

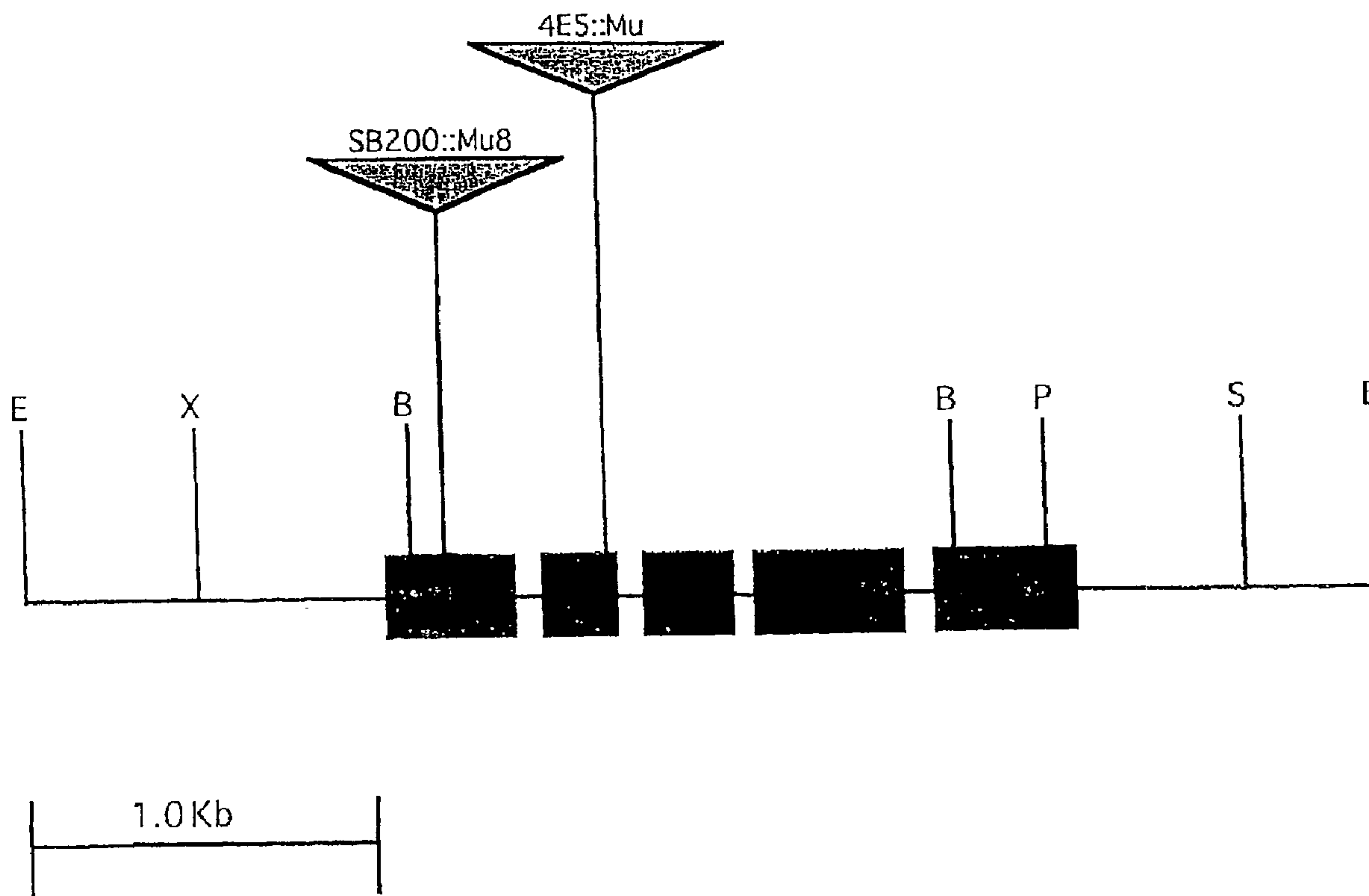


(22) Date de dépôt/Filing Date: 2001/09/25  
 (41) Mise à la disp. pub./Open to Public Insp.: 2002/04/04  
 (62) Demande originale/Original Application: 2 423 480  
 (30) Priorité/Priority: 2000/09/26 (US09/670,153)

(51) Cl.Int./Int.Cl. *C12N 15/113* (2010.01),  
*A01H 1/02* (2006.01), *A01H 5/10* (2006.01),  
*C12N 15/00* (2006.01), *C12N 15/82* (2006.01),  
*C07K 14/415* (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, US;  
 HUFFMAN, GARY, US;  
 TRIMNELL, MARY, US  
 (74) Agent: TORYS LLP

(54) Titre : SEQUENCES NUCLEOTIDIQUES MEDIATRICES DE LA FERTILITE MALE ET PROCEDE D'UTILISATION ASSOCIE

(54) Title: NUCLEOTIDE SEQUENCES MEDIATING 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.

**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.

## **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 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  
10 with variations in plant composition are 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 transferred to the same  
15 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 reproductive organs of a  
20 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 same plant. It can self or cross pollinate. Natural pollination occurs in maize when wind blows pollen from the tassels to  
25 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 hybrids, which relies upon some sort of male sterility system and where a female sterility system would reduce production costs.

30 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 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  
35 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 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<sub>1</sub>. In the development of hybrids only the F<sub>1</sub> hybrid plants are sought. The F<sub>1</sub> hybrid is more vigorous than its inbred parents. This 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 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<sub>1</sub>) and will form hybrid plants.

Environmental variation in plant development can result in plants tasseling 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 seed is not as productive as F<sub>1</sub> 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 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  
5 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.

10 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  
15 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  
20 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,  
25 published as WO 90/08829.

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  
30 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.



- 4 -

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

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.

#### 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.

- 4a -

An aspect of the invention is to provide an isolated nucleotide sequence comprising the SBMu200 gene, and a plant cell and plant transformed with the nucleotide sequence. The plant cell can be a maize cell and the plant can be maize.

Another aspect of the invention is to provide an isolated nucleotide sequence that  
5 mediates male fertility in plants comprising a nucleotide sequence encoding any of the amino acid sequences of SEQ ID Nos: 2, 4 or 8 and those sequences which hybridize to the nucleotide sequences encoding any of the amino acid sequences of SEQ ID Nos: 2, 4 or 8 under highly stringent conditions. Also provided, is a plant cell and plant transformed with the nucleotide sequence. The plant cell can be a maize cell and the  
10 plant can be maize.

Another aspect of the invention is to provide an isolated DNA molecule that mediates fertility in plants comprising a nucleotide sequence of any of SEQ ID Nos: 1, 3, or 7 and those sequences which hybridize to the nucleotide sequences of SEQ ID Nos: 1, 3, or 7 under highly stringent conditions. Also provided is a plant cell and plant transformed  
15 with the nucleotide sequence. The plant cell can be a maize cell and the plant can be maize.

Another aspect of the invention is to provide a method of impacting fertility of a plant comprising impacting the SBMu200 gene. The sequence expression can be repressed. The expression of the nucleotide sequence is repressed by mutation of the nucleotide  
20 sequence. The method can further comprise delivering into the plant a second nucleotide sequence which represses expression of the nucleotide sequence. The method can further comprise delivering into the plant a second nucleotide sequence molecule oriented in the antisense direction relative to the DNA molecule thereby repressing expression of the DNA molecule. A native SBMu200 gene in a plant can be silenced and a second  
25 SBMu200 gene linked to an inducible promoter can be introduced into the plant such that the plant is constitutively male sterile and fertility is induced by inducing the promoter. Also provided is a male fertility mediated plant produced according to this method.

Another aspect of the invention is to provide a method of impacting fertility of a plant  
30 comprising impacting a nucleotide sequence in the plant encoding the amino acid



- 4b -

sequence of any of SEQ ID Nos 2, 4 or 8 the nucleotide sequences of any of SEQ. ID Nos. 1, 3, or 7 and those sequences which hybridize to any of said sequences under highly stringent conditions.

Another aspect of the invention is to provide a method of producing hybrid seed,  
5 comprising: (a) planting in cross-pollinating juxtaposition, a first seed from a selected male fertile parent line and a second seed selected from a female parent line having male sterility produced according to the method described above; (b) growing the seed to mature plants under conditions which do not induce expression of the second DNA molecule; (c) cross-pollinating the male sterile female plant with pollen from the male  
10 fertile plant; and (d) harvesting seed from the male sterile female plant. The method can further comprise cross-fertilizing the male sterile plant with a second plant, the second plant comprising a second exogenous gene, the product of the second gene preventing disruption of the male tissue by the first exogenous gene, producing a male fertile hybrid plant. The gene impacting male fertility can be dominant and the method can further  
15 comprise growing the hybrid seed to produce a third male sterile parent plant; producing a fourth parent plant comprising one or more genes controlling a desired gene trait and cross-fertilizing the third and fourth parent plants to produce second hybrid seed.

Another aspect of the invention is to provide a method of providing heritable externally controllable male sterility in a plant comprising linking SBMu200 in an expression  
20 sequence with an inducible promoter responsive to external control; delivering the expression sequence into the genome of the plant; and inactivating a second DNA molecule which codes for the product of SBMu200 from the native genome of the plant. The nucleotide sequences can comprise the nucleotide sequences encoding the amino acids of SEQ ID Nos. 2, 4 or 8 or having the nucleotide sequence of any of SEQ. ID Nos.  
25 1, 3, or 7 and those nucleotide sequences which hybridize to any of said sequences. The method can further comprise planting seed of the plant to provide growing male sterile plants; inducing conversion of the growing plants to male fertile form under conditions which induce the promoter to express the first DNA molecule; and open-pollinating the growing plants in isolation to produce seed; and harvesting the seed. Also provided is a  
30 controllably male sterile plant produced according to this method.

