



(86) Date de dépôt PCT/PCT Filing Date: 2003/07/10
(87) Date publication PCT/PCT Publication Date: 2004/01/29
(45) Date de délivrance/Issue Date: 2014/09/30
(85) Entrée phase nationale/National Entry: 2005/01/12
(86) N° demande PCT/PCT Application No.: US 2003/021551
(87) N° publication PCT/PCT Publication No.: 2004/009761
(30) Priorité/Priority: 2002/07/18 (US60/396,665)

(51) Cl.Int./Int.Cl. *C12N 15/54* (2006.01),
A01H 5/00 (2006.01), *C12N 15/09* (2006.01),
C12N 15/82 (2006.01), *C12N 5/10* (2006.01),
C12N 9/10 (2006.01), *C12P 19/34* (2006.01),
C12Q 1/68 (2006.01)
(72) Inventeur/Inventor:
FLASINSKI, STANISLAW, US
(73) Propriétaire/Owner:
MONSANTO TECHNOLOGY LLC, US
(74) Agent: OSLER, HOSKIN & HARCOURT LLP

(54) Titre : PROCEDES PERMETTANT D'UTILISER DES POLYNUCLEOTIDES ARTIFICIELS ET COMPOSITIONS
DESDITS POLYNUCLEOTIDES AFIN DE REDUIRE LE SILENCAGE TRANSGENIQUE
(54) Title: METHODS FOR USING ARTIFICIAL POLYNUCLEOTIDES AND COMPOSITIONS THEREOF TO REDUCE
TRANSGENE SILENCING

(57) **Abrégé/Abstract:**

The materials and methods disclosed provide for polynucleotide molecules sufficiently divergent from polynucleotides naturally contained in plants, or polynucleotides previously introduced into plants as transgenes to permit trait stacking in plant breeding methods or plant transformation methods. The disclosure also provides for methods and compositions to detect the polynucleotides of the invention in plants.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
29 January 2004 (29.01.2004)

PCT

(10) International Publication Number
WO 2004/009761 A3

(51) International Patent Classification⁷: C12N 15/09,
15/63, 15/82, 15/87, 15/90

(21) International Application Number:
PCT/US2003/021551

(22) International Filing Date: 10 July 2003 (10.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/396,665 18 July 2002 (18.07.2002) US

(71) Applicant (for all designated States except US): MON-
SANTO TECHNOLOGY LLC [US/US]; 800 N. Lind-
bergh Boulevard, St. Louis, MO 63167 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): FLASINSKI, Stanis-
law [PL/US]; 2325 Laurenwood Drive, Chesterfield, MO
63017 (US).

(74) Agent: KAMMERER, Patricia, A.; Howrey Simon
Arnold & White, LLP, 750 Bering Drive, Houston, TX
77057-2198 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC,
SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,
SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

(88) Date of publication of the international search report:
29 July 2004

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: METHODS FOR USING ARTIFICIAL POLYNUCLEOTIDES AND COMPOSITIONS THEREOF TO REDUCE
TRANSGENE SILENCING

(57) Abstract: The materials and methods disclosed provide for polynucleotide molecules sufficiently divergent from polynu-
cleotides naturally contained in plants, or polynucleotides previously introduced into plants as transgenes to permit trait stacking
in plant breeding methods or plant transformation methods. The disclosure also provides for methods and compositions to detect the
polynucleotides of the invention in plants.



WO 2004/009761 A3

-1-

METHODS FOR USING ARTIFICIAL POLYNUCLEOTIDES AND COMPOSITIONS THEREOF TO REDUCE TRANSGENE SILENCING

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to plant genetic engineering. More particularly, to a method for constructing an artificial polynucleotide and methods of use to reduce transgene silencing in plants. The invention also relates to the plant cells containing the artificial polynucleotide in which a plant cell is transformed to express the artificial polynucleotide and the plant regenerated therefrom.

