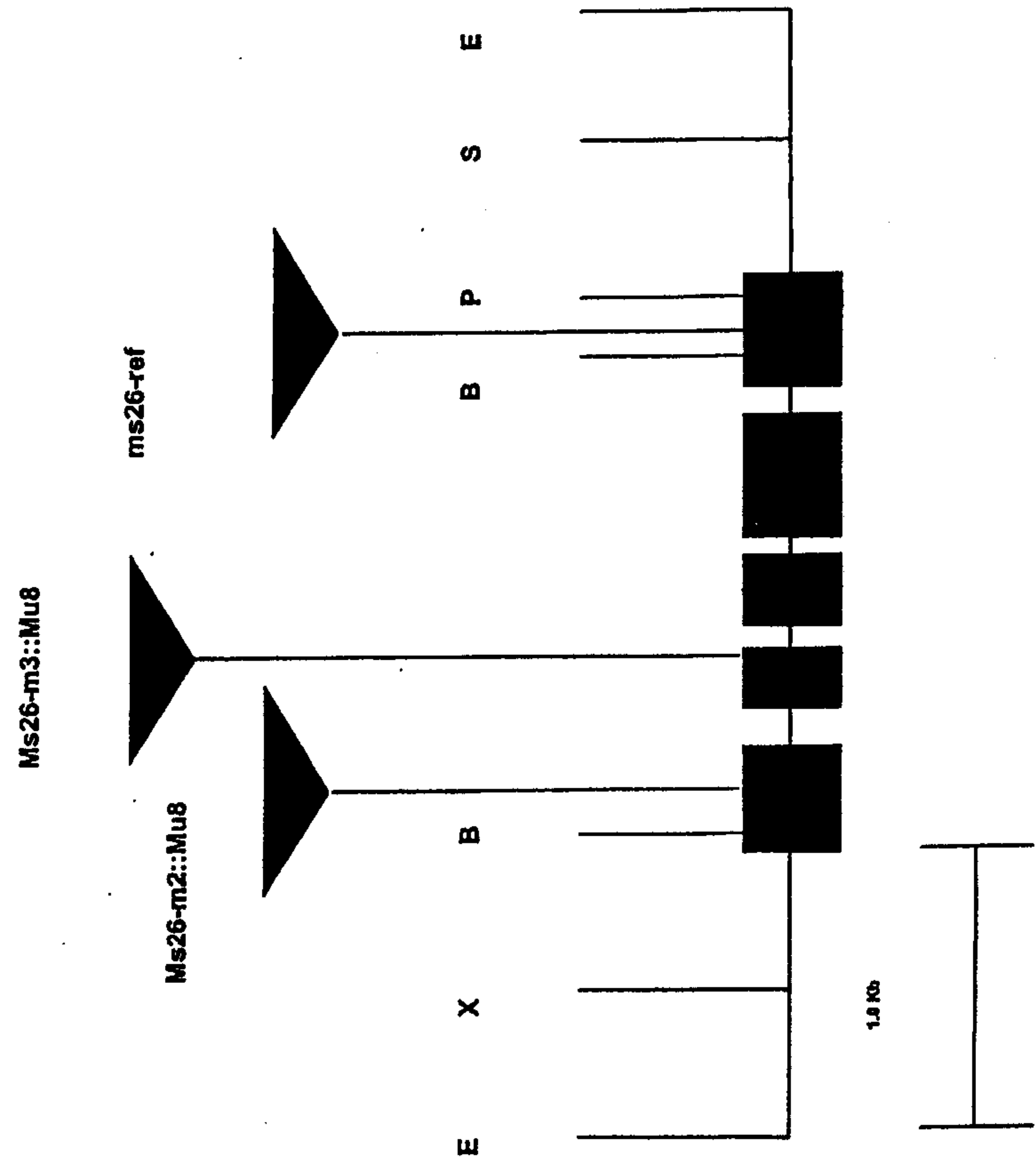
Figures: _____

Pages: 2, 3, 16, 28

Unscannable items
received with this application
(Request original documents in File Prep. Section on the 10th floor)

Documents reçu avec cette demande ne pouvant être balayés
(Commander les documents originaux dans la section de préparation des dossiers au
10ème étage)

Figure 1



4/34

EcoRI

GAATTCGGCAGGAGGAAGCTCACCTCAGCCGGCGACGCCATCGCCATTCTTCCCACTA 60
 CTTAAGCCGTGCTCCCTTCGAGTGGAGTGGCGCCGCTGCGGTAGCGGTAAGAAGGGTGAT
 E F G T R E A H L T P A T P S P F F P L -
 GCAGGGCCTCACAAGTACATCGCGCTCCTTCTGGTTGTCCTCTCATGGATCCTGGTCCAG 120
 61 CGTCCCGGAGTGTTTCATGTAGCGCGAGGAAGACCAACAGGAGAGTACCTAGGACCAGGTC
 A G P H K Y I A L L L V V L S W I L V Q -
 AGGTGGAGCCTGAGGAAGCAGAAAGGCCCGAGATCATGGCCAGTCATCGGCGCAACGGTG 180
 121 TCCACCTCGGACTCCTTCGTCTTCCGGGCTCTAGTACCGGTCAGTAGCCGCGTTGCCAC
 R W S L R K Q K G P R S W P V I G A T V -
 GAGCAGCTGAGGAACTACCACCGGATGCACGACTGGCTTGTGCGGTACCTGTCACGGCAC 240
 181 CTCGTCGACTCCTTGATGGTGGCCTACGTGCTGACCGAACAGCCCATGGACAGTGCCGTG
 E Q L R N Y H R M H D W L V G Y L S R H -
 AGGACAGTGACCGTCGACATGCCGTTCACTTCCTACACCTACATCGCTGACCCGGTGAAT 300
 241 TCCTGTCACTGGCAGCTGTACGGCAAGTGAAGGATGTGGATGTAGCGACTGGGCCACTTA
 R T V T V D M P F T S Y T Y I A D P V N -
 GTCGAGCATGTCCTCAAGACTAACTTCACCAATTACCCCAAGGGAATCGTGTACAGATCC 360
 301 CAGCTCGTACAGGAGTTCTGATTGAAGTGGTTAATGGGGTTCCCTTAGCACATGTCTAGG
 V E H V L K T N F T N Y P K G I V Y R S -
 TACATGGACGTGCTCCTCGGTGACGGCATCTTCAACGCCGACGGCGAGCTGTGGAGGAAG 420
 361 ATGTACCTGCACGAGGAGCCACTGCCGTAGAAGTTGCGGCTGCCGCTCGACACCTCCTTC
 Y M D V L L G D G I F N A D G E L W R K -
 CAGAGGAAGACGGCGAGTTTCGAGTTCGCCTCCAAGAACCTGAGGGATTTCAGCGCCATT 480
 421 GTCTCCTTCTGCCGCTCAAAGCTCAAGCGGAGGTTCTTGACTCCCTAAAGTCGCGGTAA
 Q R K T A S F E F A S K N L R D F S A I -

FIG. 4A

5/34

```

481 GTGTTTCAGAGAGTACTCCCTGAAGCTGTCGGGTATACTGAGCCAGGCATCCAAGGCAGGC 540
-----+-----+-----+-----+-----+-----+
CACAAGTCTCTCATGAGGGACTTCGACAGCCCATATGACTCGGTCCGTAGGTTCCGTCCG
V F R E Y S L K L S G I L S Q A S K A G -

541 AAAGTTGTGGACATGCAGGAACTTTACATGAGGATGACGCTGGACTCCATCTGCAAGGTT 600
-----+-----+-----+-----+-----+-----+
TTTCAACACCTGTACGTCCTTGAAATGTACTCCTACTGCGACCTGAGGTAGACGTTCCAA
K V V D M Q E L Y M R M T L D S I C K V -

601 GGGTTCGGGGTCGAGATCGGCACGCTGTCGCCAGATCTCCCCGAGAACAGCTTCGCGCAG 660
-----+-----+-----+-----+-----+-----+
CCCAAGCCCCAGCTCTAGCCGTGCGACAGCGGTCTAGAGGGGCTCTTGTCGAAGCGCGTC
G F G V E I G T L S P D L P E N S F A Q -

661 GCGTTCGATGCCGCCAACATCATCATCACGCTGCGGTTTCATCGACCCGCTGTGGCGCATC 720
-----+-----+-----+-----+-----+-----+
CGCAAGCTACGGCGGTTGTAGTAGTAGTGCGACGCCAAGTAGCTGGGCGACACCGCGTAG
A F D A A N I I I T L R F I D P L W R I -

721 AAGAGGTTCTTCCACGTCGGGTCAGAGGCCCTCCTAGCGCAGAGCATCAAGCTCGTGGAC 780
-----+-----+-----+-----+-----+-----+
TTCTCCAAGAAGGTGCAGCCCAGTCTCCGGGAGGATCGCGTCTCGTAGTTCGAGCACCTG
K R F F H V G S E A L L A Q S I K L V D -

781 GAGTTCACCTACAGCGTGATCCGCCGGAGGAAGGCCGAGATCGTCGAGGTCCGGGGCCAGC 840
-----+-----+-----+-----+-----+-----+
CTCAAGTGGATGTCGCACTAGGCGGCCTCCTTCCGGCTCTAGCAGCTCCAGGCCCGGTCTG
E F T Y S V I R R R K A E I V E V R A S -

841 GGCAAACAGGAGAAGATGAAGCACGACATCCTGTCACGGTTCATCGAGCTGGGCGAGGCC 900
-----+-----+-----+-----+-----+-----+
CCGTTTGTCTCTTCTACTTCGTGCTGTAGGACAGTGCCAAGTAGCTCGACCCGCTCCGG
G K Q E K M K H D I L S R F I E L G E A -

901 GGCGACGACGGCGGCGGCTTCGGGGACGATAAGAGCCTCCGGGACGTGGTGCTCAACTTC 960
-----+-----+-----+-----+-----+-----+
CCGCTGCTGCCGCCGCCGAAGCCCCTGCTATTCTCGGAGGCCCTGCACCACGAGTTGAAG
G D D G G G F G D D K S L R D V V L N F -

961 GTGATCGCCGGGCGGGACACGACGGCGACGACGCTGTCGTGGTTCACGCACATGGCCATG 1020
-----+-----+-----+-----+-----+-----+
CACTAGCGGCCCCGCTGTGCTGCCGCTGCTGCGACAGACCAAGTGCCTGTACCGGTAC
V I A G R D T T A T T L S W F T H M A M -

```

FIG. 4B

6/34

```

TCCCACCCGGACGTGGCCGAGAAGCTGCGCCGCGAGCTGTGCGCGTTTCGAGGCGGAGCGC 1080
1021 -----+-----+-----+-----+-----+-----+
AGGGTGGGCCTGCACCGGCTCTTCGACGCGGCGCTCGACACGCGCAAGCTCCGCCTCGCG
S H P D V A E K L R R E L C A F E A E R -
GCGCGCGAGGAGGGCGTCACGCTCGTGCTCTGCGGCGGCGCTGACGCCGACGACAAGGCG
1081 -----+-----+-----+-----+-----+-----+ 1140
CGCGCGCTCCTCCCGCAGTGCAGACGAGACGCCGCCGCGACTGCGGCTGCTGTTCCGC
A R E E G V T L V L C G G A D A D D K A -
TTCGCCGCCCGCGTGGCGCAGTTCGCGGGCCTCCTCACCTACGACAGCCTCGGCAAGCTG
1141 -----+-----+-----+-----+-----+-----+ 1200
AAGCGGCGGGCGCACCGCGTCAAGCGCCCGAGGAGTGGATGCTGTGCGAGCCGTTTCGAC
F A A R V A Q F A G L L T Y D S L G K L -
GTCTACCTCCACGCCTGCGTCACCGAGACGCTCCGCCTGTACCCCGCCGTCCTCAGGAC
1201 -----+-----+-----+-----+-----+-----+ 1260
CAGATGGAGGTGGGACGCAGTGGCTCTGCGAGGCGGACATGGGGCGGCAGGGAGTCCTG
V Y L H A C V T E T L R L Y P A V P Q D -
CCCAAGGGGATCCTGGAGGACGACGTGCTGCCGGACGGGACGAAGGTGAGGGCCGGCGGG
1261 -----+-----+-----+-----+-----+-----+ 1320
GGGTTCCCCTAGGACCTCCTGCTGCACGACGGCCTGCCCTGCTTCCACTCCCGGCCGCC
P K G I L E D D V L P D G T K V R A G G -
ATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGTACAACCTGGGGCCCCGACGCG
1321 -----+-----+-----+-----+-----+-----+ 1380
TACCACTGCATGCACGGGATGAGCTACCCCGCCTACCTCATGTTGACCCCGGGGCTGCGC
M V T Y V P Y S M G R M E Y N W G P D A -
GCGAGCTTCCGGCCGGAGCGGTGGATCAACGAGGATGGCGCGTTCCGCAACGCGTCGCCG
1381 -----+-----+-----+-----+-----+-----+ 1440
CGCTCGAAGGCCGGCCTCGCCACCTAGTTGCTCCTACCGCGCAAGGCGTTGCGCAGCGGC
A S F R P E R W I N E D G A F R N A S P -
TTCAAGTTCACGGCGTTCCAGGCGGGGCGGAGGATCTGCCTGGGCAAGGACTCGGCGTAC
1441 -----+-----+-----+-----+-----+-----+ 1500
AAGTTCAAGTGCCGCAAGGTCCGCCCCGGCTCCTAGACGGACCCGTTCTGAGCCGCATG
F K F T A F Q A G P R I C L G K D S A Y -
CTGCAGATGAAGATGGCGCTGGCCATCCTCTTCGCTTCTACAGCTTCCGGCTGCTGGAG
1501 -----+-----+-----+-----+-----+-----+ 1560
GACGTCTACTTCTACCGCGACCGGTAGGAGAAGGCGAAGATGTCGAAGGCCGACGACCTC
L Q M K M A L A I L F R F Y S F R L L E -