- 4c -

Also provided is an expression vector comprising the DNA sequence of an SBMu200 gene. The expression vector can further comprise an exogenous gene, wherein the exogenous gene is operably linked to the promoter. The promoter can be selected from any one of CaMV35S, SGB6, SBMu200, MS45 and 5126. The product of the  
5 exogenous gene can disrupt the function of male tissue. Also provided are plant cells comprising the vector.

Another aspect of the invention is to provide a method of mediating male fertility in a plant comprising introducing into a plant the expression vector described above wherein the exogenous gene impacts male fertility of the plant and the promoter control  
10 expression of the exogenous gene. The regulatory element in conjunction with the promoter can be inducible.

Another aspect of the invention is to provide a nucleotide sequence as represented in ATCC deposit no. 98931.

Another aspect of the invention is to provide an isolated nucleic acid sequence encoding  
15 the promoter region of the gene SBMu200.

Another aspect of the invention is to provide an isolated nucleic acid sequence comprising SEQ ID NO. 5 or those nucleotide sequences which hybridize to SEQ ID NO. 5 under conditions of high stringency.

Another aspect of the invention is to provide a male tissue-preferred regulatory region  
20 comprising nucleotide sequences essential for initiating transcription of the SBMu200 gene. The regulatory region can comprise nucleotide sequences of about 130 contiguous base pairs from about -38 and higher upstream of the TATA box of the SBMu200 gene. Also provided is an expression vector comprising a promoter that is operably linked with the male tissue specific regulatory element. The expression vector can further comprise  
25 an exogenous gene, wherein the exogenous gene is operably linked to the promoter. The promoter can be selected from any one of CaMV35S, SGB6, SBMu200 MS45 and 5126. The product of the exogenous gene can disrupt the function of male tissue. Also provided are plant cells comprising the vector. Another aspect of the invention is to provide a method of mediating male fertility in a plant comprising introducing into a  
30 plant the expression vector wherein the exogenous gene impacts male fertility of the



- 4d -

plant and the regulatory element in conjunction with the promoter control expression of the exogenous gene. The exogenous gene can disrupt function of male tissue of the plant causing the plant to be male sterile. The regulatory element in conjunction with the promoter can be inducible. The plant can be constitutively sterile when the promoter and regulatory element are not induced and can be fertile when the promoter and regulatory element are induced. The method can further comprise cross-fertilizing the male sterile plant with a second plant, the second plant comprising a second exogenous gene, the product of the second gene preventing disruption of the male tissue by the first heterologous gene, producing a male fertile hybrid plant.

10

Another aspect of the invention is to provide a male tissue-preferred regulatory region comprising Seq. ID. NO. 6.

Another aspect of the invention is to provide an isolated nucleic acid that is a male tissue specific regulatory element comprising a fragment of the nucleotide sequence of SEQ ID No. 6 or those nucleotide sequences which hybridize to SEQ ID. NO. 6 under conditions of high stringency wherein the regulatory element is essential for initiating transcription of Seq ID Nos. 1, 3 or 7.

Another aspect of the invention is to provide a regulatory region comprising the sequences of -44 to -180 upstream of the TATA box of SEQ ID NO. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency. The regulatory region can comprise sequences -92 to -176 upstream of the TATA box of SEQ ID. No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency. The regulatory region can comprise sequences -44 to -89 upstream of the TATA box of SEQ. ID. No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency. The regulatory region can comprise sequences -52 to -131 upstream of the TATA box of SEQ ID No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency. The regulatory region can comprise sequences -52 to -71 upstream of the TATA box of SEQ ID No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency. The regulatory region



- 4e -

can comprise sequences -82 to -131 upstream of the TATA box of SEQ ID No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency.

Another aspect of the invention is to provide a method of producing hybrid seeds  
5 comprising: (a) producing a first parent plant comprising nucleotide sequences essential for initiating transcription of the SBMu200 gene operably linked with an exogenous gene impacting male fertility of the plant such that the plant is male sterile; (b) producing a second parent plant which is male fertile; (c) cross-fertilizing the first parent plant and the second parent plant to produce hybrid seeds. The gene impacting male fertility can  
10 be dominant and the method can further comprise growing the hybrid seed to produce a third male sterile parent plant; producing a fourth parent plant comprising one or more genes controlling a desired gene trait and cross-fertilizing the third and fourth parent plants to produce second hybrid seed.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

15 FIG. 1. is a locus map of the male sterility gene SBMu200, where E=EcoRI, X=XbaI, B=BamHI, P=PstI, and S=SacI.

FIG. 2. is a gel of a Southern Blot analysis of ECORI digested DNA from a Mu family segregating for male sterility and hybridized with a Mu8 probe.

FIG. 3. is a Northern Blot analysis gel hybridized with a PstI fragment isolated from the SBMu2003.1 clone.

FIG 4 is the sequence of SBMu200 (The cDNA is SEQ ID NO.1, the protein is SEQ ID NO.2).

5 FIG. 5 is the genomic SBMu200 sequence (also referred to as SEQ ID NO. 7).

FIG 6 is a comparison of the genomic SBMu200 sequence with the cDNA of SBMu200.

FIG. 7. is a Northern analysis gel showing developmental gene expression in microsporogenesis of the gene SBMu200.

FIG. 8 is the full length promoter of SBMu200 (SEQ ID No. 5)

10 FIG. 9. is a bar graph showing luciferase activity after deletions of select regions of the SbMu200 promoter.

FIG. 10 shows an essential region of the SBMu200 promoter (SEQ ID NO. 6).

Coordinates are relative to the putative TATA box (underlined). P motifs are in italic.

15 Linker scanning mutations that reduce activity to ~5% or less are in bold. Mutations with a significant but less pronounced effect are in bold italic.

FIG 11 is a bar graph showing luciferase activity after substitution by restriction site linker scanning of select small (9-10bp) regions of the SBMu200 essential promoter fragment.

FIG. 12 is a comparison of SBMu200 sorghum tassel and SBMu200 maize cDNA 8.1 (the sorghum DNA is SEQ No. 3 and the protein is SEQ ID NO. 4).

20

## 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  
25 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  
30 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. To date, published genetic male sterility research has been mostly descriptive. Some efforts have been made to establish the mechanism of sterility in maize, but few have been satisfactory. This should not be surprising given the number of genes that have been identified as being responsible for male sterility. One mechanism is unlikely to apply to all mutations.

At Patent No. 5,478,369 there is described a method by which a male sterility gene was tagged on maize chromosome 9. Previously, the only described male sterility gene on chromosome 9 was MS2, which has never been cloned and sequenced. 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*.

The SBMu200 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 SBMu200 and was found to have an approximately 5.5 Kb EcoRI fragment that hybridized with a Mu8 probe. 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. This probe was used to isolate cDNA clones from a tassel cDNA library. The cDNA for SBMu200 is 1906 bp, and the Mu insertion occurred in exon 1 of the gene. Figure 9 (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.

Plasmids containing the nucleotide sequences of the invention were deposited with the Patent Depository of the American Type Culture Collection (ATCC), Manassas, Virginia, received on October 16, 1998 and assigned ATCC Designation No. 98931. The deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. The deposit was made merely as a convenience for those of skill in the art and are not an admission that a deposit is required.

Further, 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 invention also includes those nucleotide sequences which selectively hybridize to the SBMu200 nucleotide sequences under stringent conditions. In referring to a sequence that "selectively hybridizes" with SBMu200, 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 (e.g., at least 2-fold over background) 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 sequence, to a detectably greater degree than to other sequences (e.g., at least 2-fold over background). 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 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, 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. (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% 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% sequence similarity.

Specificity is typically the function of post-hybridization washes, the critical 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 2X SSC, 0.5% (w/v) SDS, at 65°C for 30 minutes 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.

Methods of aligning sequences for comparison are well-known in the art. Gene comparisons can be determined by conducting BLAST ( Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) J. Mol. Biol. 215:403-410)

searches under default parameters for identity to sequences contained in the BLAST "GENEMBL" database. A sequence can be analyzed for identity to all publicly available DNA sequences contained in the GENEMBL database using the BLASTN algorithm under the default parameters. Identity to the sequence of the present invention would mean a 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% sequence identity, more preferably at least 90% sequence identity and most preferably at least 95% sequence identity.

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 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 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.

The isolated promoter sequence of the present invention can be modified 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 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 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 are known in the art and include the SV40 enhancer region, the 35S enhancer element, and the like:

Sequences which hybridize to the 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% 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 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. 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), 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 SBMU200 has been identified as including the 1005 bp 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-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 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 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, the promoter of the present invention can be linked with nucleotide sequences other than the SBMu200 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 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 exogenous gene of the expression vector with the SBMu200 promoter include complementary nucleotidic units such as antisense molecules (callase antisense RNA, barnase antisense RNA and chalcone synthase antisense RNA, Ms45 antisense RNA), ribozymes and external guide sequences, an aptamer or single stranded nucleotides. The exogenous nucleotide sequence can also encode auxins, rol B, cytotoxins, diphtheria toxin, DAM methylase, avidin, or may be selected from a prokaryotic regulatory system. By 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- $\beta$ -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 SBMu200 gene. A vector is prepared comprising the SBMu200 gene, 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. Selection of the promoter will depend upon the use intended of the gene. 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.