Description of the Related Art

Heterologous genes may be isolated from a source other than the plant into which it will be transformed or they may be modified or designed to have different or improved qualities. Particularly desirable traits or qualities of interest for plant genetic engineering would include but are not limited to resistance to insects, fungal diseases, and other pests and disease-causing agents, tolerances to herbicides, enhanced stability or shelf-life, yield, environmental stress tolerances, and nutritional enhancements.

Traditional molecular biological methods for generating novel genes and proteins generally involved random or directed mutagenesis. An example of random mutagenesis is a recombination technique known as "DNA shuffling" as disclosed in US Patents 5,605,793; 5,811,238; 5,830,721; 5,837,458 and International Applications WO 98/31837, WO 99/65927. An alternative method of molecular evolution involves a staggered extension process (StEP) for in vitro mutagenesis and recombination of nucleic acid molecule sequences, as disclosed in US Patent 5,965,408. An example of directed mutagenesis is the introduction of a point mutation at a specific site in a polypeptide.

An alternative approach, useful when the heterologous gene is from a non-plant source, is to design an artificial insecticidal gene that uses the most often used codon in maize plant codon usage table (Koziel *et al.*, 1993, Biotechnology 11, 194-200). Fischhoff and Perlak (US Patent

-2-

No. 5,500,365) report higher expression of *Bacillus thuringiensis* (Bt) insecticidal protein compared in crop plants when the polynucleotide sequence was modified to reduce the occurrence of destabilizing sequences. It was necessary to modify the wild type Bt polynucleotide sequence because the wild type full length Bt polynucleotide did not express sufficient levels of insecticidal protein in plants to be agronomically useful.

Heterologous genes are cloned into vectors suitable for plant transformation. Transformation and regeneration techniques useful to incorporate heterologous genes into a plant's genome are well known in the art. The gene can then be expressed in the plant cell to exhibit the added characteristic or trait. However, heterologous genes that normally express well as transgenes may experience gene silencing when more than one copy of the same genes are expressed in the same plant. This may occur when a first heterologous gene is too similar to an endogenous gene DNA sequence in the plant. Other examples include when a transgenic plant is subsequently crossed to other transgenic plants having the same or similar transgenes or when the transgenic plant is retransformed with a plant expression cassette that contains the same or similar gene. Similarly, gene silencing may occur if trait stacking employs the same genetic elements used to direct expression of the transgene gene of interest. In order to stack traits, stable transgenic lines should be done with different combinations of genes and genetic elements to avoid gene silencing.

N-phosphonomethylglycine, also known as glyphosate, is a well-known herbicide that has activity on a broad spectrum of plant species. Glyphosate is the active ingredient of Roundup® (Monsanto Co.), a safe herbicide having a desirably short half-life in the environment. When applied to a plant surface, glyphosate moves systemically through the plant. Glyphosate is phytotoxic due to its inhibition of the shikimic acid pathway, which provides a precursor for the synthesis of aromatic amino acids. Glyphosate inhibits the enzyme 5-enolpyruvyl-3-phosphoshikimate synthase (EPSPS).

Glyphosate tolerance can also be achieved by the expression of EPSPS variants that have lower affinity for glyphosate and therefore retain their catalytic activity in the presence of glyphosate (U.S. Patent No. 5,633,435). Enzymes that degrade glyphosate in plant tissues (U.S. Patent No. 5,463,175) are also capable of conferring cellular tolerance to glyphosate. Such genes are used for the production of transgenic crops that are tolerant to glyphosate, thereby allowing glyphosate to be used for effective weed control with minimal concern of crop damage. For example, glyphosate tolerance has been genetically

-3-

engineered into corn (U.S. Patent No. 5,554,798), wheat (U.S. Patent Application No. 20020062503), soybean (U.S. Patent Application No. 20020157139) and canola (WO 9204449). The transgenes for glyphosate tolerance and the transgenes for tolerance to other herbicides, *e.g.* *bar* gene, (Toki *et al.* Plant Physiol., 100:1503-1507, 1992; Thompson *et al.* EMBO J. 6:2519-2523, 1987, phosphinothricin acetyltransferase, BAR gene isolated from *Streptomyces*; DeBlock *et al.* EMBO J., 6:2513-2522, 1987, glufosinate herbicide) are also useful as selectable markers or scorable markers and can provide a useful phenotype for selection of plants linked with other agronomically useful traits.