```

FIG.4C

7/34

GGGCACCCGGTGCAGTACCGCATGATGACCATCCTCTCCATGGCGCACGGCCTCAAGGTC
1561 -----+-----+-----+-----+-----+-----+1620
CCCGTGGGCCACGTCATGGCGTACTACTGGTAGGAGAGGTACCGCGTGCCGGAGTTCCAG
G H P V Q Y R M M T I L S M A H G L K V -
CGCGTCTCTAGGGCCGTCTGATGTCATGGCGATTTGGATATGGATATCGTCCCGCTTAAT
1621 -----+-----+-----+-----+-----+-----+1680
GCGCAGAGATCCCGGCAGACTACAGTACCGCTAAACCTATACCTATAGCAGGGCGAATTA
R V S R A V * C H G D L D M D I V P L N -
CCACGACAAATAACGCTCGTGTTACAAATTTGCATGCATGCATGTAAGGGAAAGCGATGG
1681 -----+-----+-----+-----+-----+-----+1740
GGTGCTGTTTATTGCGAGCACAATGTTTAAACGTACGTACGTACATTCCCTTTTCGCTACC
P R Q I T L V L Q I C M H A C K G K R W -
GTTTCATTGGTGGCTTGGCTTAAGCCTTAAAACTCCGTCGGGTCTTGCGAACCACCACA
1741 -----+-----+-----+-----+-----+-----+1800
CAAAGTAACCACCGAACCGAATTCGGAATTTTGGAGGCAGCCCAGAACGCTTGGTGGTGT
V S L V A W L K P *
TCACTAGTGTGTTTGTACTCTACTCCTCAGTGGAAGTGAGTGACAGCATACAAGTTCATC
1801 -----+-----+-----+-----+-----+-----+1860
AGTGATCACAAAACATGAGATGAGGAGTCACCTTCACATCACTGTCGTATGTTCAAGTAG

XhoI
|
ATATATATTATCCTCTTTCTTAAAAAAAAAAAAAAAAAACTCGAG
1861 -----+-----+-----+-----+-----+1906
TATATATAATAGGAGAAAGAATTTTTTTTTTTTTTTTTTTGAGCTC

FIG. 4D

8/34

1 GAATTCCAAG CGAGGCCCTT GTAGCAGAGA GTGTTGCTGA TGCAGTCGGC
21 GGAAATGAGT GCGTGCTGAG AGCAACGCTG AGGGGTTCCA GGGATGGCAA
101 TGGCTATGGC AATCGGCTAG AGGTGGAGGA CAAGGTGGTG AGGATTGGGA
151 GGGCAACCTA TGGCAAGTTG GTGAAGAGGC ACGCAATGAG AGATCTATTC
201 AGACTTACAC TGGATGCCGC CAACAAATTC AACCTTTAGA TTTTGATACT
251 GTCACTCCTA CTTTATTCCT TGGTTGGGCA ACTTCCAATA GGCTCATGTT
301 AATCAATGAT TAGTGATTAT TCAGCAAATA TTCTTGTTTG TTTGACATTT
351 ATAATATGTG GGGTGAGACG GATTAAATAT CATCCATGAG AGCTTTATCT
401 TCATGCTCTC TTGATTTTGG TTTCAGATCA TTCTTTCAGT GTTCACAAGA
451 ATTTTCTCAG TTTGGTCCAT GTAATTTTGG AAGTGAGGTT CCTTAAATTT
501 CATTATGCTT CCTTTCTTTT CTAGACTAGC AACTGTCATGA CTTTTCACTT
551 TGGGTTTACA AATTGACTCA CAAGAAAACA AATTCACTTT TGGGTTTACA
601 AATTCCTCTT CAGGATGTAC TTTTCACTTG AACTGTCATG TATAGGAACA
651 AGGAATGGCT CAGTTTTTAA GGAACAATGT ACAGATTTCA TTTCAGAACT
701 CTTTCTGGTT GGTGAGTTT CAGACTTTTT GTACCAAGCT GATGGATCAC
751 AATACTTGTT TCCAAAGTCT GATAACAGAA ACTGGCAACT CCTAATTGAT
801 AATAAAAAGA ATAAAATACA GTATCAGATA TCTCATTTTC TTGGTTGGCA
851 GATCACAAAA AGGAACACAA AGGCTAAGCC TCCTACTTGT TCGGGAGTTA
901 GGTGAGGGAC ACCATATGAA TGAAAGAAAT CTTAATTTGG GGTACACCA
951 AGATTGTCTC TCTCGAGGTT GGGGGGTCCC TAAGGTTGGT AGTAGCAATA
1001 CCCAATATAT CACCTAACAA ACCCAATCCA TGCTACATAC ATACATAGCA
1051 TCCATCACTT GTAGACTGGA CCCTTCATCA AGAGCACCAT GGAGGAAGCT
1101 CACATCACGC CGGCGACGCC ATCGCCATTC TTCCCACTAG CAGGGCCTCA
1151 CAAGTACATC GCGCTCCTCC TGGTTGTCCT CTCATGGATC CTGGTCCAGA
1201 GGTGGAGCCT GAGGAAGCAG AAAGGCCCGA GATCATGGCC AGTCATCGGT
1251 GCAACGGTGG AGCAGCTGAG GAACTACCAC CGGATGCACG ACTGGCTTGT
1301 CGGGTACCTG TCACGGCACA GGACAGTGAC CGTCGACATG CCGTTCACTT
1351 CCTACACCTA CATCGCTGAC CCGGTGAATG TCGAGCATGT CCTCAAGACT

FIG. 5A

9/34

1401 AACTTCACCA ATTACCCCAA GGTAAATGAC CTGAACTCAC TGATGTTTCAG
1451 TCTTCGGAAA TCAGAGCTGA AAGCTGAATC GAATGTGCCT GAACACCGTG
1501 TAGGGAATCG TGTACAGATC CTACATGGAC GTGCTCCTCG GTGACGGCAT
1551 CTTCAACGCC GACGGCGAGC TGTGGAGGAA GCAGAGGAAG ACGGCGAGTT
1601 TCGAGTTCGC CTCCAAGAAC CTGAGGGATT TCAGCGCCAT TGTGTTTCAGA
1651 GAGTACTCCC TGAAGCTGTC GGGTATACTG AGCCAGGCAT CCAAGGCAGG
1701 CAAAGTTGTG GACATGCAGG TGAGATCACT GCTCCCTTGC CATTGCCAAC
1751 ATGAGCATTT CAACCTGAGA CACGAGAGCT ACCTTGCCGA TTCAGGAACT
1801 TTACATGAGG ATGACGCTGG ACTCCATCTG CAAGGTTGGG TTCGGGGTCG
1851 AGATCGGCAC GCTGTCGCCG GATCTCCCCG AGAACAGCTT CGCGCAGGCG
1901 TTCGATGCCG CCAACATCAT CGTCACGCTG CGGTTCATCG ACCCGCTGTG
1951 GCGCATCAAG AGGTTCTTCC ACGTCGGGTC AGAGGCCCTC CTAGCGCAGA
2001 GCATCAAGCT CGTGGACGAG TTCACCTACA GCGTGATCCG CCGGAGGAAG
2051 GCCGAGATCG TCGAGGCCCG GGCCAGCGGC AAACAGGAGA AGGTACGTGC
2101 ACATGACTGT TTCGATTCTT CAGTTCATCG TCTTGCCCGG GATGGACCTG
2151 ATCCTGATTG ATTATATATC CGTGTGACTT GTGAGGACAA ATTAAATGG
2201 GCAGATGAAG CACGACATCC TGTACGGTT CATCGAGCTA GGCGAGGCCG
2251 GCGACGACGG CGGCGGCTTC GGGGACGACA AGAGCCTCCG GGACGTGGTG
2301 CTCAACTTCG TGATCGCCGG GCGGGACACG ACGGCGACGA CGCTGTCGTG
2351 GTTCACGCAC ATGGCCATGT CCCACCCGGA CGTGGCCGAG AAGCTGCGCC
2401 GCGAGCTGTG CGCGTTCGAG GCGGAGCGCG CGCGCGAGGA GGGCGTCGCG
2451 CTCGTGCCCT GCGGCGGCGC TGACGCCGAC GACAAGGCGT TCGCCGCCCG
2501 CGTGGCGCAG TTCGCGGGCC TCCTCACCTA CGACAGCCTC GGCAAGCTGG
2551 TCTACCTCCA CGCCTGCGTC ACCGAGACGC TCCGCCTGTA CCCCGCCGTC
2601 CCTCAGGTGA GCGCGCCCGA CACGCGACCT CCGGTCCAGA GCACAGCATG
2651 CAGTGAGTGG ACCTGAATGC AATGCACATG CACTTGCGCG CGCGCAGGAC
2701 CCCAAGGGGA TCCTGGAGGA CGACGTGCTG CCGGACGGGA CGAAGGTGAG
2751 GGCCGGCGGG ATGGTGACGT ACGTGCCCTA CTCGATGGGG CGGATGGAGT

FIG.5B

10/34

2801 ACAACTGGGG CCCCAGCGG GCGAGCTTCC GGCCGGAGCG GTGGATCAAC
2851 GAGGATGGCG CGTTCCGCAA CGCGTCGCCG TTCAAGTTCA CGGCGTTCCA
2901 GGCGGGGCCG AGGATCTGCC TGGGCAAGGA CTCGGCGTAC CTGCAGATGA
2951 AGATGGCGCT GGCCATCCTC TTGCGCTTCT ACAGCTTCCG GCTGCTGGAG
3001 GGGCACCCGG TGCAGTACCG CATGATGACC ATCCTCTCCA TGGCGCACGG
3051 CCTCAAGGTC CGCGTCTCTA GGGCCGTCTG ATGTCATGGC GATTGTTGGAT
3101 ATCATCCCGC TTAATCCTTA AAAATTTGCA TGCATGCATG TAAGGGAAAG
3151 CGATGGGTTT CATTGGTGGC TTGGCTTAAG CCTTAAAAAC TCCGTCGGGT
3201 CTTGCGAACC ACCACATCAC TAGTGTTTTG TACTCTACTC CTCAGTGGAA
3251 GTGTAGTGAC AGCATAAAG TTCATCATAT ATATTATCCT CTTTCTTCGC
3301 CGGATGCTTC CCGGGACCTT TTGGAGACCA TTAAGTACAG GCGTGTGAAA
3351 AAAAGGCTTC TTCTGCGGCG AAGTTTTGGG TTCAGAGTCT TGGCGTCTTT
3401 GCAGCAGAAA AAAGGTTTGG AAGGATCTGA ACCCTGAACC GAAAATGGCT
3451 TCGGAAATAT GCTCGCATCG GGGCGGGGCC GTCAGTGGG ATGACGACAA
3501 GCCCACAAGC AGTGAGAGCG AAGCGATCTT TGGAGTTTGG AGACACTCTC
3551 GGACCCCTCG GCGCTCCGCG AGCTCATCTT CGCCTCCTCT GTCGTGTCCG
3601 TGGCGGCACC GCGCCCGCCC GCCTCGTGTT CGACCAAATC CCGCGCCCCG
3651 ACCGGTTCGT GTACAACACC CTCATCCGCG GCGCCGCGCG CAGTGACACG
3701 CCCCAGGACG CCGTATACAT CTATAAATCA TGGTATTGTA CTTTATTTTC
3751 AAACGGCCTT AACACAACCA TATTTTATG GTAAACACGT TCAAAATTGA
3801 CACAAATTTA AAACAGGCAC AAACCGTAGC TAAACATAAG AGAATGAGAG
3851 ACAACCCAAA GGTAGAGAT GAAATAAGCT GAGTAAACGA CGAATTC