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, 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: picornavirus leaders, for example, 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, 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 under the control of the promoter. 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*).

The transformation vector comprising the promoter sequence of the present invention operably linked to a heterologous nucleotide sequence in an expression cassette, 5 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.

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 10 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 (1995); and Chiu et al. Current Biology 6:325-330 (1996).

Selectable marker genes for selection of transformed cells or tissues can be 15 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 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, 20 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. 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. 25 Science 233:478-481(1986); phosphinothricin, DeBlock et al. EMBO J. 6:2513-2518 (1987).

The method of transformation/transfection is not critical to the instant invention; various methods of transformation or transfection are currently available. As newer methods are available to transform crops or other host cells they may be directly applied. 30 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 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 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); 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 202:179-185 (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 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. 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 for transforming corn or rice. Sorghum transformation is described at Casas et al, *supra* and sorghum by Wan et al, Plant Physicol. 104:37 (1994). Soybean



transformation is described in a number of publications, including U.S. Patent no. 5,015,580.

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

5

### Example 1

#### Identification and Cosegregation of SBMu200

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  
10 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 its expression. Plants from a segregating F<sub>2</sub> family in which the male sterile mutation was designated SBMu200, were grown and classified for male fertility/sterility based on the above criteria. Leaf samples were taken and DNA  
15 subsequently isolated on approximately 20 plants per phenotypic classification, that is male fertility vs. male sterility.

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,  
20 fractioning the cut DNA by molecular weight on an agarose gel, and transferring to nylon membranes to fix the separated DNA. These membranes are subsequently hybridized with a probe fragment that was radioactively labeled with P<sup>32</sup>P-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 F<sub>2</sub> SBMu200  
25 family were grown and classified for male fertility/sterility. Leaf 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  
30 from the Mu8 transposon. Autoradiography of the membrane revealed cosegregation of a 5.5 Kb EcoRI fragment with the sterility phenotype as shown in Figure 2. This EcoRI band segregated in the fertile plants suggesting a heterozygous wild type condition for the allele.

- 17 -

Example 2Library Construction and Screening

The process of cDNA 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 EcoRI 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 PstI border fragment from this clone was isolated and used to reprobe the original EcoRI cosegregation blot. The 5.5 Kb EcoRI 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 EcoRI band and a 4.3 Kb EcoRI band. Two of the fertile plants are homozygous for the 4.3 Kb EcoRI band, presumably the wild type allele.

Example 3Expression 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 PstI 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 SBMu200-8.1, was isolated from the library.

5

#### Example 4

##### Sequence Analysis

The SBMu200-3.1 genomic clone and the cDNA clone, SBMu200-8.1, 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 74:5463-10 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 15 genomic clone. The cDNA homology with respect to the genomic clone begins at nucleotide 1094. Thus SBMu200-8.1 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. A comparison of 20 the predicted protein from SBMu200-8.1 to the consensus amino acid sequence of these domains showed 1) 92% identity to the dioxygen binding domain, 2) 85% identity to the tridecapeptide domain (steroid binding), and 3) 100% identity to the C-terminal heme attachment domain. Further expression studies were done using the SBMu200-8.1 cDNA probe against a northern containing mRNA at discrete stages of microsporogenesis. Signal 25 is detected from meiosis II/quartet to late-uninucleate, with maximal signal being observed from early-uninucleate through late-uninucleate as shown at Figure 7.

#### Example 5

##### Identification of Promoter and its Essential Regions

30

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).

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  
5 “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 35S promoters have been identified in relatively small fragments derived  
10 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 the art. For eukaryotes, the regions in the vector include regions that control initiation of transcription and control  
15 processing. These regions are operably linked to a reporter gene such as UidA, encoding  $\beta$ -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” 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  
20 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 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).

25 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 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  
30 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 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  $\mu$ m 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., *supra*).

Upstream of the likely translational start codon of SBMu200, 1088 bp of DNA was present in the genomic clone SBMu200-3.1. Translational fusions via an engineered *NcoI* site were generated with reporter genes encoding luciferase and  $\beta$ -glucuronidase to test whether this fragment of DNA had promoter activity in 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 SBMu200-3.1 thus 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 SBMu200 mRNA a 5'-nontranslated leader of about 48-58 nt. The total SBMu200 subgenomic fragment of 1088 bp, including nontranslated leader, start of transcription, TATA box and sequences upstream of the TATA box, was 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.

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' 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 SBMu200 plus a core promoter. The core promoter can be any one of known core



- 21 -

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 SBMu200 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 SBMu200 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



- 22 -

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 TAATA box, particularly the -52 to -71 region.

#### Example 6

##### 10 SBMu200 Sorghum Tassel RT-PCR and SBMu200 Maize cDNA Comparison

As noted above, SBMu200 is a male fertility gene. When it is mutated, male sterility will result. A homologue of SBMu200 was identified in sorghum. The sorghum-SBMu200 cDNA was isolated by using the maize SBMu200 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 SBMu200 cDNA from maize. Nucleotide sequence comparisons are set forth in Figure 10 and show 90% identity.

As is evident from the above, the SBMu200 gene is critical to male fertility in plants.

20 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. An isolated nucleic acid sequence encoding the promoter region of the gene SBMu200.
2. An isolated nucleic acid sequence comprising SEQ ID NO. 5 or those nucleotide sequences which hybridize to SEQ ID NO. 5 under conditions of high stringency.
3. A male tissue-preferred regulatory region comprising nucleotide sequences essential for initiating transcription of the SBMu200 gene.
4. A male tissue-preferred regulatory region comprising Seq. ID. NO. 6.
5. An isolated nucleic acid that is a male tissue specific regulatory element comprising a fragment of the nucleotide sequence of SEQ ID No. 6 or those nucleotide sequences which hybridize to SEQ ID. NO. 6 under conditions of high stringency wherein the regulatory element is essential for initiating transcription of Seq ID Nos. 1, 3 or 7.
6. The regulatory region of claim 3 comprising nucleotide sequences of about 130 contiguous base pairs from about -38 and higher upstream of the TATA box of the SBMu200 gene.
7. A regulatory region comprising the sequences of -44 to -180 upstream of the TATA box of SEQ ID NO. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency.
8. The regulatory region of claim 7 comprising sequences -92 to -176 upstream of the TATA box of SEQ ID. No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency.
9. The regulatory region of claim 7 comprising sequences -44 to -89 upstream of the TATA box of SEQ. ID. No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency.
10. The regulatory region of claim 7 comprising sequences -52 to -131 upstream of the TATA box of SEQ ID No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency.
11. The regulatory region of claim 7 comprising sequences -52 to -71 upstream of the TATA box of SEQ ID No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency.
12. The regulatory region of claim 7 comprising sequences -82 to -131 upstream of the TATA box of SEQ ID No. 7 or those nucleotide sequences which hybridize to said sequences under conditions of high stringency.
13. An expression vector comprising a promoter that is operably linked with the male tissue specific regulatory element of claim 3.

14. The expression vector of claim 13 further comprising a exogenous gene, wherein the exogenous gene is operably linked to the promoter.
15. The expression vector of claim 13 wherein the promoter is selected from any one of CaMV35S, SGB6, SBMu200 MS45 or 5126.
16. The expression vector of claim 13 wherein the product of the exogenous gene disrupts the function of male tissue.
17. Plant cells comprising the vector of claim 13.
18. A method of mediating male fertility in a plant comprising introducing into a plant the expression vector of claim 13 wherein the exogenous gene impacts male fertility of the plant and the regulatory element in conjunction with the promoter control expression of the exogenous gene.
19. The method of claim 18 wherein the exogenous gene disrupts function of male tissue of the plant causing the plant to be male sterile.
20. The method of claim 19 wherein the regulatory element in conjunction with the promoter is inducible.
21. The method of claim 18 wherein the plant is constitutively sterile when the promoter and regulatory element are not induced and is fertile when the promoter and regulatory element are induced.
22. The method of claim 18 further comprising cross-fertilizing the male sterile plant with a second plant, the second plant comprising a second exogenous gene, the product of the second gene preventing disruption of the male tissue by the first heterologous gene, producing a male fertile hybrid plant.
23. A method of producing hybrid seeds comprising: (a) producing a first parent plant comprising nucleotide sequences essential for initiating transcription of the SBMu200 gene operably linked with an exogenous gene impacting male fertility of the plant such that the plant is male sterile; (b) producing a second parent plant which is male fertile;(c) cross-fertilizing the first parent plant and the second parent plant to produce hybrid seeds.
24. The method of claim 23 wherein the gene impacting male fertility is dominant and further comprising growing the hybrid seed to produce a third male sterile parent plant; producing a fourth parent plant comprising one or more genes controlling a desired gene trait and cross-fertilizing the third and fourth parent plants to produce second hybrid seed.