What is needed in the art are methods to design genes for expression in plants to improve agronomically useful traits that avoid gene silencing when multiple copies are inserted and recombination with endogenous plant genes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Pileup comparison of the polynucleotide sequences changes of two artificial rice EPSPS versions (OsEPSPS_AT, OsEPSPS_ZM) and a native rice EPSPS (OsEPSPS_Nat) the polypeptide of each modified to be glyphosate resistant.

Figure 2. Pileup comparison of the polynucleotide sequences of a native (ZmEPSPS_Nat) and an artificial corn EPSPS (ZmEPSPS_ZM) the polypeptide of each modified to be glyphosate resistant.

-4-

Figure 3. Pileup comparison of the polynucleotide sequences of a soybean native EPSPS (GmEPSPS_Nat) and artificial version (GmEPSPS_GM) the polypeptide of each modified to be glyphosate resistant.

Figure 4. Pileup comparison of the polynucleotide sequences of a native BAR gene (BAR1_Nat) and two artificial versions with Zea mays (BAR1_ZM) and Arabidopsis thaliana (BAR1_AT) codon bias.

Figure 5. Pileup comparison of the polynucleotide sequences of CTP2 and CP4EPSPS native (CTP2CP4_Nat) and artificial versions (CTP2CP4_AT, CTP2CP4_ZM, and CTP2CP4_GM).

Figure 6. Plasmid map of pMON54949.

Figure 7. Plasmid map of pMON54950.

Figure 8. Plasmid map of pMON30151.

Figure 9. Plasmid map of pMON59302.

Figure 10. Plasmid map of pMON59307.

Figure 11. Plasmid map of pMON42411.

Figure 12. Plasmid map of pMON58400.

Figure 13. Plasmid map of pMON58401.

Figure 14. Plasmid map of pMON54964.

Figure 15. Plasmid map of pMON25455.

Figure 16. Plasmid map of pMON30152.

Figure 17. Plasmid map of pMON54992.

Figure 18. Plasmid map of pMON54985.

Figure 19. Plasmid map of pMON20999.

Figure 20. Plasmid map of pMON45313.

Figure 21. Plasmid map of pMON59308.

Figure 22. Plasmid map of pMON59309.

Figure 23. Plasmid map of pMON59313.

Figure 24. Plasmid map of pMON59396.

Figure 25. Plasmid map of pMON25496.

-5-

BRIEF DESCRIPTION OF SEQUENCE LISTING

SEQ ID NO:1	OsEPSPS_TIPS	A rice EPSPS protein sequence modified to be glyphosate resistant, with chloroplast transit peptide.
SEQ ID NO:2	OsEPSPS_Nat	Polynucleotide sequence of a rice native EPSPS polynucleotide modified to encode a glyphosate resistant protein.
SEQ ID NO:3	OsEPSPS_AT	Polynucleotide sequence of an artificial rice EPSPS polynucleotide using the <i>Arabidopsis</i> codon usage table and the methods of the present invention, and further modified to encode a glyphosate resistant protein.
SEQ ID NO:4	OsEPSPS_ZM	Polynucleotide sequence of an artificial rice EPSPS polynucleotide using the <i>Zea mays</i> codon usage table and the methods of the present invention, and further modified to encode a glyphosate resistant protein.
SEQ ID NO:5	GmEPSPS_IKS	A soybean EPSPS protein sequence modified to be glyphosate resistant, with chloroplast transit peptide.
SEQ ID NO:6	GmEPSPS_Nat	Polynucleotide sequence of a soybean native EPSPS polynucleotide modified to encode a glyphosate resistant protein.
SEQ ID NO:7	GmEPSPS_GM	Polynucleotide sequence of an artificial soybean EPSPS polynucleotide using the <i>Glycine max</i> codon usage table and the methods of the present invention, and further modified to encode a glyphosate resistant protein.
SEQ ID NO:8	ZmEPSPS_TIPS	A corn EPSPS protein sequence modified to be glyphosate resistant, with chloroplast transit peptide.