Fig. 5C

11/34

```

1051 TCCATCACTTGTAGACTGGACCCTTCATCAAGAGCACCATGGAGGAAGCT 1100
      1 .....GAATTCGGCACGAGGGAAGCT 21
1101 CACATCACGCCGGCGACGCCATCGCCATTCTTCCCCTAGCAGGGCCTCA 1150
      22 CACCTCACGCCGGCGACGCCATCGCCATTCTTCCCCTAGCAGGGCCTCA 71
1151 CAAGTACATCGCGCTCCTCTGGTTGTCCTCTCATGGATCCTGGTCCAGA 1200
      72 CAAGTACATCGCGCTCCTCTGGTTGTCCTCTCATGGATCCTGGTCCAGA 121
1201 GGTGGAGCCTGAGGAAGCAGAAAGGCCCGAGATCATGGCCAGTCATCGGT 1250
      122 GGTGGAGCCTGAGGAAGCAGAAAGGCCCGAGATCATGGCCAGTCATCGGC 171
1251 GCAACGGTGGAGCAGCTGAGGAACTACCACCGGATGCACGACTGGCTTGT 1300
      172 GCAACGGTGGAGCAGCTGAGGAACTACCACCGGATGCACGACTGGCTTGT 221
1301 CGGGTACCTGTCACGGCACAGGACAGTGACCGTCGACATGCCGTTCACTT 1350
      222 CGGGTACCTGTCACGGCACAGGACAGTGACCGTCGACATGCCGTTCACTT 271
1351 CCTACACCTACATCGCTGACCCGGTGAATGTCGAGCATGTCCTCAAGACT 1400
      272 CCTACACCTACATCGCTGACCCGGTGAATGTCGAGCATGTCCTCAAGACT 321
1401 AACTTCACCAATTACCCCAAGGTAAATGACCTGAACTCACTGATGTTTCA 1450
      322 AACTTCACCAATTACCCCA..... 340
      .
      .
1501 TAGGGAATCGTGTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCAT 1550
      341 .AGGGAATCGTGTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCAT 389
1551 CTTCAACGCCGACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTT 1600
      390 CTTCAACGCCGACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTT 439
1601 TCGAGTTCGCCTCCAAGAACCTGAGGGATTTGAGCGCCATTGTGTTTCA 1650
      440 TCGAGTTCGCCTCCAAGAACCTGAGGGATTTGAGCGCCATTGTGTTTCA 489
1651 GAGTACTCCCTGAAGCTGTCGGGTATACTGAGCCAGGCATCCAAGGCAGG 1700
      490 GAGTACTCCCTGAAGCTGTCGGGTATACTGAGCCAGGCATCCAAGGCAGG 539
1701 CAAAGTTGTGGACATGCAGGTGAGATCACTGCTCCCTTGCCATTGCCAAC 1750
      540 CAAAGTTGTGGACATG.....

```

FIG. 6A

12/34

1751 ATGAGCATTTCAACCTGAGACACGAGAGCTACCTTGCCGATTCAGGAACT 1800
556 CAGGAACT 563
1801 TTACATGAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTTCGGGGTTCG 1850
564 TTACATGAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTTCGGGGTTCG 613
1851 AGATCGGCACGCTGTCGCCGGATCTCCCCGAGAACAGCTTCGCGCAGGCG 1900
614 AGATCGGCACGCTGTCGCCAGATCTCCCCGAGAACAGCTTCGCGCAGGCG 663
1901 TTCGATGCCGCCAACATCATCGTCACGCTGCGGTTTCATCGACCCGCTGTG 1950
664 TTCGATGCCGCCAACATCATCATCACGCTGCGGTTTCATCGACCCGCTGTG 713
1951 GCGCATCAAGAGGTTCTTCCACGTCGGGTCAGAGGCCCTCCTAGCGCAGA 2000
714 GCGCATCAAGAGGTTCTTCCACGTCGGGTCAGAGGCCCTCCTAGCGCAGA 763
2001 GCATCAAGCTCGTGGACGAGTTCACCTACAGCGTGATCCGCCGGAGGAAG 2050
764 GCATCAAGCTCGTGGACGAGTTCACCTACAGCGTGATCCGCCGGAGGAAG 813
2051 GCCGAGATCGTCGAGGCCCGGGCCAGCGGCAAACAGGAGAAGGTACGTGC 2100
814 GCCGAGATCGTCGAGGTCCGGGCCAGCGGCAAACAGGAGA..... 853
:
:
2201 GCAGATGAAGCACGACATCCTGTACGGTTCATCGAGCTAGGCGAGGCCG 2250
854 ..AGATGAAGCACGACATCCTGTACGGTTCATCGAGCTAGGCGAGGCCG 901
2251 GCGACGACGGCGGGCTTCGGGGACGACAAGAGCCTCCGGGACGTGGTG 2300
902 GCGACGACGGCGGGCTTCGGGGACGATAAGAGCCTCCGGGACGTGGTG 951
2301 CTCAACTTCGTGATCGCCGGGCGGGACACGACGGCGACGCTGTCTGTG 2350
952 CTCAACTTCGTGATCGCCGGGCGGGACACGACGGCGACGCTGTCTGTG 1001
2351 GTTCACGCACATGGCCATGTCCCACCCGGACGTGGCCGAGAAGCTGCGCC 2400
1002 GTTCACGCACATGGCCATGTCCCACCCGGACGTGGCCGAGAAGCTGCGCC 1051
2401 GCGAGCTGTGCGCGTTCGAGGCGGAGCGCGCGCGAGGAGGGCGTCGCG 2450
1052 GCGAGCTGTGCGCGTTCGAGGCGGAGCGCGCGCGAGGAGGGCGTCACG 1101
2451 CTCGTGCCCTGCGGCGGCGCTGACGCCGACGACAAGGCGTTCGCCGCCCG 2500
1102 CTCGTGCTCTGCGGCGGCGCTGACGCCGACGACAAGGCGTTCGCCGCCCG 1151

Fig. 6B

13/34

```
2501 CGTGGCGCAGTTCGCGGGCCTCCTCACCTACGACAGCCTCGGCAAGCTGG 2550
      |||
1152 CGTGGCGCAGTTCGCGGGCCTCCTCACCTACGACAGCCTCGGCAAGCTGG 1201
2551 TCTACCTCCACGCCTGCGTCACCGAGACGCTCCGCCTGTACCCCGCCGTC 2600
      |||
1202 TCTACCTCCACGCCTGCGTCACCGAGACGCTCCGCCTGTACCCCGCCGTC 1251
2601 CCTCAGGTGAGCGCGCCCGACACGCGACCTCCGGTCCAGAGCACAGCATG 2650
      |||
1252 CCT..... 1254
2651 CAGTGAGTGGACCTGAATGCAATGCACATGCACTTGCGCGCGCGCAGGAC 2700
      |||
1255 .....CAGGAC 1260
2701 CCCAAGGGGATCCTGGAGGACGACGTGCTGCCGGACGGGACGAAGGTGAG 2750
      |||
1261 CCCAAGGGGATCCTGGAGGACGACGTGCTGCCGGACGGGACGAAGGTGAG 1310
2751 GGCCGGCGGGATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGT 2800
      |||
1311 GGCCGGCGGGATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGT 1360
2801 ACAACTGGGGCCCCGACGCGGCGAGCTTCCGGCCGGAGCGGTGGATCAAC 2850
      |||
1361 ACAACTGGGGCCCCGACGCGGCGAGCTTCCGGCCGGAGCGGTGGATCAAC 1410
2851 GAGGATGGCGCGTTCCGCAACGCGTCGCCGTTCAAGTTCACGGCGTTCCA 2900
      |||
1411 GAGGATGGCGCGTTCCGCAACGCGTCGCCGTTCAAGTTCACGGCGTTCCA 1460
2901 GGCGGGGCGGAGGATCTGCCTGGGCAAGGACTCGGCGTACCTGCAGATGA 2950
      |||
1461 GGCGGGGCGGAGGATCTGCCTGGGCAAGGACTCGGCGTACCTGCAGATGA 1510
2951 AGATGGCGCTGGCCATCCTCTTGCGCTTCTACAGCTTCCGGCTGCTGGAG 3000
      |||
1511 AGATGGCGCTGGCCATCCTCTTGCGCTTCTACAGCTTCCGGCTGCTGGAG 1560
3001 GGGCACCCGGTGACGTACCGCATGATGACCATCCTCTCCATGGCGCACGG 3050
      |||
1561 GGGCACCCGGTGACGTACCGCATGATGACCATCCTCTCCATGGCGCACGG 1610
3051 CCTCAAGGTCCGCGTCTCTAGGGCCGTCTGATGTCATGGCGATTG.... 3096
      |||
1611 CCTCAAGGTCCGCGTCTCTAGGGCCGTCTGATGTCATGGCGATTGATA 1660
3097 .GGATATCATCCCGCTTAATCC.....TTAAAAATT 3126
      |||
1661 TGGATATCGTCCCGCTTAATCCACGACAAATAACGCTCGTGTACAAATT 1710
3127 TGCATGCATGCATGTAAGGGAAAGCGATGGGTTTCATTGGTGGCTTGGCT 3176
      |||
1711 TGCATGCATGCATGTAAGGGAAAGCGATGGGTTTCATTGGTGGCTTGGCT 1760
```

Fig.6C

14/34

3177 TAAGCCTTAAAACTCCGTCGGGTCTTGCGAACCACCACATCACTAGTGT 3226
|||||
1761 TAAGCCTTAAAACTCCGTCGGGTCTTGCGAACCACCACATCACTAGTGT 1810
3227 TTTGTACTCTACTCCTCAGTGGAAGTGTAGTGACAGCATACAAGTTCATC 3276
|||||
1811 TTTGTACTCTACTCCTCAGTGGAAGTGTAGTGACAGCATACAAGTTCATC 1860
3277 ATATATATTATCCTCTTTCTTCGCCGGATGCTTCCCGGGACCTTTGGAG 3326
|||||
1861 ATATATATTATCCTCTTTCTTAAAAAAAAAAAAAAAAAACTCGAG.... 1906

Fig. 6D

16/34

1 GAATTCCAAG CGAGGCCCTT GTAGCAGAGA GTGTTGCTGA TGCAGTCGGC
51 GGAAATGAGT GCGTGCTGAG AGCAACGCTG AGGGGTTCCA GGGATGGCAA
101 TGGCTATGGC AATCGGCTAG AGGTGGAGGA CAAGGTGGTG AGGATTGGGA
151 GGGCAACCTA TGGCAAGTTG GTGAAGAGGC ACGCAATGAG AGATCTATTC
201 AGACTTACAC TGGATGCCGC CAACAAATTC AACCTTTAGA TTTTGATACT
251 GTCACTCCTA CTTTATTCCT TGGTTGGGCA ACTTCCAATA GGCTCATGTT
301 AATCAATGAT TAGTGATTAT TCAGCAAATA TTCTTGTTTG TTTGACATTT
351 ATAATATGTG GGGTGAGACG GATTAAATAT CATCCATGAG AGCTTTATCT
401 TCATGCTCTC TTGATTTTGG TTTCAGATCA TTCTTTCAGT GTTCACAAGA
451 ATTTTCTCAG TTTGGTCCAT GTAATTTTGG AAGTGAGGTT CCTTAAATTT
501 CATTATGCTT CCTTTCTTTT CTAGACTAGC AACTGCATGA CTTTTCACCT
551 TGGGTTTACA AATTGACTCA CAAGAAAACA AATTCACCTT TGGGTTTACA
601 AATTCCTCTT CAGGATGTAC TTTTCACTTG AACTGTCATG TATAGGAACA
651 AGGAATGGCT CAGTTTTTAA GGAACAATGT ACAGATTTCA TTTCAGAACT
701 CTTTCTGGTT GGTTGAGTTT CAGACTTTTT GTACCAAGCT GATGGATCAC
751 AATACTTGTT TCCAAAGTCT GATAACAGAA ACTGGCAACT CCTAATTGAT
801 AATAAAAAGA ATAAAATACA GTATCAGATA TCTCATTTTC TTGGTTGGCA
851 GATCACAAAA AGGAACACAA AGGCTAAGCC TCCTACTTGT TCGGGAGTTA
901 GGTGAGGGAC ACCATATGAA TGAAAGAAAT CTTAATTTGG GGTCACACCA
951 AGATTGTCTC TCTCGAGGTT GGGGGGTCCC TAAGGTTGGT AGTAGCAATA
1001 CCCAATATAT CACCTAACAA ACCCAATCCA TGCTACATAC ATACATAGCA
1051 TCCATCACTT GTAGACTGGA CCCTTCATCA AGAGCACCAT GG