1/21

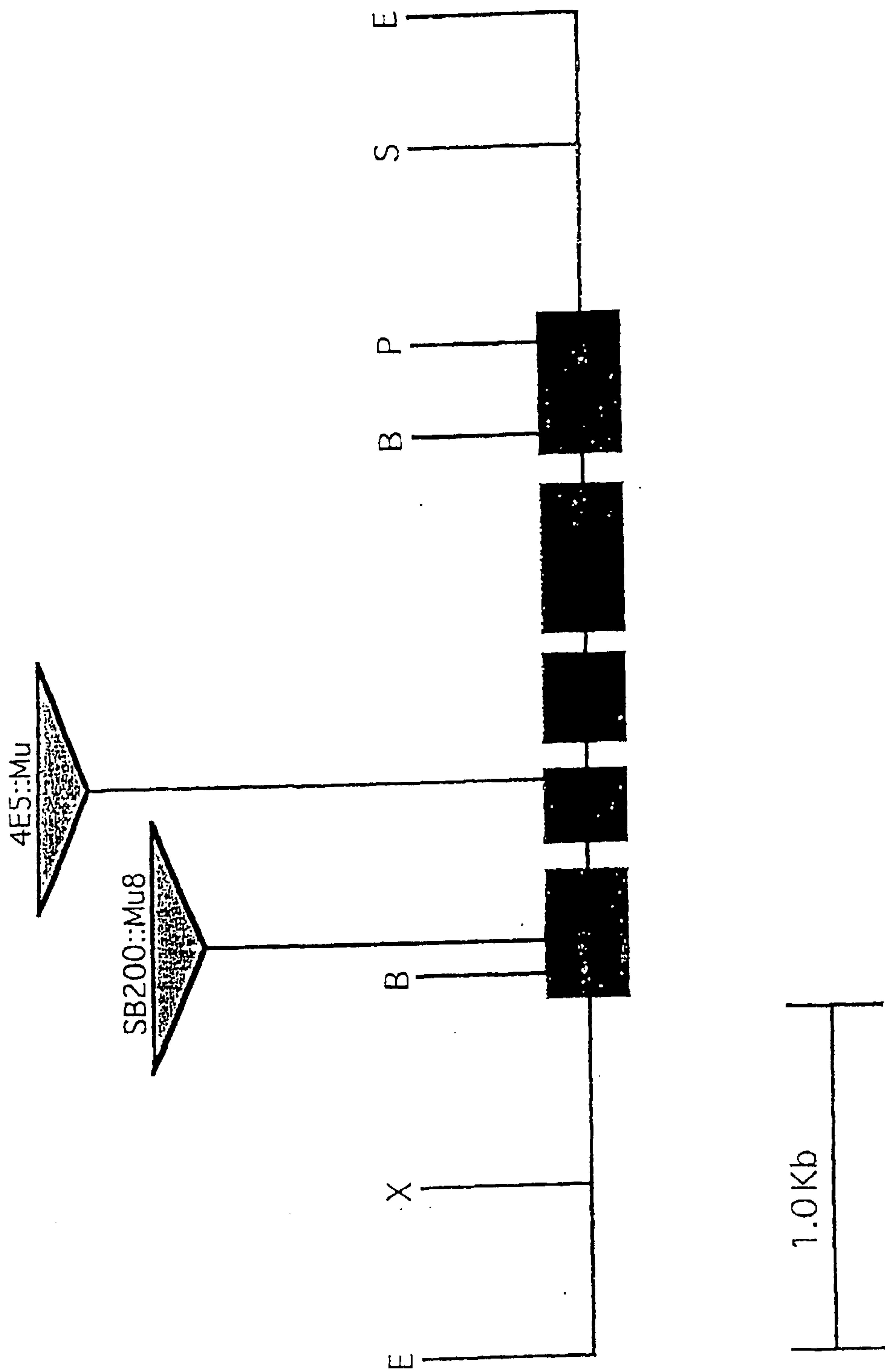


FIGURE 1

2/21

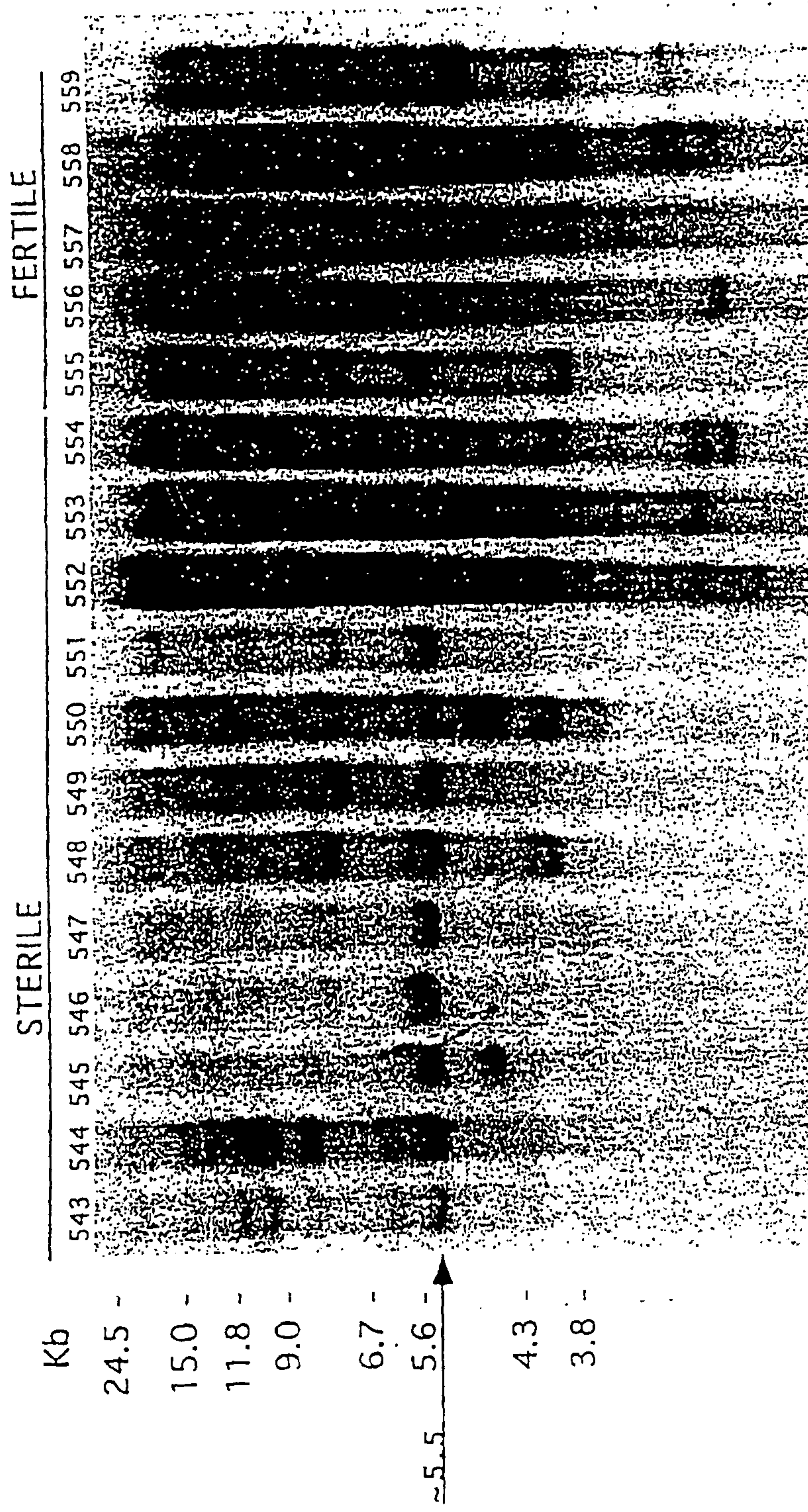


FIGURE 2





4/21

Figure 4

```

EcoRI
|
GAATTCGGCACGAGGGAAGCTCACCTCACGCCGGCGACGCCATCGCCATTCTTCCCCTA
1 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 60
CTTAAGCCGTGCTCCCTTCGAGTGGAGTGCGGCCGCTGCGGTAGCGGTAAGAAGGGTGAT

a   E F G T R E A H L T P A T P S P F F P L -

GCAGGGCCTCACAAGTACATCGCGCTCCTTCTGGTTGTCCTCTCATGGATCCTGGTCCAG
61 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 120
CGTCCCGGAGTGTTTCATGTAGCGCGAGGAAGACCAACAGGAGAGTACCTAGGACCAGGTC

a   A G P H K Y I A L L L V V L S W I L V Q -

AGGTGGAGCCTGAGGAAGCAGAAAGGCCCGAGATCATGGCCAGTCATCGGCGCAACGGTG
121 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 180
TCCACCTCGGACTCCTTCGTCTTCCGGGCTCTAGTACCGGTCAGTAGCCGCGTTGCCAC

a   R W S L R K Q K G P R S W P V I G A T V -

GAGCAGCTGAGGAACTACCACCGGATGCACGACTGGCTTGTCGGGTACCTGTCACGGCAC
181 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 240
CTCGTGCCTCCTTGATGGTGGCCTACGTGCTGACCGAACAGCCCATGGACAGTGCCGTG

a   E Q L R N Y H R M H D W L V G Y L S R H -

AGGACAGTGACCGTGCACATGCCGTTCACTTCCTACACCTACATCGCTGACCCGGTGAAT
241 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 300
TCCTGTCACTGGCAGCTGTACGGCAAGTGAAGGATGTGGATGTAGCGACTGGGCCACTTA

a   R T V T V D M P F T S Y T Y I A D P V N -

GTCGAGCATGTCCTCAAGACTAATTACCAATTACCCCAAGGGAATCGTGTACAGATCC
301 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 360
CAGCTCGTACAGGAGTTCTGATTGAAGTGGTTAATGGGGTTCCCTTAGCACATGTCTAGG

a   V E H V L K T N F T N Y P K G I V Y R S -

TACATGGACGTGCTCCTCGGTGACGGCATCTTCAACGCCGACGGCGAGCTGTGGAGGAAG
361 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 420
ATGTACCTGCACGAGGAGCCACTGCCGTAGAAGTTGCGGCTGCCGCTCGACACCTCCTTC

a   Y M D V L L G D G I F N A D G E L W R K -

CAGAGGAAGACGGCGAGTTTCGAGTTCGCCTCCAAGAACCTGAGGGATTTTCAGCGCCATT
421 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 480
GTCTCCTTCTGCCGCTCAAAGCTCAAGCGGAGGTTCTTGACTCCCTAAAGTCGCGGTAA

a   Q R K T A S F E F A S K N L R D F S A I -
    
```