Fig.8

17/34

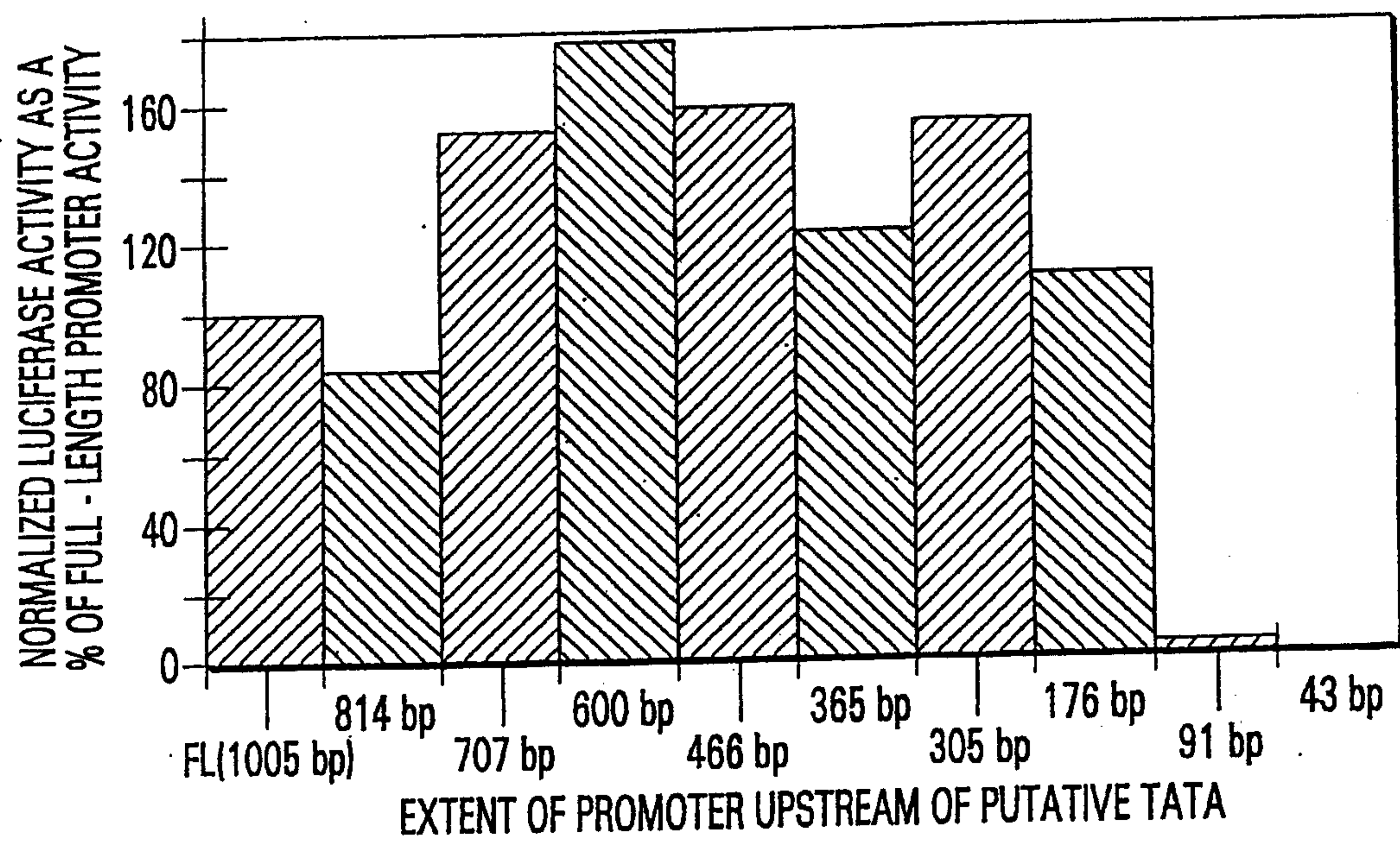


Fig.9

-180	CCCQATCTCA	TTTTCTTGGT	TGGCAGATCA	CAAAAAGGAA	CACAAAGGCT
	LS01	LS02	LS03	LS04	LS05
-130	AAGCCTCCTA	CTTGTTCTGGG	AGTTAGGTCA	GGGACACCAT	ATGAATGAAA
	LS06	LS07	LS08	LS09	LS10
-80	GAAATCTTAA	TTTGGGGTCA	CACCAAGATT	GTCTCTCTCG	AGGTTGGGGG
	LS11	LS12	LS13	LS14	LS15
-30	GTCCCTAAGG	TTGGTAGTAG	CAATACCCAA	TATATCACCT	AACAAACCCA
	LS16	LS17	LS18		
20	ATCCATGCTA	CATACATACA	TAGCATCCAT	CACCTGTAGA	CTGGACCCCTT
70	CATCAAGAGC	ACCATGG			

Fig.10

	=	DEL	-176/-92
	=	DEL	-89/-44
	=	DEL	-39/-8

19/34

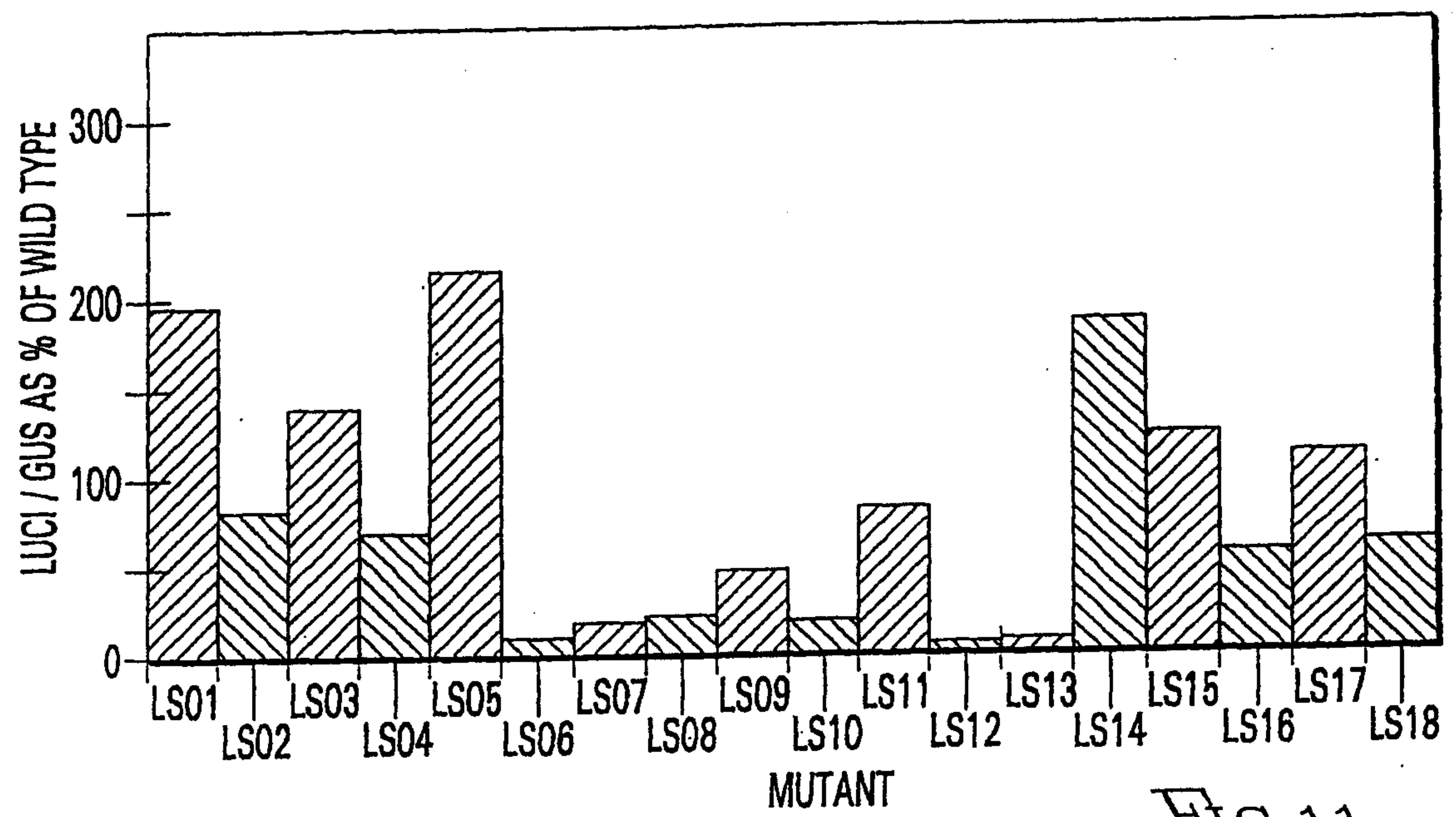


FIG. 11

20/34

```

1 .....GGAA 4
201 CCGGATGCACGACTGGCTTGTCTGGGTACCTGTCACGGCACAGGACAGTGA 250
5 TTCGGCTTATGCCGTTCACTTCCTACACCTACATCGCTGACCCGGTGAAT 54
251 CCGTCGACATGCCGTTCACTTCCTACACCTACATCGCTGACCCGGTGAAT 300
55 GTCGAGCATGTCCTCAAGACTAACTTCACCAATTACCCCAAGGGGGACGT 104
301 GTCGAGCATGTCCTCAAGACTAACTTCACCAATTACCCCAAGGGAATCGT 350
105 GTACAGATCCTACATGGATGTGCTCCTCGGTGACGGCATATTCAACGCTG 154
351 GTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCATCTTCAACGCCG 400
155 ACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCGCC 204
401 ACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCGCC 450
205 TCCAAGAACCTGAGGGATTTAGTGCCAATGTTTTAGAGAGTACTCCCT 254
451 TCCAAGAACCTGAGGGATTTAGCGCCATTGTGTTAGAGAGTACTCCCT 500
255 GAAGCTGTCGGGCATACTGAGTCAGGCATCCAAGGCAGGCAAAGTTGTTG 304
501 GAAGCTGTCGGGTATACTGAGCCAGGCATCCAAGGCAGGCAAAGTTGTTG 550
305 ACATGCAGGAACCTTTACATGAGGATGACACTGGACTCGATCTGCAANGTT 354
551 ACATGCAGGAACCTTTACATGAGGATGACGCTGGACTCCATCTGCAAGGTT 600
355 GGGTTCGGGGTCNANATCGGCACGCTGTCTNCCGGATCTCCCCGAGAACAG 404
601 GGGTTCGGGGTCGAGATCGGCACGCTGTGCGCCAGATCTCCCCGAGAACAG 650
405 CTTNCCCAAGCGTTTCGATGCCGCTAACATCATCGTCACNCTGCGGTTCA 454
651 CTTGCGCAGGCGTTTCGATGCCGCCAACATCATCATCAGCTGCGGTTCA 700
455 TCCACCCNCTGTGGCGCATCCAGAAGTTCTTCCCCNGTCA..... 494
701 TCGACCCGCTGTGGCGCATCAAGAGTTCTTCCACGTCGGGTCAGAGGCC 750

```

Percent Similarity: 92.510 Percent Identity: 90.891
 Sb200-Sorghr.Pep x. Sb20081.Pep February 13, 1997 11:29 ..

```

5 MPFTSYTYIADPVNVEHVLKTNFTNYPKGDVYRSYMDVLLGDGIFNADGE 54
87 MPFTSYTYIADPVNVEHVLKTNFTNYPKGIVYRSYMDVLLGDGIFNADGE 136
55 LWRKQRKTASF EFASKNLRDFSANVFREYSLKLSGILSQASKAGKVVD MQ 104
137 LWRKQRKTASF EFASKNLRDFS AIVFREYSLKLSGILSQASKAGKVVD MQ 186

```

FIG.12A

21/34

105 ELYMRMTLDSICXVGFGVXIGTLSPDLPENSEFXQAFDAANIIVTLRFIHP 154
|||||:|||||.|
187 ELYMRMTLDSICKVGFGVEIGTLSPDLPENSEFAQAFDAANIIITLRFIDP 236