5/21

Figure 4B

```

GTG TTCAGAGAGTACTCCCTGAAGCTGTCGGGTATACTGAGCCAGGCATCCAAGGCAGGC
481 -----+-----+-----+-----+-----+-----+-----+ 540
CACAAGTCTCTCATGAGGGACTTCGACAGCCCATATGACTCGGTCCGTAGGTTCCGTCCG
a   V F R E Y S L K L S G I L S Q A S K A G -
AAAGTTGTGGACATGCAGGAAC TTTACATGAGGATGACGCTGGACTCCATCTGCAAGGTT
541 -----+-----+-----+-----+-----+-----+-----+ 600
TTTCAACACCTGTACGTCCTTGAAATGTACTCCTACTGCGACCTGAGGTAGACGTTCCAA
a   K V V D M Q E L Y M R M T L D S I C K V -
GGG TTCGGGGTTCGAGATCGGCACGCTGTCGCCAGATCTCCCCGAGAACAGCTTCGCGCAG
601 -----+-----+-----+-----+-----+-----+-----+ 660
CCCAAGCCCCAGCTCTAGCCGTGCGACAGCGGTCTAGAGGGGCTCTTGTCGAAGCGCGTC
a   G F G V E I G T L S P D L P E N S F A Q -
GCG TTCGATGCCGCCAACATCATCATCACGCTGCGGTTTCATCGACCCGCTGTGGCGCATC
661 -----+-----+-----+-----+-----+-----+-----+ 720
CGCAAGCTACGGCGGTTGTAGTAGTAGTGCGACGCCAAGTAGCTGGGCGACACCGCGTAG
a   A F D A A N I I I T L R F I D P L W R I -
AAGAGTTCTTCCACGTCGGGTTCAGAGGCCCTCCTAGCGCAGAGCATCAAGCTCGTGGAC
721 -----+-----+-----+-----+-----+-----+-----+ 780
TTCTCCAAGAAGGTGCAGCCCAGTCTCCGGGAGGATCGCGTCTCGTAGTTCGAGCACCTG
a   K R F F H V G S E A L L A Q S I K L V D -
GAG TTCACCTACAGCGTGATCCGCCGGAGGAAGGCCGAGATCGTCGAGGTCCGGGCCAGC
781 -----+-----+-----+-----+-----+-----+-----+ 840
CTCAAGTGGATGTCGCACTAGGCCGGCCTCCTCCGGCTCTAGCAGCTCCAGGCCCGGTCCG
a   E F T Y S V I R R R K A E I V E V R A S -
GGCAAACAGGAGAAGATGAAGCACGACATCCTGTCACGGTTCATCGAGCTGGGCGAGGCC
841 -----+-----+-----+-----+-----+-----+-----+ 900
CCGTTTGTCTCTTCTACTTTCGTGCTGTAGGACAGTGCCAAGTAGCTCGACCCGCTCCGG
a   G K Q E K M K H D I L S R F I E L G E A -
GGCGACGACGGCGGGCTTCGGGGACGATAAGAGCCTCCGGGACGTGGTGCTCAACTTC
901 -----+-----+-----+-----+-----+-----+-----+ 960
CCGCTGCTGCCGCCCGAAGCCCCTGCTATTCTCGGAGGCCCTGCACCACGAGTTGAAG
a   G D D G G G F G D D K S L R D V V L N F -
GTGATCGCCGGGCGGGACACGACGGCGACGACGCTGTCGTGGTTCACGCACATGGCCATG
961 -----+-----+-----+-----+-----+-----+-----+ 1020
CACTAGCGGCCCGCCCTGTGCTGCGCTGCTGCGACAGCACCAAGTGCGTGTACCGGTAC
a   V I A G R D T T A T T L S W F T H M A M -

```

6/21

Figure 4C

```

TCCCACCCGGACGTGGCCGAGAAGCTGCGCCGAGCTGTGCGCGTTCGAGGCGGAGCGC
1021 -----+-----+-----+-----+-----+-----+-----+ 1080
AGGGTGGGCTGCACCGGCTCTTCGACGCGGCGCTCGACACGCGCAAGCTCCGCCTCGCG
a   S H P D V A E K L R R E L C A F E A E R -
GCGCGGAGGAGGGCGTCACGCTCGTGCTCTGCGGGCGGCGCTGACGCCGACGACAAGGCG
1081 -----+-----+-----+-----+-----+-----+-----+ 1140
CGCGCGCTCCTCCCGCAGTGCAGACGAGACGCCGCCGCGACTGCGGCTGCTGTTCCGC
a   A R E E G V T L V L C G G A D A D D K A -
TTCGCCGCCCGCGTGGCGCAGTTCGCGGGCCTCCTCACCTACGACAGCCTCGGCAAGCTG
1141 -----+-----+-----+-----+-----+-----+-----+ 1200
AAGCGGCGGGCGCACCGCGTCAAGCGCCCGGAGGAGTGGATGCTGTTCGGAGCCGTTTCGAC
a   F A A R V A Q F A G L L T Y D S L G K L -
GTCTACCTCCACGCCTGCGTCACCGAGACGCTCCGCCTGTACCCCGCCGTCCCTCAGGAC
1201 -----+-----+-----+-----+-----+-----+-----+ 1260
CAGATGGAGGTGCGGACGCAGTGGCTCTGCGAGGCGGACATGGGGCGGCAGGGAGTCCTG
a   V Y L H A C V T E T L R L Y P A V P Q D -
CCCAAGGGGATCCTGGAGGACGACGTGCTGCCGGACGGGACGAAGGTGAGGGCCGGCGGG
1261 -----+-----+-----+-----+-----+-----+-----+ 1320
GGGTTCCCCTAGGACCTCCTGCTGCACGACGGCCTGCCCTGCTTCCACTCCCAGCCGCC
a   P K G I L E D D V L P D G T K V R A G G -
ATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGTACAACCTGGGGCCCCGACGCG
1321 -----+-----+-----+-----+-----+-----+-----+ 1380
TACCACTGCATGCACGGGATGAGCTACCCCGCCTACCTCATGTTGACCCCGGGGCTGCGC
a   M V T Y V P Y S M G R M E Y N W G P D A -
GCGAGCTTCCGGCCGGAGCGGTGGATCAACGAGGATGGCGCGTTCGCAACGCGTTCGCCG
1381 -----+-----+-----+-----+-----+-----+-----+ 1440
CGCTCGAAGGCCGGCCTCGCCACCTAGTTGCTCCTACCGCGCAAGGCGTTGCGCAGCGGC
a   A S F R P E R W I N E D G A F R N A S P -
TTCAAGTTCACGGCGTTCCAGGCGGGGCCGAGGATCTGCCTGGGCAAGGACTCGGCGTAC
1441 -----+-----+-----+-----+-----+-----+-----+ 1500
AAGTTCAGTGCCGCAAGGTCCGCCCGGCTCCTAGACGGACCCGTTCTGAGCCGCATG
a   F K F T A F Q A G P R I C L G K D S A Y -
CTGCAGATGAAGATGGCGCTGGCCATCCTCTTCGCTTCTACAGCTTCCGGCTGCTGGAG
1501 -----+-----+-----+-----+-----+-----+-----+ 1560
GACGTCTACTTCTACCGCGACCGGTAGGAGAAGGCGAAGATGTCGAAGGCCGACGACCTC
a   L Q M K M A L A I L F R F Y S F R L L E -

```



Figure 4D

```

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

```

XhoI

```

ATATATATTATCCTCTTTCTTAAAAAAAAAAAAAAAAAACTCGAG
1861 -----+-----+-----+-----+-----+ 1906
TATATATAATAGGAGAAAGAATTTTTTTTTTTTTTTTTTTTGGAGCTC

```

8/21

Figure 5

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 TTCTTGTTG TTTGACATTT  
 351 ATAATATGTG GGGTGAGACG GATTAAATAT CATCCATGAG AGCTTTATCT  
 401 TCATGCTCTC TTGATTTTGG TTTCAGATCA TTCTTTCAGT GTTCACAAGA  
 451 ATTTTCTCAG TTTGGTCCAT GTAATTTTTG AAGTGAGGTT CCTTAAATTT  
 501 CATTATGCTT CCTTTCTTTT CTAGACTAGC AACTGCATGA CTTTTCACTT  
 551 TGGGTTCAACA AATTGACTCA CAAGAAAACA AATTCACTTT TGGGTTCAACA  
 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 GGTCAGGGAC ACCATATGAA TGAAAGAAAT CTTAATTTGG GGTACACCA  
 951 AGATTGTCTC TCTCGAGGTT GGGGGTCCC 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