155 LWRIQKFF 162
|||:|
237 LWRIKREF 244

Fig.12B

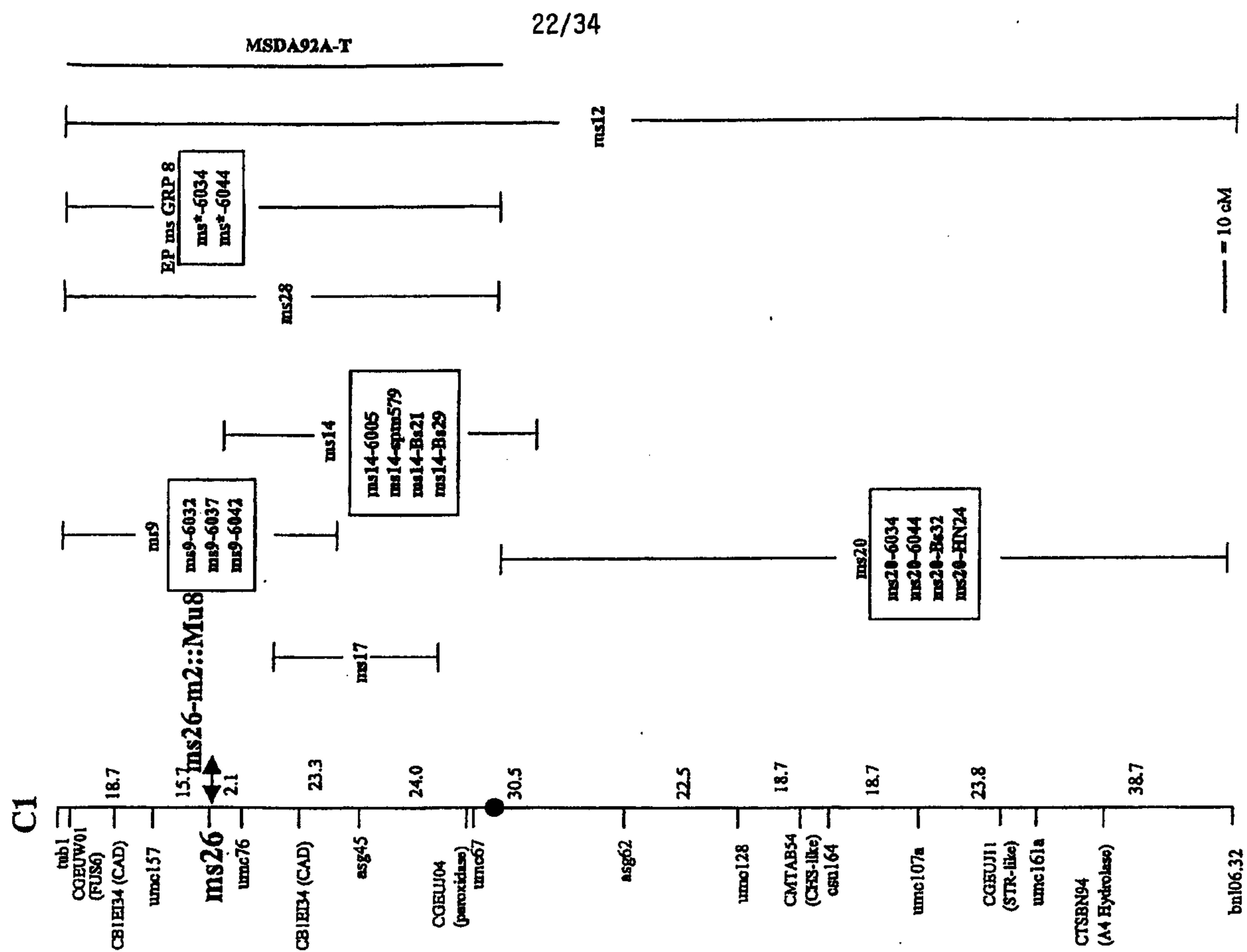


Figure 13.

23/34

Figure 14

Comparison of *ms26* excision and wild-type *Ms26*

```

excis: 798  caggaccccaaggggatcctggaggacgacgtgctgccggacgggacgaaggtgagggcc 857
            |||
Wtype: 1246 caggaccccaaggggatcctggaggacgacgtgctgccggacgggacgaaggtgagggcc 1305
            |||

excis: 858  ggcgggatggtgacgtacgtgccctactcgatggggcggatggagtacaactggggcccc 917
            |||
Wtype: 1306 ggcgggatggtgacgtacgtgccctactcgatggggcggatggagtacaactggggcccc 1365
            |||

                        NciI
                        -----

excis: 918  gacgcggcgagcttccggcgggaggacccggagcgggtggatcaacgaggatggcgcggttcc 977
            |||
Wtype: 1366 gacgcggcgagcttccggcggg-----agcgggtggatcaacgaggatggcgcggttcc 1417
            |||

excis: 978  gcaacgcgtcgccgttcaagttcacggcggttccaggcggggccgaggatctgcctgggca 1037
            |||
Wtype: 1418 gcaacgcgtcgccgttcaagttcacggcggttccaggcggggccgaggatctgcctgggca 1477
            |||

excis: 1038 aggactcggcgtagctgcagatgaagatggcgctggccatcctcttgcgcttctacagct 1097
            |||
Wtype: 1478 aggactcggcgtagctgcagatgaagatggcgctggccatcctcttcgcttctacagct 1537
            |||

excis: 1098 tccggctgctggagggggcaccgggtgcagtaccgcatgatgaccatcctctccatggcg 1157
            |||
Wtype: 1538 tccggctgctggagggggcaccgggtgcagtaccgcatgatgaccatcctctccatggcg 1597
            |||

```

24/34

Figure 15

ms26 transposon

CGGAGCTAGGGGTGAAAACGGGTAGGGTACCCGAAACGGGTACCGGATACGGATACTGAT
TCGGGACCATTTTTCGGATACGGATACGGGTATTTTTCAGATTTCGGGACGGATACGGGTA
ATACCCGGATAGTATGGCTTCGGATTCGGGTCGGATACGGAGCGAGTACTACCCGGTAAA
TACCCGGATACTCGGGTCGGATACCCGGTACCCGGAATTCGGGTACCCGTTTTTCTTTT
TCTGCAAAATAATATAGTTATAAAATCATAACTTTTACATATGAAATCGGATGAAGATAA
AGTTTATATGAAAATTGTAGAGCTCGAAGAGATCTATAACTTTGTAGTACATCACATTTT
TGTTTAAACATATCTTTAGGCCAAAATCATTAAAATAATGTCTAAATTTATATCAAATA
ATAGACTTTATCATTTTCATGTGGGGACTTAAGATTATATCCATGTGGGAACTTAGGATT
ATCTTTTATAAACTATTTATTAATATTGGTAACTTATTTGCAATTTTCGGTCGACGCTA
CAATATTTTATGAATTTAATTGTATTTTGATGATTTTCTACAACAAGAAATTAATAATA
CACCAAATAGCCTAAAAAATTCATGGATTTTACGGGGACACAACATATATCCACATATA
GTTCTCAAAAACATTTGGACTATAAAATCCACAAGATGTTGGTGTTTCTTCCATTCTACT
CCCACTTATTGCGTGAGTTACATGTGAAATCATTTTATGTATCGAAGTTTCAACATAATT
AATATTTCACTTATCATTTTCATGTGGCGACTTGAGGTTTTATTTGAATAGAATGTTTAT
TTGTTTTGGTAAGCTTTTGCATTTTGGATCAAACTAGTGTATTTATGAATTTTAATTAT
ACTTTGATGATTTTATGTAGAAAGAAATTAATAATGTATAAATAGCCTCAGAAATCTATG
AAATTATACGAAGGTACAACATATGGCCACATATAGTCATAACAAATAATGGGACCATAA
AATCCACAGGATGTCAACGTTTCTTCTATTTTATTTCCACTTATTGCGTGAGTTACACGT
GAAATCACTCTAAGTATCCAAGTTTCAACATAATCAATACTTCACTTTACCATTTTACG
TGGGAACTTGAGATTATCTTCTATTAAATGCTTATTAGTATTAATTTACTTGCAATTTTCG
TGGTCGAACAAGAATATTTTTTGATAACCAATTAATGCATTATCCGACAAGTATCCGATA
TCCGATCAAATAATATCCGTATCCGTCACTTATCCGCTCGGATAAATATCCGGTCCCTGT
ATCCGTATCCGTCCCGTTTCTAACTATCCGTATCCGATCCCGAATCCGTTTTAAATACAT
TAGGGTAGGATACAGGATGAGCTAATATCCGTCCGTATCCGCCCGTTTTACCCCTAGCC

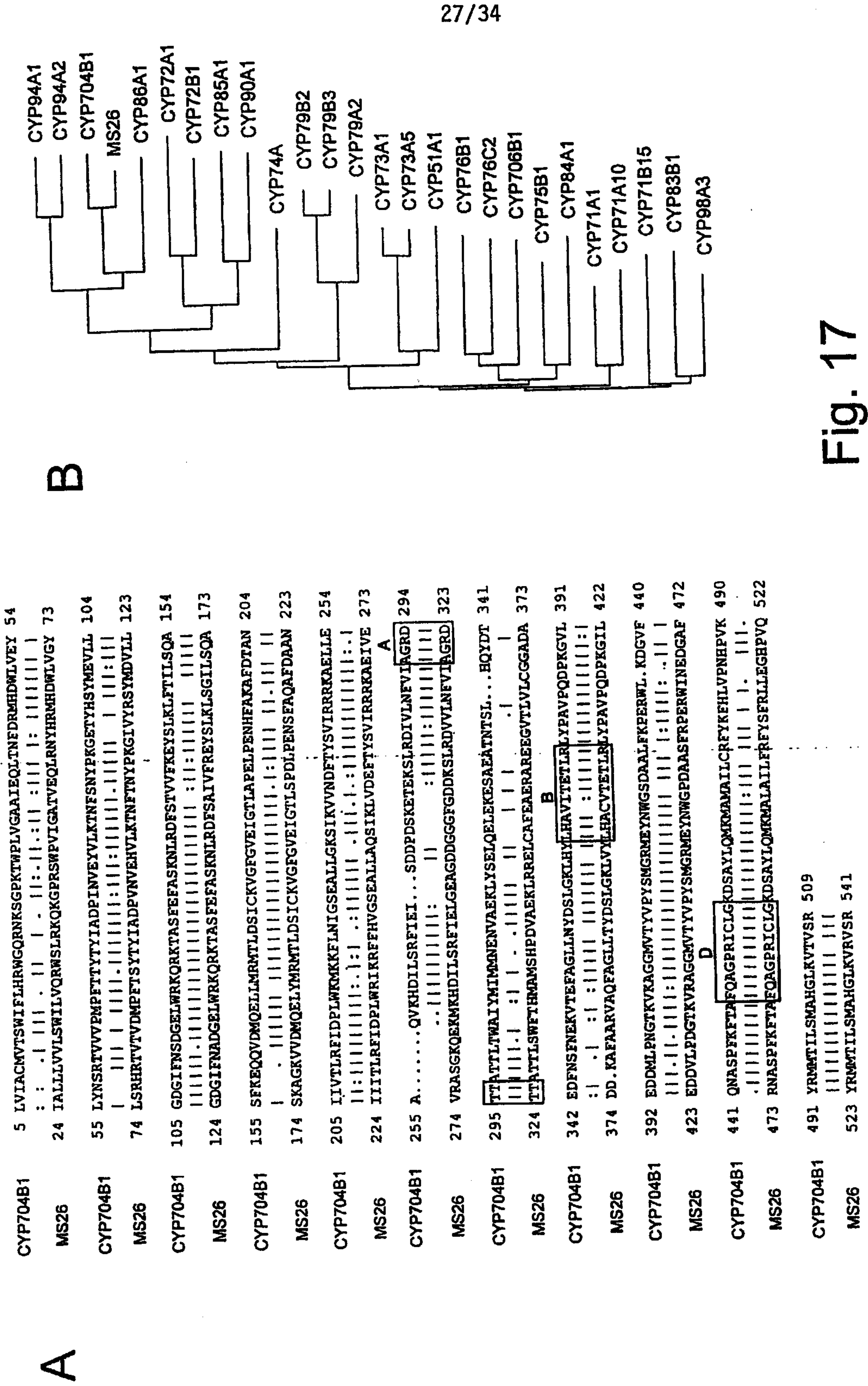
Figure 16A

ms26

AAC TGC ATG ACT TTT CACT TTT GGG TTC ACAA ATT GACT CACA AGAAA CAA ATT CACT TTT
TGG GTT CACA AAT TCCT CTC CAGG ATGT ACT TTT CACT TGAA CTGT CAT GTAT AGGA AC
AAG GAAT GGCT CAG TTT TTA AGGA ACA ATGT ACAG ATTT CAT TTT CAG AACT CTT TCT GGT
TGG TTG AG TTT CAG ACT TTT TGT ACCA AGCT GAT GGT CACA ATACT TGT TCC AAAG TC
TGATA ACAG AA CTGG CAA CTCTA ATTG ATAATA AAA AGAATA AAA ATACAGTATCAGAT
ATCT CAT TTT TCT TGG TTGGC AGATC AAAAA AGGAACACA AAGGCTAAGCCTCCTACTTG
TTCGGGAGTTAGGTCAGGGACACCATATGAATGAAAGAAATCTTAATTTGGGGTCACACC
AAGATTGTCTCTCTCGAGGTTGGGGGGTCCCTAAGGTTGGTAGTAGCAATACCCAATATA
TCACCTAACAAACCCAATCCATGCTACATACATACATAGCATCCATCACTTGTAGACTGG
ACCCTTCATCAAGAGCACCATGGAGGAAGCTCACATCACGCCGGCGACGCCATCGCCATT
CTTCCC ACTAGCAGGGCCTCACAAGTACATCGCGCTCCTCCTGGTTGTCTCTCATGGAT
CCTGGTCCAGAGGTGGAGCCTGAGGAAGCAGAAAGGCCCGAGATCATGGCCAGTCATCGG
TGCAACGGTGGAGCAGCTGAGGAAC TACCACGGATGCACGACTGGCTTGT CGGGTACCT
GTCGCGGCACAGGACAGTGACCGTCGACATGCCGTTCACTTCCTACACCTACATCGCTGA
CCCGGTGAATGTCGAGCATGTCTCAAGACTAACTTCACCAATTACCCAAGGTAAATGA
CCTGAACTCACTGATGTT CAGTCTTCGGAAATCAGAGCTGAAAGCTGAATCGAATGTGCC
TGAACACCGTGTAGGGAATCGTGTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCA
TCTTCAACGCCGACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCG
CCTCCAAGAACCTGAGGGATTT CAGCGCCATTGTGTT CAGAGAGTACTCCCTGAAGCTGT
CGGGTATACTGAGCCAGGCATCCAAGGCAGGCAAAGTTGTGGACATGCAGGTGAGATCAC
TGCTCCCTTGCCATTGCCAACATGAGCATTTC AACCTGAGACACGAGAGCTACCTTGCCG
ATTCAGGAAC TTTACATGAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTCCGGGTC
GAGATCGGCACGCTGTGCGCGGATCTCCCCGAGAACAGCTTCGCGCAGGCGTTTCGATGCC
GCCAACATCATCGTCACGCTGCGGTT CATCGACCCGCTGTGGCGCATCAAGAGGTTCTTC
CACGTCCGGT CAGAGGCCCTCCTAGCGCAGAGCATCAAGCTCGTGGACGAGTTCACCTAC
AGCGTGATCCGCCGGAGGAAGGCCGAGATCGTCGAGGTCCGGGCCAGCGGCAAACAGGAG
AAGGTACGTGTACATGACTGTTTCGATTCTTCAGTT CATCGTCTTGCCGGGATGGACCT
GATCCTGATTGATTATATATCCGTGTGACTTGTGAGGACAAATTAAAATGGGCAGATGAA
GCACGACATCCTGTACGGTTCATCGAGCTAGGCGAGGCCGCGACGACGGCGCGGCTT
CGGGGACGACAAGAGCCTCCGGGACGTGGTGCTCAACTTCGTGATCGCCGGGCGGGACAC
GACGGCGACGACGCTGTGCTGGTTCACGCACATGGCCATGTCCCACCCGGACGTGGCCGA
GAAGCTGCGCCGCGAGCTGTGCGCGTTTCGAGGCGGAGCGCGCGCGAGGAGGGCGTCGC
GCTCGTGCCCTGCGGCGGCGCTGACGCCGACGACAAGGCGTTTCGCCGCCCGCGTGGCGCA
GTTTCGCCGGGCTCCTCACCTACGACAGCCTCGGCAAGCTGGTCTACCTCCACGCCTGCGT
CACCGAGACGCTCCGCCTGTACCCCGCCGTCCCTCAGGTGAGCGCGCCCGACACGACCTC
CGGTCCGCGATGCAACGCATATGTGGCTGTCCGCAGAGCACAGCATGCAGTGAGTGGACC
TGAATGCACTATGCAATGCACTTGC GCGCGCGCAGGACCCCAAGGGGATCCTGGAGGACG
ACGTGCTGCCGGACGGGACGAAGGTGAGGGCCGGCGGGATGGTGACGTACGTGCCCTACT
CGATGGGGCGGATGGAGTACA ACTGGGGCCCCGACGCGCGAGCTTCCGGCCGGAGCTAG
GGGTGAAAACGGGTAGGGTACCCGAAACGGGTACCGGATACGGATACTGATTCCGGGACCA
TTTTTCGGATACGGATACGGGTATTTTTTAGATTCCGGGACGGATACGGGTAATACCCGGA
TAGTATGGCTTCGGATTCCGGTCCGATACGGAGCGAGTACTACCCGGTAATAACCCGGAT
ACTCGGGTCCGATACCCGGTACCCGGAATTCCGGTACCCGTTTTTTCTTTTTCTGCAAAA
TAATATAGTTATAAAATCATAACTTTTACATATGAAATCGGATGAAGATAAAGTTTATAT
GAAAATTGTAGAGCTCGAAGAGATCTATAACTTTGTAGTACATCACATTTTTGTTTAAAC
ATATCTTTAGGCCAAAATCATTAATAATGTCTAAATTTATATCAAAATAATAGACTTT
ATCATTTTCATGTGGGACTTAAGATTATATCCATGTGGGAACTTAGGATTATCTTTTAA
TAACTATTTATTAATATTGGTAAC TTTTGAATTTTCGGTCCGACGCTACAATATTTT
TATGAATTTAATTGATTTTGTATTTTCTACAACAAGAAATTAATAATACACCAAATA
GCCTAAAAAATTCATGGATTTTACGGGGACACAACATATATCCACATATAGTTCTCAA
AACATTTGGACTATAAAATCCACAAGATGTTGGTGGTTCTTCCATTCTACTCCCACTTAT
TGCGTGAGTTACATGTGAAATCATTTTATGTATCGAAGTTTCAACATAATTAATATTTCA

Figure 16B

CTTATCATTTCATGTGGCGACTTGAGGTTTTATTTGAATAGAATGTTTATTTGTTTTGG
TAAGCTTTTTGCATTTTGGATCAAAC TAGTGTATTTATGAATTTTAATTATACTTTGATG
ATTTTATGTAGAAAGAAATTAATAATGTATAAATAGCCTCAGAAATCTATGAAATTATAC
GAAGGTACAACATATGGCCACATATAGTCATAACAAATAATGGGACCATAAAATCCACAG
GATGTCAACGTTTCTTCTATTTTATTTCCACTTATTGCGTGAGTTACACGTGAAATCACT
CTAAGTATCCAAGTTTCAACATAATCAATACTTCACTTTACCATTTTACGTGGGAACTT
GAGATTATCTTCTATTAAATGCTTATTAGTATTAATTTACTTGCAATTTTCGTGGTCTGAAC
AAGAATATTTTTTGATAACCAATTAATGCATTATCCGACAAGTATCCGATATCCGATCAA
ATAATATCCGTATCCGTCACTTATCCGCTCGGATAAATATCCGGTCCCTGTATCCGTATC
CGTCCCGTTTCTAACTATCCGTATCCGATCCCGAATCCGTTTTAAATACATTAGGGTAGG
ATACAGGATGAGCTAATATCCGTCCGTATCCGCCCGTTTTACCCCTAGCCGGAGCGGTG
GATCAACGAGGATGGCGCGTTCCGCAACGCGTCGCCGTTCAAGTTCACGGCGTTCCAGGC
GGGGCCGAGGATCTGCCTGGGCAAGGACTCGGCGTACCTGCAGATGAAGATGGCGCTGGC
CATCCTTCTTGCGCTTCTACAGCTTCCGGCTGCTGGAGGGGCACCCGGTGCAGTACCGCA
TGATGACCATCCTCTCCATGGCGCACGGCCTCAAGGTCCGCGTCTCTAGGGCCGTCTGAT
GTCATGGCGATTTGGGATATCATCCCGCTTAATCCACGACAAATAACGTTTCGTGTTACAA
ATTTGCATGCATGCATGTAAGGGAAAGCGATGGGTTTCATTGGTGGCTTGGCTTAAGCCT
TAAAAACTCCGTCCGGTCTTGCGAACCACCACATCACTAGA



Rice Ms26

>GRMT00000036836
ATGAAGAGCCCATGGAGGAAGCTCATGCAATGCCAGTGACATCATTTCTCCAGTAGCAGGAATCCACAAGCTCATGCTATCTTCTTGTGCTCTCATGGATCTTG
GTCCACAAGTGGAGCCTCAGGAACCAAGCAAGGCCAAGATCATGGCCAATCATCGGCGGCACAGTGGAGCAACTGAAGAACTACCAACAGGATGCATGACTGGCTTGTCGA
GTACTTGTGCGAAGGACAGGACGGTGACCGTCGACATGCCCTTACACCTACACCTACATTCGCCGACCGGTGAACGTGCGAGCATGTCCTGAAGACCAACTTCACCAATTA
CCCCAAGGGTGAAGTGTACAGGTCTTACATGGATGTGCTCGGTGATGGCATATTCAATGCCGACGGCGAGATGTGGAGGAAGCAAGGAAGACGGCGAGCTTCGAGT
TTGCCCTCCAAAGAACTTGAGAGACTTCAGCACTGTGGTGTTCAGGGAGTACTCCCTGAAGCTATCAAGCATTTCTGAGCCAAAGCATGCAAGGCCGGCAGAGTTGTAGACATGC
AGGAATTGTTTATGAGGATGACACTGGACTCGATCTGCAAGGTGCGGTTTGGGTTGAGATCGGAGCGCTGTACCTGATCTCCGGAGAACAGCTTTGCCAGGCATTTC
GACGCTGCCAACAATCATCGTACGCTGCGGTTCATCGATCCTCTGTGGCTCTGAAGAAAGTTCTTGACAGTCCGCAAGCAAGCAAGATCAAGCACGACATACTGTGCGGTTTCAATCGA
GATGACTTCACCTACAGCTGATCCGCCCGCGCAAGCTTCGGGAGCACAAGAGCTCCGCGACGTGGTCTCAACTTCGTGATCGCCGGGCTGACACGACGCGGACG
GCTCGGGAGCGCGGCGACGAGGGGGCGGAGCTTCGCGGCGGCTCGCGGCGGCTCGAGGATGAGCGCGCGCGGCGGAGGCGGCTCG
ACGCTGTCGTGTTACGTACATGGCGATGACGACCCGCGGCTCGCGGCGGCGGTGCGGAGTTGCGTCTGAGTACGACGCGGTGGGAAGCTGGTGTACCTGCACGCGTGCGT
CGCTCGCGGACGCGCGGAGGCGTCTCGCGGCGGCGGTGCGGAGTTGCGTCTGAGTACGACGCGGTGGGAAGCTGGTGTACCTGCACGCGTGCGT
GACGGAGACGCTCCGCTTACCCGCGGTGCGCGAGGACCCCAAGGGATCGTGAGGACGACGTCTCCCGACGGCAAGGTGCGCGCGCGGCGGATGTTGACG
TACGTGCCCTACTCCATGGGAGGATGGAGTACAACTGGGCCCCGACGCGGAGCTTCGCGCGGAGCGGTGCTCAGGGCGACGGCGGCGGCTTCGCGGAACGCGT
CGCGCTTCAAGTTACCGCGTTCCAGCGCGCGGATCTGCTCGGCAAGGACTCCGCTACCTCCAGATGAAGATGGCGTCCGCTCCTCTCGCTTCGCTCTACACC
TTCGACCTCGTAGGACCAACCCCGTCAAGTACCGGATGATGACCATCCTCTCCATGGCTCAAGGTCCGCGTCTCCACCTCCGCTCTGA