9/21

Figure 5B

1401 AACTTCACCA ATTACCCCAA GGTAATGAC 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 TTCGGGGTTCG  
1851 AGATCGGCAC GCTGTCCCGG GATCTCCCCG AGAACAGCTT CGCGCAGGCG  
1901 TTCGATGCCG CCAACATCAT CGTCACGCTG CCGTTCATCG 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 ATTAAAATGG  
2201 GCAGATGAAG CACGACATCC TGTCACGGTT CATCGAGCTA GGCGAGGCCG  
2251 GCGACGACGG CGGCGGCTTC GGGGACGACA AGAGCCTCCG GGACGTGGTG  
2301 CTCAACTTCG TGATCGCCGG GCGGGACACG ACGGCGACGA CGCTGTCCGTG  
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 CCCC GCCGTC  
2601 CCTCAGGTGA GCGCGCCCGA CACGCGACCT CCGGTCCAGA GCACAGCATG  
2651 CAGTGAGTGG ACCTGAATGC AATGCACATG CACTTGCGCG CGCGCAGGAC  
2701 CCCAAGGGGA TCCTGGAGGA CGACGTGCTG CCGGACGGGA CGAAGGTGAG  
2751 GGCCGGCGGG ATGGTGACGT ACGTGCCCTA CTCGATGGGG CGGATGGAGT

10/21

Figure 5C

2801 ACAACTGGGG CCCCGACGCG 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 CCTTAAAAC TCCGTCGGGT  
3201 CTTGCGAACC ACCACATCAC TAGTGTTTTG TACTCTACTC CTCAGTGGAA  
3251 GTGTAGTGAC AGCATAACAAG TTCATCATAT ATATTATCCT CTTTCTTCGC  
3301 CGGATGCTTC CCGGGACCTT TTGGAGACCA TTACTGACAG GCGTGTGAAA  
3351 AAAAGGCTTC TTCTGCGGCG AAGTTTTGGG TTCAGAGTCT TGGCGTCTTT  
3401 GCAGCAGAAA AAAGGTTTGG AAGGATCTGA ACCCTGAACC GAAAATGGCT  
3451 TCGGAAATAT GCTCGCATCG GGGCGGGGCC GTCACTCGGG ATGACGACAA  
3501 GCCACAAGC AGTGAGAGCG AAGCGATCTT TGGAGTTTGG AGACACTCTC  
3551 GGACCCCTCG GCGCTCCGCG AGCTCATCTT CGCCTCCTCT GTCGTGTCCG  
3601 TGGCGGCACC GCGCCCGCCC GCCTCGTGTT CGACCAAATC CCGCGCCCCG  
3651 ACCGGTTCGT GTACAACACC CTCATCCGCG GCGCCGCGCG CAGTGACACG  
3701 CCCCGGGACG CCGTATACAT CTATAAATCA TGGTATTGTA CTTTATTTTC  
3751 AAACGGCCTT AACACAACCA TATTTTTATG GTAAACACGT TCAAATTGA  
3801 CACAAATTTA AAACAGGCAC AAACCGTAGC TAAACATAAG AGAATGAGAG  
3851 ACAACCCAAA GGTTAGAGAT GAAATAAGCT GAGTAAACGA CGAATTC



11/21

Figure 6

1051 TCCATCACTTGTAGACTGGACCCTTCATCAAGAGCACCATGGAGGAAGCT 1100  
 1 .....GAATTCGGCACGAGGGAAGCT 21

1101 CACATCACGCCGGCGACGCCATCGCCATTCTTCCCCTAGCAGGGCCTCA 1150  
 22 CACCTCACGCCGGCGACGCCATCGCCATTCTTCCCCTAGCAGGGCCTCA 71

1151 CAAGTACATCGCGCTCCTCCTGGTTGTCCTCTCATGGATCCTGGTCCAGA 1200  
 72 CAAGTACATCGCGCTCCTCCTGGTTGTCCTCTCATGGATCCTGGTCCAGA 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 AACTTCACCAATTACCCCAAGGTAAATGACCTGAACTCACTGATGTTTCAG 1450  
 322 AACTTCACCAATTACCCCA..... 340

1501 TAGGGAATCGTGTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCAT 1550  
 341 .AGGGAATCGTGTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCAT 389

1551 CTTCAACGCCGACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTT 1600  
 390 CTTCAACGCCGACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTT 439

1601 TCGAGTTCGCCTCCAAGAACCTGAGGGATTTAGCGCCATTGTGTTTCAGA 1650  
 440 TCGAGTTCGCCTCCAAGAACCTGAGGGATTTAGCGCCATTGTGTTTCAGA 489

1651 GAGTACTCCCTGAAGCTGTCCGGTATACTGAGCCAGGCATCCAAGGCAGG 1700  
 490 GAGTACTCCCTGAAGCTGTCCGGTATACTGAGCCAGGCATCCAAGGCAGG 539

1701 CAAAGTTGTGGACATGCAGGTGAGATCACTGCTCCCTTGCCATTGCCAAC 1750  
 540 CAAAGTTGTGGACATG..... 555

12/21

Figure 6B

1751 ATGAGCATTTC AACCTGAGACACGAGAGCTACCTTGCCGATTCAGGAACT 1800  
 556 .....CAGGAACT 563

1801 TTACATGAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTCTGGGGTCTG 1850  
 564 TTACATGAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTCTGGGGTCTG 613

1851 AGATCGGCACGCTGTCGCCGGATCTCCCCGAGAACAGCTTCGCGCAGGCG 1900  
 614 AGATCGGCACGCTGTCGCCAGATCTCCCCGAGAACAGCTTCGCGCAGGCG 663

1901 TTCGATGCCGCCAACATCATCATCAGCTGCGGTTTCATCGACCCGCTGTG 1950  
 664 TTCGATGCCGCCAACATCATCATCAGCTGCGGTTTCATCGACCCGCTGTG 713

1951 GCGCATCAAGAGGTTCTTCCACGTCCGGTTCAGAGGCCCTCCTAGCGCAGA 2000  
 714 GCGCATCAAGAGGTTCTTCCACGTCCGGTTCAGAGGCCCTCCTAGCGCAGA 763

2001 GCATCAAGCTCGTGGACGAGTTCACCTACAGCGTGATCCGCCGGAGGAAG 2050  
 764 GCATCAAGCTCGTGGACGAGTTCACCTACAGCGTGATCCGCCGGAGGAAG 813

2051 GCCGAGATCGTTCGAGGCCCGGGCCAGCGGCAAACAGGAGAAGGTACGTGC 2100  
 814 GCCGAGATCGTTCGAGGTCCGGGCCAGCGGCAAACAGGAGA..... 853

2201 GCAGATGAAGCACGACATCCTGTACCGTTCATCGAGCTAGGCGAGGCCG 2250  
 854 ..AGATGAAGCACGACATCCTGTACCGTTCATCGAGCTAGGCGAGGCCG 901

2251 GCGACGACGGCGGGCTTCGGGGACGACAAGAGCCTCCGGGACGTGGTG 2300  
 902 GCGACGACGGCGGGCTTCGGGGACGATAAGAGCCTCCGGGACGTGGTG 951

2301 CTCAACTTCGTGATCGCCGGGCGGGACACGACGGCGACGACGCTGTCGTG 2350  
 952 CTCAACTTCGTGATCGCCGGGCGGGACACGACGGCGACGACGCTGTCGTG 1001

2351 GTTCACGCACATGGCCATGTCCCACCCGGACGTGGCCGAGAAGCTGCGCC 2400  
 1002 GTTCACGCACATGGCCATGTCCCACCCGGACGTGGCCGAGAAGCTGCGCC 1051

2401 GCGAGCTGTGCGCGTTCGAGGCGGAGCGCGCGCGGAGGAGGGCGTCGCG 2450  
 1052 GCGAGCTGTGCGCGTTCGAGGCGGAGCGCGCGCGGAGGAGGGCGTCACG 1101

2451 CTCGTGCCCTGCGGCGGCGCTGACGCCGACGACAAGGCGTTCGCCGCCCG 2500  
 1102 CTCGTGCTCTGCGGCGGCGCTGACGCCGACGACAAGGCGTTCGCCGCCCG 1151



13/21

Figure 6C

```

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 GAGGATGGCGCGTTCGCAACGCGTCGCCGTTCAAGTTCACGGCGTTCCA 2900
      |||
1411 GAGGATGGCGCGTTCGCAACGCGTCGCCGTTCAAGTTCACGGCGTTCCA 1460

2901 GGCGGGGCCGAGGATCTGCCTGGGCAAGGACTCGGCGTACCTGCAGATGA 2950
      |||
1461 GGCGGGGCCGAGGATCTGCCTGGGCAAGGACTCGGCGTACCTGCAGATGA 1510

2951 AGATGGCGCTGGCCATCCTCTTTCGCTTCTACAGCTTCCGGCTGCTGGAG 3000
      |||
1511 AGATGGCGCTGGCCATCCTCTTTCGCTTCTACAGCTTCCGGCTGCTGGAG 1560

3001 GGGCACCCGGTGCAGTACCGCATGATGACCATCCTCTCCATGGCGCACGG 3050
      |||
1561 GGGCACCCGGTGCAGTACCGCATGATGACCATCCTCTCCATGGCGCACGG 1610

3051 CCTCAAGGTCCGCGTCTCTAGGGCCGTCTGATGTCATGGCGATTTG.... 3096
      |||
1611 CCTCAAGGTCCGCGTCTCTAGGGCCGTCTGATGTCATGGCGATTTGGATA 1660

3097 .GGATATCATCCCGCTTAATCC.....TTAAAAATT 3126
      |||
1661 TGGATATCGTCCCGCTTAATCCACGACAAATAACGCTCGTGTTACAAATT 1710

3127 TGCATGCATGCATGTAAGGGAAAGCGATGGGTTTCATTGGTGGCTTGGCT 3176
      |||
1711 TGCATGCATGCATGTAAGGGAAAGCGATGGGTTTCATTGGTGGCTTGGCT 1760
    
```

14/21

Figure 6D

3177 TAAGCCTTAAAAACTCCGTCCGGTCTTGCGAACCACCACATCACTAGTGT 3226  
|||||  
1761 TAAGCCTTAAAAACTCCGTCCGGTCTTGCGAACCACCACATCACTAGTGT 1810  
3227 TTTGTACTCTACTCCTCAGTGGAAGTGTAGTGACAGCATAACAAGTTCATC 3276  
|||||  
1811 TTTGTACTCTACTCCTCAGTGGAAGTGTAGTGACAGCATAACAAGTTCATC 1860  
3277 ATATATATTATCCTCTTTCTTCGCCGGATGCTTCCCGGGACCTTTTGGAG 3326  
||||| | ||  
1861 ATATATATTATCCTCTTTCTTAAAAAAAAAAAAAAAAAACTCGAG.... 1906  
.  
.  
.