>GRMT00000036836 Rice Ms26 protein

MKSPMEEAHAMPVTSFFPVAGIHKLIAIFLVLSMILVHKWSLRNQKPRSWPIIGATVE QLKNYHRMHHDWLVEYLSKDRTVTVDMPFTSYTYIADPVNVEHVLKTNFTNYPKGEVYRSY
MDVLLGDGIFNADGEMWRKQRKTASFEFASKNLRFSTVWFREYSLKLSILSQACKAGR VVDMQELFMRMTLDSICKVGFGEIGTLPDLPENSFAQAFAAANIIVTLRFIDPLWRLK
KFLHVGSEALLEQSMKLVDDFTYSVIRRRKAEILOARASGKQEKIKHDILSRFIELGEAG GDEGGSGFDDKSLRDVVLNFIAGRDTTATTLWSFTYMAAMTHPAVADKLRRELAAFEDE
RAREEGVALADAAGEASFARVAQFASLLSYDAVGKLVYLHACVTETLRLYPVPQDPKG
IVEDDVLPDGTKVRAGGMVTYVPYSMGRMEYNWGPDAASFPRPERWLSGDGGAFRNASPFK
FTAFAQGPRICLGKDSAYLQMKMALAILFRFYTFDLVEDHPVKYRMMTILSMAHGLKVRV STSV

Figure 19

Maize, rice and sorghum MS26 promoter alignment

gi_14030555_gb_AF366296.1_AF366296	(650)	651	700
sorghum ms26 promoter	(114)	AACGAATGATCA--TTGTCCTAAATTTTAAAGAAATTTTCAATTTCAAGAAC	
riceMS26promoter	(1)	AAGCCTGGTTTCAG----TTGGTGACAATTTACAGAAATTCAGATG-GATA	
Consensus	(651)	AAGGAATGT TCAG TTTTG GAAATTTTACAGAAATCA T GAAC	701
gi_14030555_gb_AF366296.1_AF366296	(700)	750	
sorghum ms26 promoter	(157)	AATTTCTGTAGGCTGAGTTT-CAGACTTTTGTACCAAGCTGATGGATC	
riceMS26promoter	(47)	TGGTTCGATATTAGAGGTGGCATACCTTTAGTCGCT-GCAAGGCTTC	
Consensus	(701)	T TTTCTGCTAGG TGAGTTT CAGACTTTTAGTACCAAGCTGATGGATC	751
gi_14030555_gb_AF366296.1_AF366296	(749)	800	
sorghum ms26 promoter	(206)	ACATT--CTGGATCCGAAGTATGATACATATATCTGGCACTCTCTAATTTG	
riceMS26promoter	(96)	AG-TTATCTGAA--CAAA-----ACAC-GNACTTGGCTGAGC--AGGGG	
Consensus	(751)	ACATTA CTG ATCCAAAGT TGATAACAGAA CTGGCAACTCTCTAATTTG	801
gi_14030555_gb_AF366296.1_AF366296	(799)	850	
sorghum ms26 promoter	(254)	-TAATAACATGATTAACCTGCAATACAGTATTAAGAGTGGC--TCATTT	
riceMS26promoter	(135)	AAAAAATACTGTAGCATTCATTTTGTGTTTACATGAGTAACGATTCCTTT	
Consensus	(801)	ATAATAA ATGTAATAA ANTACAGTATTAAGATGA C TCATTT	851
gi_14030555_gb_AF366296.1_AF366296	(839)	900	
sorghum ms26 promoter	(301)	TCTTGGTTGGCAGATCACAAAAGGAAAGACACAAAGGCTAAGCCTCTACTT	
riceMS26promoter	(185)	TCTAGGTGGACAGATCACAAAAG-AAAATAAGCTAAGAT-CCAACTC	
Consensus	(851)	TCTTGGTTGGCAGATCACAAAAGGAAAGACACAAAGGCTAAGC CCAACTT	901
gi_14030555_gb_AF366296.1_AF366296	(889)	950	
sorghum ms26 promoter	(350)	GTCCGGGAGTTAGGTGATGATACCATTAATGATGAAGAAATCTTAATTT	
riceMS26promoter	(233)	CTAAGGTTGTTAGGTAGGGACACCATTAATGAATGAGACAA-TCTTAATTC	
Consensus	(901)	GT CGGGAGTTAGGTGAGGACACCATTAATGAATGAAGAAATCTTAATTT	951
gi_14030555_gb_AF366296.1_AF366296	(939)	1000	
sorghum ms26 promoter	(400)	CCGCTCACACCAAGATTTGCTCTCTC-----AAGGTTG	
riceMS26promoter	(282)	TTGCTCACACAAGATTGCTCTC-----AAGGTTG	
Consensus	(951)	GGTCACACCAAGATTGCTCTCTC AAGGTTG	1001
gi_14030555_gb_AF366296.1_AF366296	(989)	1050	
sorghum ms26 promoter	(433)	GTACAGCAATACCAATATATACCTAACAAACCCAGACACACTACAT	
riceMS26promoter	(311)	GTAGCATCAGTGCCCAATATATACCTNACTATGCCA-TCCAAAATGC-T	
Consensus	(1001)	GTAGCAGCAATACCAATATATATACCTNACTATGCCA TCCA ACTACAT	1051
gi_14030555_gb_AF366296.1_AF366296	(1039)	1100	
sorghum ms26 promoter	(483)	ACATACATGCAATCCATCATTGTAGCTGGACCCCTTCATCAAGAGCAAC	
riceMS26promoter	(359)	ACATA-----G---CACTCTTTAGACTGAGCCCTTCATGAAGAGCCCC	
Consensus	(1051)	ACATA CATTCCATCATTGTAGACTGGACCCCTTCATCAAGAGCAAC	1101
gi_14030555_gb_AF366296.1_AF366296	(1089)	1150	
sorghum ms26 promoter	(529)	ATGGAGGAGCTCACTCATG-----	
riceMS26promoter	(401)	ATGGAGGAGCTCATGCAATGCCAGTGACATCTTCTTCCAGTAGCAGG	
Consensus	(1101)	ATGGAGGAGCTCA ATG	

Figure 20

Maize, rice and
sorghum MS26
alignment

Figure 21

1			50
	MS26	(1)	--MEEHLPATPSPFFPLAGPHKYTALLLVLSWILVQRWSLRKQKQPR
	RICE-MS26	(1)	MSSPGCEAHAMPVTSFFPVAGIHKLTATFLVLVLSWILVHKNSLRNQKQPR
sorghum	contig-made-by-aligning-top7hits	(1)	-----MPATPLPFLAGLHKYTAILLVLSWALVHRWSLRKQKQPR
	Consensus	(1)	MP TPEFPLAGLHKYTAILLVLSWILVHRWSLRKQKQPR
	MS26	51	100
	RICE-MS26	(49)	SWPVIGATVQQLRNYHRMHDWLVGYLSRHRTVTVDPFTSYTYTADPVNV
	RICE-MS26	(51)	SWPIIGATVQQLRNYHRMHDWLVGYLSKORTVTVDPFTSYTYTADPVNV
sorghum	contig-made-by-aligning-top7hits	(41)	SWPVIGATVQQLRNYHRMHDWLVGYLSRHRTVTVDPFTSYTYTADPVNV
	Consensus	(51)	SWPVIGATVQQLRNYHRMHDWLVGYLSRHRTVTVDPFTSYTYTADPVNV
	MS26	101	150
	RICE-MS26	(99)	EHVLKTNFTNYPKGIYRSYNDVLLGDGIFNADGELWRKQRKTASFEFAS
	RICE-MS26	(101)	EHVLKTNFTNYPKGEVYRSYNDVLLGDGIFNADGENWRKQRKTASFEFAS
sorghum	contig-made-by-aligning-top7hits	(91)	EHVLKTNFTNYPKGDVYRSYNDVLLGDGIFNADGELWRKQRKTASFEFAS
	Consensus	(101)	EHVLKTNFTNYPKGDVYRSYNDVLLGDGIFNADGELWRKQRKTASFEFAS
	MS26	151	200
	RICE-MS26	(149)	KNLRDPSAIVFREYSKLGLSILSOASKAGKVVDMQELYNRMNTLDSICKVG
	RICE-MS26	(151)	KNLRDFTVTVFREYSKLGLSILSOACKAGKVVDMQELFMNTLDSICKVG
sorghum	contig-made-by-aligning-top7hits	(141)	KNLRDPSANVTVFREYSKLGLSILSOASKAGKVVDMQELYNRMNTLDSICKVG
	Consensus	(151)	KNLRDPSAIVFREYSKLGLSILSOASKAGKVVDMQELYNRMNTLDSICKVG
	MS26	201	250
	RICE-MS26	(199)	FGVEIGTILSPDLFENSFAQADDAANIITILRFIDPLWRKRFFHVGSEAL
	RICE-MS26	(201)	FGVEIGTILSPDLFENSFAQADDAANIIVTLRFIDPLWRKKKTLFVGSEAL
sorghum	contig-made-by-aligning-top7hits	(191)	FGVEIGTILSPDLFENSFAQADDAANIIVTLRFIDPLWRKRFFHVGSEAL
	Consensus	(201)	FGVEIGTILSPDLFENSFAQADDAANIIVTLRFIDPLWRKRFFHVGSEAL
	MS26	251	300
	RICE-MS26	(249)	LAQSIKLVDDEFTYSVIRRRKAEIWEVRASGKQEKMKHDILSRFIELGEAG
	RICE-MS26	(251)	LEQSMKLVDDFTYSVIRRRKAEILOARASGKQEKIKHDILSRFIELGEAG
sorghum	contig-made-by-aligning-top7hits	(241)	LAQSIKLVDDEFTYSVIRRRKAEIWEVRASGKQEKMKHDILSRFIELGEAG
	Consensus	(251)	LAQSIKLVDDEFTYSVIRRRKAEIWEVRASGKQEKMKHDILSRFIELGEAG
	MS26	301	350
	RICE-MS26	(299)	DDGGG--FGDDKSLRDVVLNFVYIAGRDTTATTLTSLWFTTHQMSHPDVAEKL
	RICE-MS26	(301)	GDEGGGFGDDKSLRDVVLNFVYIAGRDTTATTLTSLWFTYNDANTHPAVADKL
sorghum	contig-made-by-aligning-top7hits	(291)	DDGGF---GDDKSLRDVVLNFVYIAGRDTTATTLTSLWFTTHQMSHPDVAEKL
	Consensus	(301)	DDGGG FGDDKSLRDVVLNFVYIAGRDTTATTLTSLWFTTHQMSHPDVAEKL
	MS26	351	400
	RICE-MS26	(347)	RRELCAFEAEARAREEGVTIVLCGGADADDKAFARVAQFAGLLTYDSLGK
	RICE-MS26	(351)	RRELANFEDERAREEGVALADNA----GEASEFAARVAQFASLLSYDAVGK
sorghum	contig-made-by-aligning-top7hits	(338)	RRELCAFEAEARAREEGVAVPCCG--PDDKAFARVAQFAGLLTYDSLGK
	Consensus	(351)	RRELCAFEAEARAREEGVAL CG DOKAFARVAQFAGLLTYDSLGK
	MS26	401	450
	RICE-MS26	(397)	LVYLHACVTETLRLYPVPQDPKGILEDVLPDCTKVRAGGMVTVVPYSM
	RICE-MS26	(397)	LVYLHACVTETLRLYPVPQDPKGIVEDDVLPDCTKVRAGGMVTVVPYSM
sorghum	contig-made-by-aligning-top7hits	(386)	LVYLHACVTETLRLYPVPQDPKGILEDVLPDCTKVRAGGMVTVVPYSM
	Consensus	(401)	LVYLHACVTETLRLYPVPQDPKGILEDVLPDCTKVRAGGMVTVVPYSM
	MS26	451	500
	RICE-MS26	(447)	GRMEYNWGPDAASFRPERWINEG-AFRNASPFKFTAFQAGPRICLGKDS
	RICE-MS26	(447)	GRMEYNWGPDAASFRPERWLSGCGGAFRNASPFKFTAFQAGPRICLGKDS
sorghum	contig-made-by-aligning-top7hits	(436)	GRMEYNWGPDAASFRPERWINEG-AFRNASPFKFTAFQAGPRICLGKDS
	Consensus	(451)	GRMEYNWGPDAASFRPERWINEG AFRNASPFKFTAFQAGPRICLGKDS
	MS26	501	548
	RICE-MS26	(496)	ATLQMKMALAILFRFYSFRLLEGHPVQYRNMTILSNRHGLKVRVSRVAV
	RICE-MS26	(497)	ATLQMKMALAILFRFTTDLVEDRFPVKYRNMTILSNRHGLKVRVSTSV
sorghum	contig-made-by-aligning-top7hits	(485)	ATLQMKMALAILFRFYSFOLLLEGHPVQYRNMTILSNRHGLKVRVSRVAV
	Consensus	(501)	ATLQMKMALAILFRFYSF LLEGHPVQYRNMTILSNRHGLKVRVSRVAV