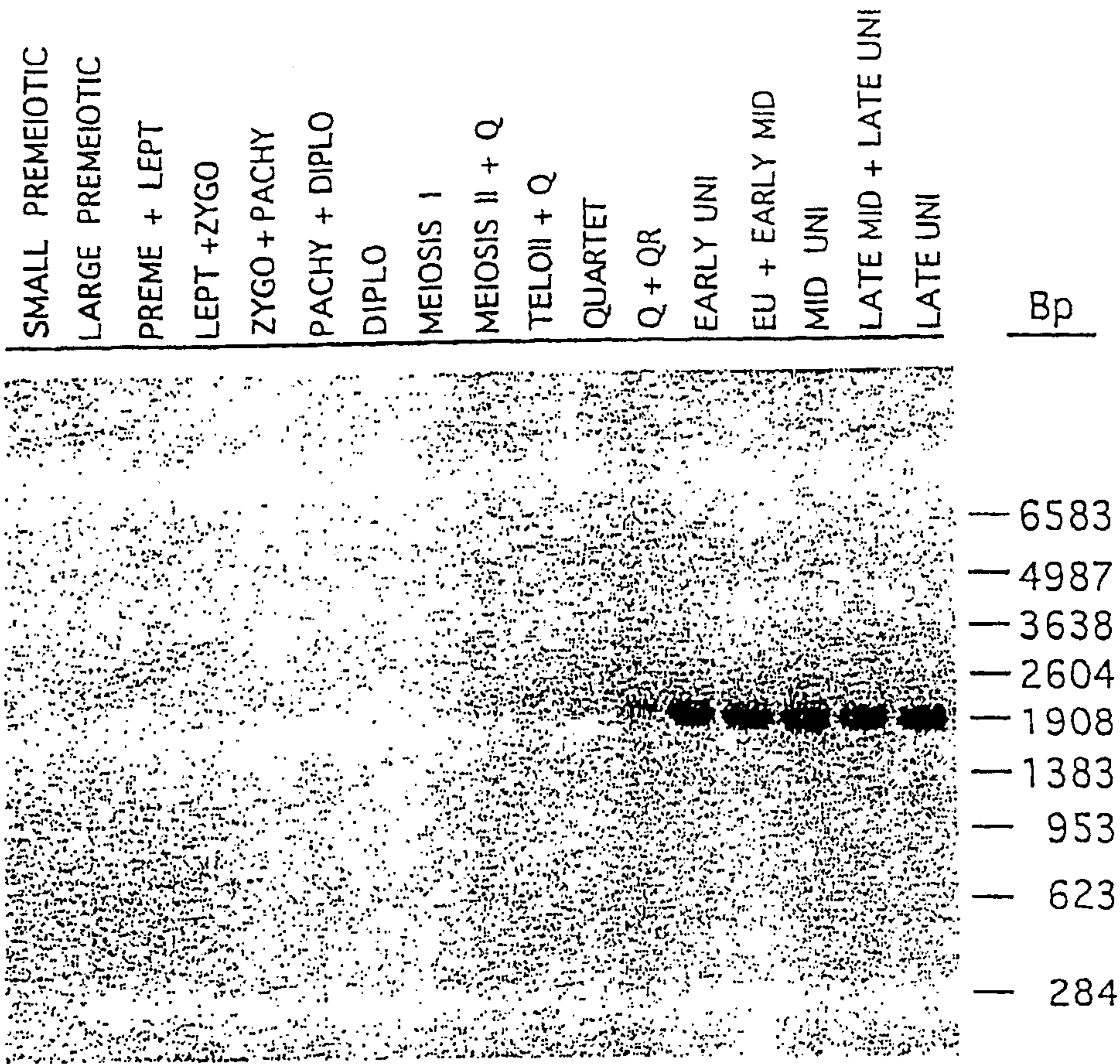


Figure 7

16/21

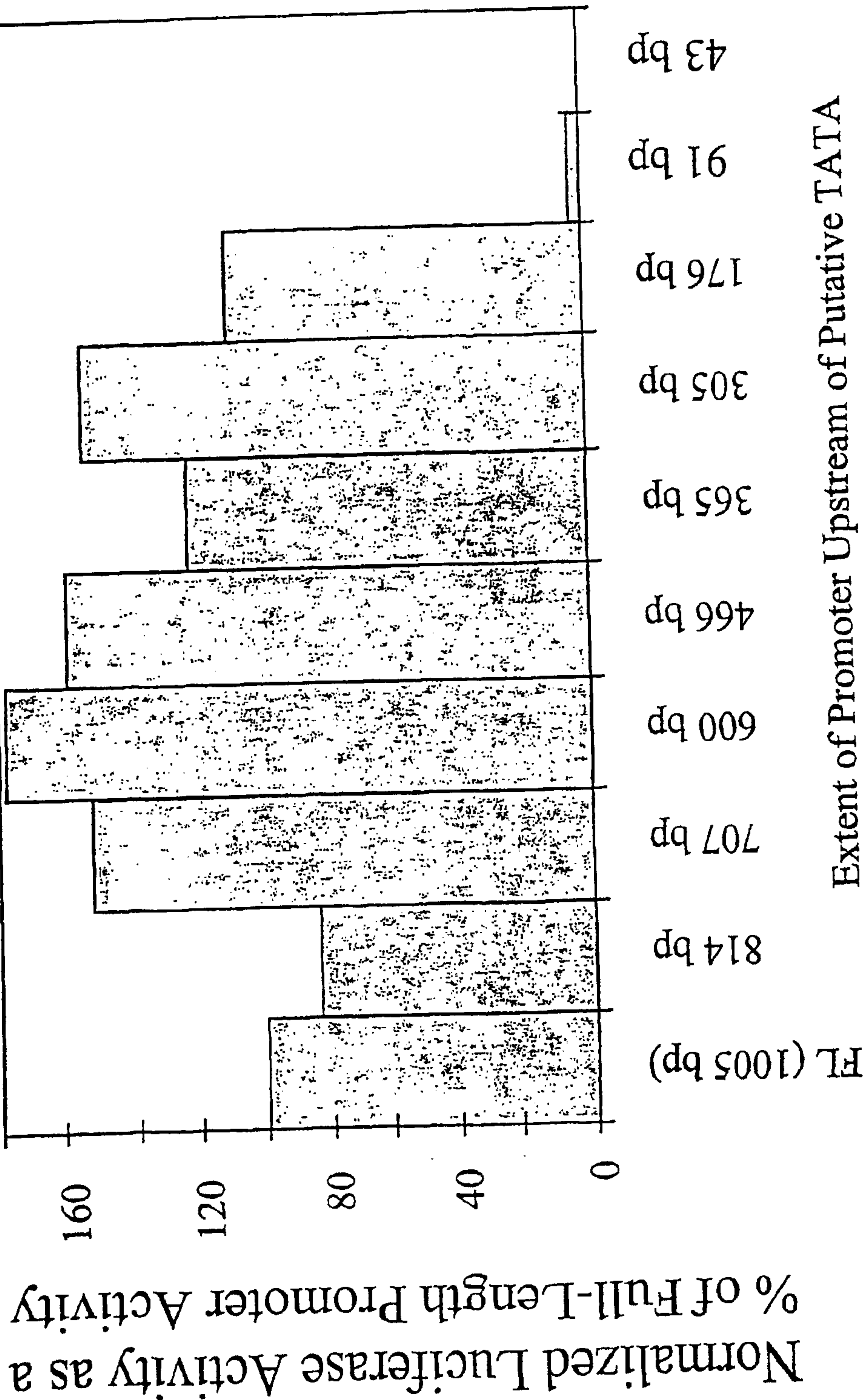
Figure 8

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 CTTTTCACTT  
 551 TGGGTTCACA AATTGACTCA CAAGAAAACA AATTCACCTT TGGGTTCACA  
 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 GGTCAGGGAC ACCATATGAA TGAAAGAAAT CTTAATTTGG GGTCACACCA  
 951 AGATTGTCTC TCTCGAGGTT GGGGGTCCC TAAGGTTGGT AGTAGCAATA  
 1001 CCCAATATAT CACCTAACAA ACCCAATCCA TGCTACATAC ATACATAGCA  
 1051 TCCATCACTT GTAGACTGGA CCCTTCATCA AGAGCACCAT GG



# SBMu200 Promoter Analysis: 5'

## Deletions



Extent of Promoter Upstream of Putative TATA  
Figure 9

# SBMu200 ‘Minimal’ Promoter

```

-180 CCCATCTCA TTTCTTGGT TGGCAGATCA CAAAAGGAA CCAAAGGCT
      LS01 | LS02 | LS03 | LS04 | LS05
-130 AAGCCTCCTA CTGTTCGGG AGTTAGGTCA GGGACACCT ATGAATGAAA
      LS06 | LS07 | LS08 | LS09 | LS10
-80  GAAATCTTAA TTTGGGGTCA CACCAAGATT GTCCTCTCG AGGTTGGGG
      LS11 | LS12 | LS13 | LS14 | LS15
-30  GTCCCTAAGG TTGGTAGTAG CAATACCCAA TATATCACCT AACAAACCCA
      LS16 | LS17 | LS18
20  ATCCATGCTA CATAATACA TAGCATCCAT CACTTGTAGA CTGGACCCTT

70  CATCAAGAGC ACCATGG
    
```

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

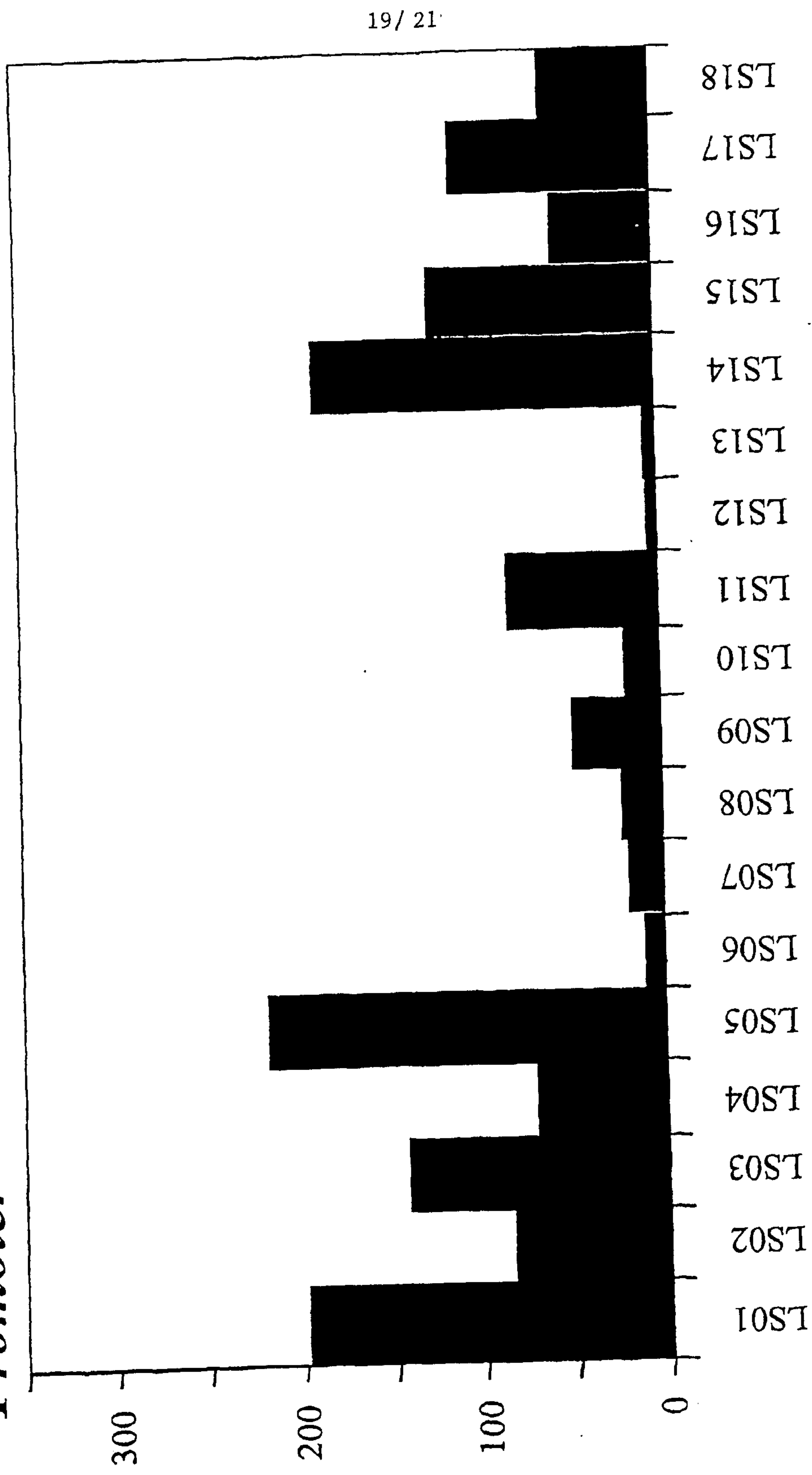
Figure 10



# Linker Scanning Analysis of SBMu200

Promoter

Luci/Gus as % of Wild Type



Mutant

Figure 11

20/21

Figure 12

```

1 .....GGAA 4
201 CCGGATGCACGACTGGCTTGTCCGGTACCTGTCACGGCACAGGACAGTGA 250
5 TTCGGCTTATGCCGTTCACTTCTACACCTACATCGCTGACCCGGTGAAT 54
251 CCGTCGACATGCCGTTCACTTCTACACCTACATCGCTGACCCGGTGAAT 300
55 GTCGAGCATGTCCTCAAGACTAACTTCACCAATTACCCCAAGGGGGACGT 104
301 GTCGAGCATGTCCTCAAGACTAACTTCACCAATTACCCCAAGGGAATCGT 350
105 GTACAGATCCTACATGGATGTGCTCCTCGGTGACGGCATATTCAACGCTG 154
351 GTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCATCTTCAACGCCG 400
155 ACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCGCC 204
401 ACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCGCC 450
205 TCCAAGAACCTGAGGGATTTTCAGTGCCAATGTTTTTCAGAGAGTACTCCCT 254
451 TCCAAGAACCTGAGGGATTTTCAGCGCCATTGTGTTTCAGAGAGTACTCCCT 500
255 GAAGCTGTCGGGCATACTGAGTCAGGCATCCAAGGCAGGCAAAGTTGTTG 304
501 GAAGCTGTCGGGTATACTGAGCCAGGCATCCAAGGCAGGCAAAGTTGTTG 550
305 ACATGCAGGAACTTTACATGAGGATGACACTGGACTCGATCTGCAANGTT 354
551 ACATGCAGGAACTTTACATGAGGATGACGCTGGACTCCATCTGCAAGGTT 600
355 GGGTTCGGGGTCNANATCGGCACGCTGTNCNCCGGATCTCCCCGAGAACAG 404
601 GGGTTCGGGGTCGAGATCGGCACGCTGTGCCAGATCTCCCCGAGAACAG 650
405 CTTNCCCAAGCGTTCGATGCCGCTAACATCATCGTCACTGCGGTTCA 454
651 CTTGCGCAGGCGTTCGATGCCGCCAACATCATCATCACGCTGCGGTTCA 700
455 TCCACCCNCTGTGGCGCATCCAGAAGTTCTTCCCCNGTCA..... 494
701 TCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTCAGAGGCC 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 LWRKQRKTASFEFASKNLRDFSANVFREYSLKLSGILSQASKAGKVVDQM 104
137 LWRKQRKTASFEFASKNLRDFS AIVFREYSLKLSGILSQASKAGKVVDQM 186

```



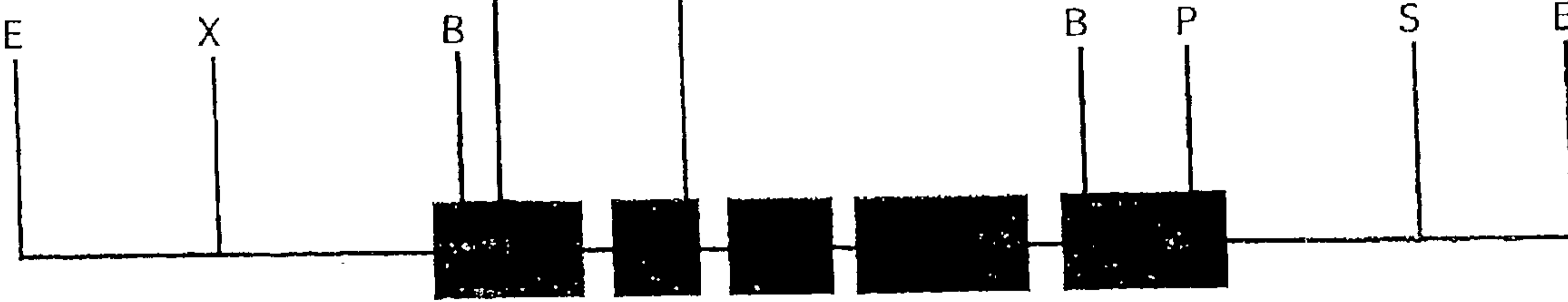
21/21.

Figure 12B

105 ELYMRMTLDSICKXVGFVXIGTLPDLPENSFXQAFDAANIIVTLRFIHP 154  
|||||:|||||. |  
187 ELYMRMTLDSICKVGFVXIGTLPDLPENSFAQAFDAANIIITLRFIDP 236  
  
155 LWRIQKFF 162  
|||.:||  
237 LWRIKRF 244

4E5::Mu

SB200::Mu8



1.0 Kb