Figure 22

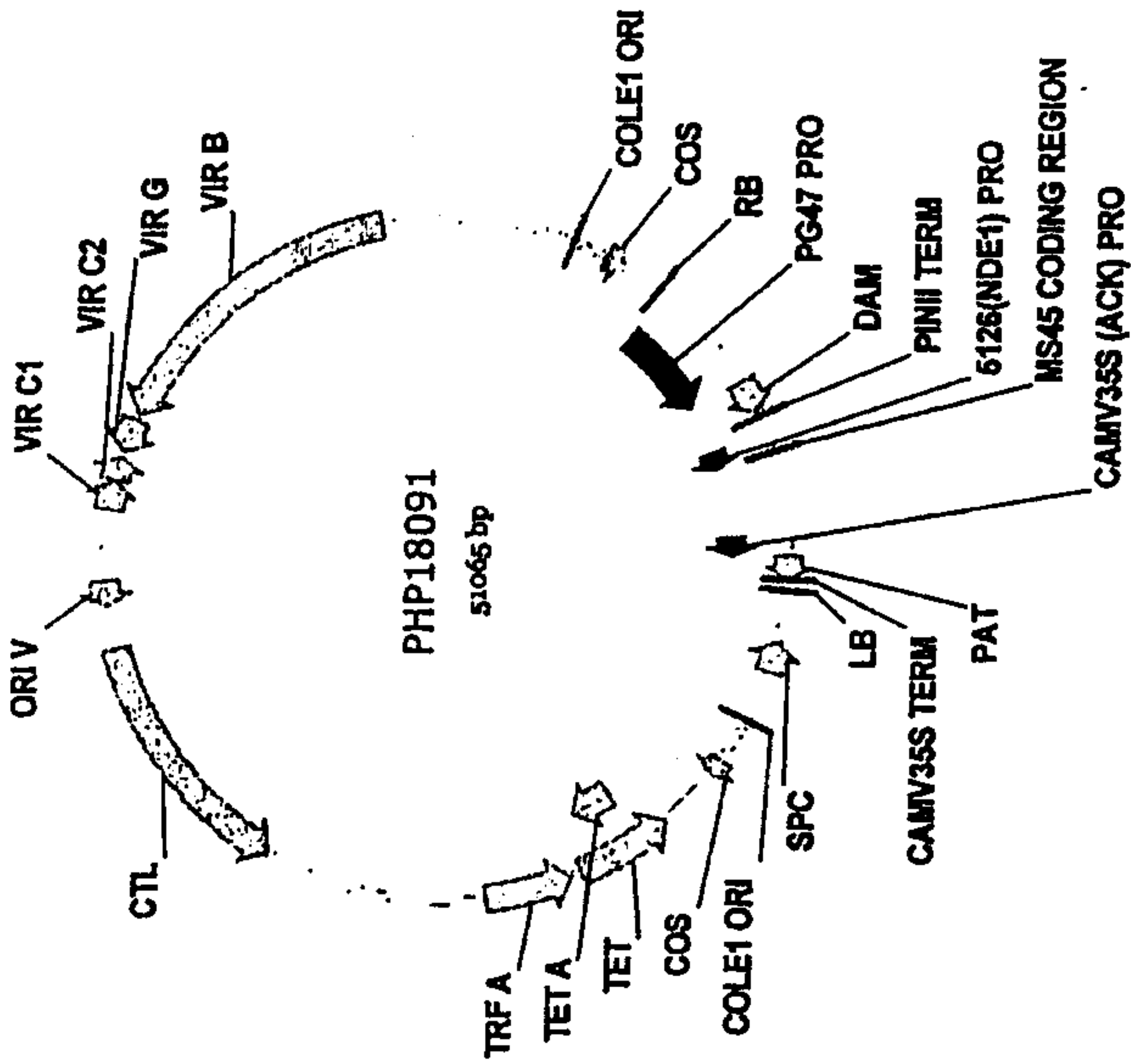
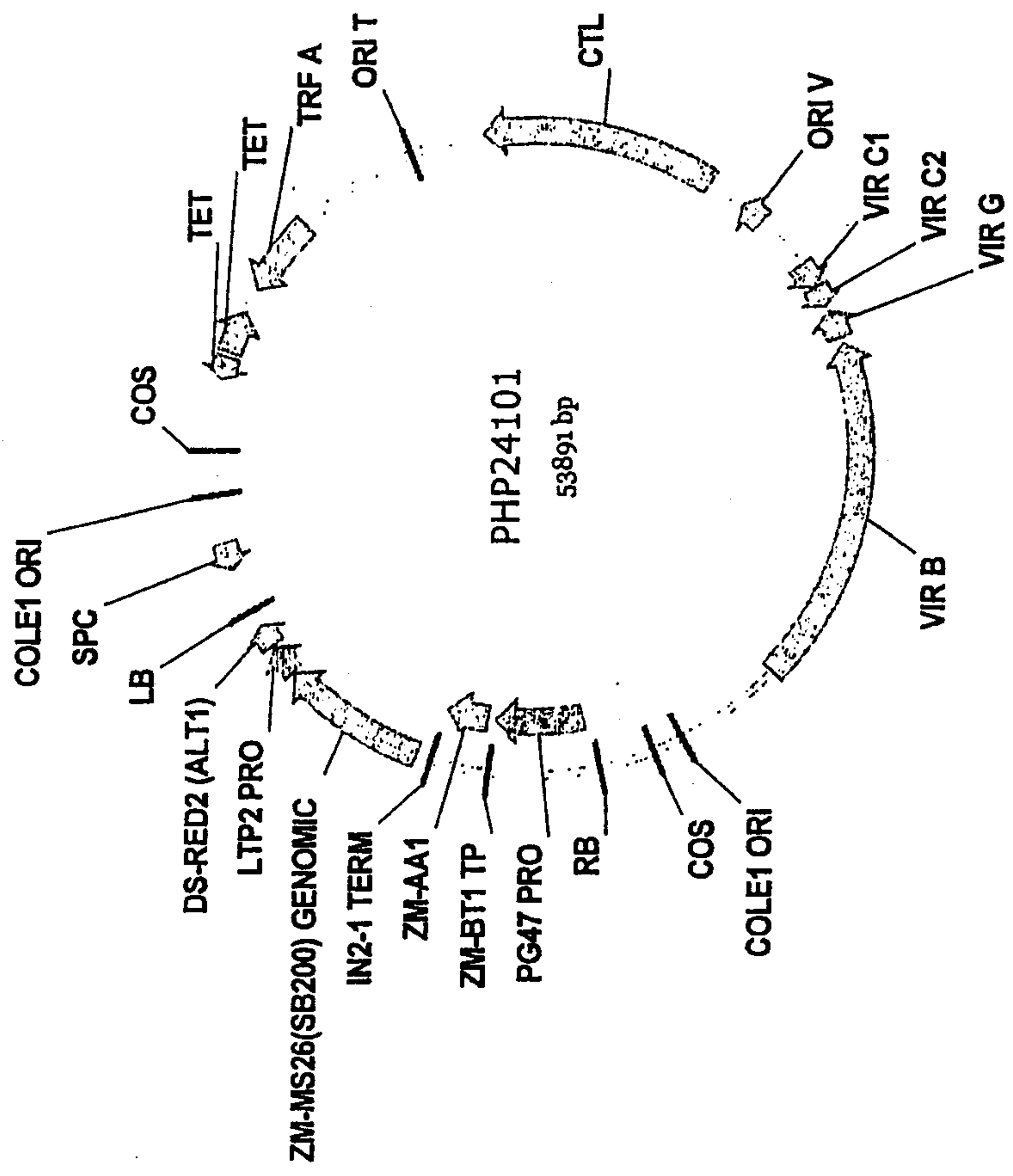


Figure 23



34/34

Figure 24

1 GGCACGAGCC GCGAGGCCA CTCGGCAGTC GGCACAAACCA CACACACCTC CACCCACTCT CTGAGATAAG TGAAGCATCT CGCGCACTGT CGCAGTCGCA
 101 M M K H S S S L C L L F L L A L C T T L L L A C G L V Q A Q V L
 GACGAGATG ATGAAGCACT CGAGCAGCTT GTGCTTGCTC TTCCTCTTGG CGCTCTGCAC CACCCTGTG GCTGGGCGC TGGTCCAGGC ACAAGTCCTC
 201 F Q G F N W E S C K Q Q G G W Y N R L K A Q V D I A K A G V T H V.
 TTCCAGGGGT TTAAGTGGG GTCGTGCAAG CAGCAGGGAG GCTGGTACAA CAGGCTCAAG GCCCAGGTG ACGACATCGC CAGGCGCGGC GTCACGCAAG
 301 . W L P P S H S V S P Q G Y M P G R L Y D L D A S K Y G T A A E L.
 TCTGGCTGCC TCCACCCCTG CACTCCGTCT CGCCACAAGG CTACATGCCA GGCCGCTAT ACACCTGGA CGCTCCAAAG TACGGCACGG CGGCGGAGCT
 401 . K S L I A A F H G R G V Q C V A D I V I N H R C A E K K D A R G V
 CAAGTCCCTG ATAGCGCGT TCCACGGCAG GGGCGTGCAG TCGTGGCGG ACATCGTCAT CAACCAACCGG TGGCGGAAA AGAAGGACGC GCGCGGCGTG
 501 Y C I F E G G T P D D R L D W G P G M I C S D D T Q Y S D G T G H R.
 TACTGCATCT TCGAGGGCGG GACTCCCGAC GACCCCTGG ACTGGGGCC CGGATGATC TGCAGCGAG ACACGCAGTA CTCGGACGGG ACGGGGCACC
 601 . D T G E G F A A A P D I D H L N P R V Q R E L S A W L N W L R S D.
 GCGACACGGG CGAGGGGTTG GCGGCGGCG CCGACATCGA CCACCTCAAC CGCGCGTGC AGCGGGAGCT CTCGGCTTGG CTCAACTGGC TCAGGTCCGA
 701 . A V G F D G W R L D F A K G Y S P A V A R M Y V E S T G P P S F V
 CGCGGTGGG TTCGACGGT GCGGCTCGA CTTGCGCAAG GGCTACTGC CGCGCGTGC CAGATGTAC GTGGAGAGCA CCGGGCGGCC GAGCTTCGTC
 801 V A E I W N S L S Y S G D G K P A P N Q D Q C R Q E L L D W T R A V.
 GTCGCGGAGA TATGGAATC GCTGAGCTAC AGCGGGGACG GCAAGCGGC GCCCAACCAAG GACCACTGCC GCGAGGAGCT GCTGGACTGG ACGCGGGCGG
 901 . G G P A M A F D F P T K G L L Q A G V Q G E L W R L R D S S G N A.
 TCGGCGGCGC CGCCATGGG TTCGACTTCC CCACCAAGG CTTGCTGCAG GCGGGGTGC AGGGGAGCT GTGGCGGTG CCGACAGCT CCGGCAACGC
 1001 . A G L I G W A P E K A V T F V D N H D T G S T Q K L W P F P S D K
 GGCGGCGCTG ATCGGGTGG CGCCCGAGAA GGCGTCCACC TTCGTGACA ACCATGACAC CGGTCGACG CAGAAGCTCT GGCGGTCCC ATCCGACAAG
 1101 V M Q G Y A Y I L T H P G V P C I F Y D H M F D W N L K Q E I S T L.
 GTCATGCAGG GCTACGCCA CATCCTCACC CATCCAGGAG TCCCTGCTAT TTCTACGAC CACATGTTG ACTGGAACCT GAAGCAGGAG ATATCCACGC
 1201 . S A I R A R N G I R A G S K L R I L V A D A D A Y V A V V D E K V.
 TGTCTGCCAT CAGGCGCGG AACGGCATCC GCGCGGGAG CAAGCTGCGG ATCTCTGTCG CCGTACGTG GCCGTGTCG ACGAGAGGT
 1301 . M V K I G T R Y G V S S V V P S D F H P A A H G K D Y C V W E K A
 CATGGTGAAG ATCGGGACAA GGTACGGCGT GAGCAGCGTG GTCCCGTCGG ATTTCCACCC GGCGGCGCAC GGCAAGGACT ACTGCGTCTG GGAGAAAGCG
 1401 S L R V P A G R H L
 AGCCTCCGG TCCCGGCGG GCGCCACCTC TAGCAGCTCA GATTGCTCAG TCTTGTGCTG CATTGCAAC ACAGCAGCAC GACACTGCAT AACGTCTTTT
 1501 CCTTAATTTC CTGAATTTTA CCTTTTCTTA GTTCAATTTC ATATATGTAT TTCTACATGT ACACACTATC ACAATCAGAT AATAAACAA GCTTGGTCAA
 1601 AAAAAAAAA AAAAAA