-6-

SEQ ID NO:9	ZmEPSPS_Nat	Polynucleotide sequence of a corn native EPSPS polynucleotide modified to encode a glyphosate resistant protein.
SEQ ID NO:10	ZmEPSPS_ZM	Polynucleotide sequence of an artificial corn EPSPS polynucleotide using the <i>Zea mays</i> codon usage table and the methods of the present invention, and further modified to encode a glyphosate resistant protein.
SEQ ID NO:11	CTP2	Protein sequence of the chloroplast transit peptide 2 from <i>Arabidopsis</i> EPSPS gene.
SEQ ID NO:12	CTP2_Nat	Polynucleotide sequence of the chloroplast transit peptide from <i>Arabidopsis</i> EPSPS.
SEQ ID NO:13	CTP2_AT	Polynucleotide sequence of an artificial polynucleotide encoding the CTP2 using the <i>Arabidopsis</i> codon usage table and the methods of the present invention.
SEQ ID NO:14	CTP2_ZM	Polynucleotide sequence of an artificial polynucleotide encoding the CTP2 using the <i>Zea mays</i> codon usage table and the methods of the present invention.
SEQ ID NO:15	CP4EPSPS	The protein sequence of the glyphosate resistant EPSPS protein from <i>Agrobacterium</i> strain CP4.
SEQ ID NO:16	CP4EPSPS_Nat	Polynucleotide sequence of the native polynucleotide encoding the CP4EPSPS protein (U.S. Patent No. 5,633,435).
SEQ ID NO:17	CP4EPSPS_AT	Polynucleotide sequence of an artificial polynucleotide encoding the CP4EPSPS protein using the <i>Arabidopsis</i> codon usage table and the methods of the present invention.

-7-

SEQ ID NO:18	CP4EPSPS_ZM	Polynucleotide sequence of an artificial polynucleotide encoding the CP4EPSPS protein using the <i>Zea mays</i> codon usage table and the methods of the present invention.
SEQ ID NO:19	BAR1	The protein sequence of a phosphinothricin acetyltransferase.
SEQ ID NO:20	BAR1_Nat	Polynucleotide sequence of the native polynucleotide isolated from <i>Streptomyces</i> encoding the phosphinothricin acetyltransferase.
SEQ ID NO:21	BAR1_AT	Polynucleotide sequence of an artificial polynucleotide encoding the phosphinothricin acetyltransferase using the <i>Arabidopsis</i> codon usage table and the methods of the present invention.
SEQ ID NO:22	BAR1_ZM	Polynucleotide sequence of an artificial polynucleotide encoding the phosphinothricin acetyltransferase using the <i>Zea mays</i> codon usage table and the methods of the present invention.
SEQ ID NO:23	CP4EPSPS_Syn	Polynucleotide sequence of an artificial polynucleotide with dicot codon bias.
SEQ ID NO:24	CP4EPSPS_AT_p1	DNA primer molecule diagnostic for the CP4EPSPS_AT polynucleotide.
SEQ ID NO:25	CP4EPSPS_AT_p2	DNA primer molecule diagnostic for the CP4EPSPS_AT polynucleotide.
SEQ ID NO:26	CP4EPSPS_ZM_p1	DNA primer molecule diagnostic for the CP4EPSPS_ZM polynucleotide.
SEQ ID NO:27	CP4EPSPS_ZM_p2	DNA primer molecule diagnostic for the CP4EPSPS_ZM polynucleotide.
SEQ ID NO:28	CP4EPSPS_Nat_p1	DNA primer molecule diagnostic for the CP4EPSPS_Nat polynucleotide.

-8-

SEQ ID NO:29	CP4EPSPS_Nat_p2	DNA primer molecule diagnostic for the CP4EPSPS_Nat polynucleotide.
SEQ ID NO:30	CP4EPSPS_Syn_p1	DNA primer molecule diagnostic for the CP4EPSPS_Syn polynucleotide.
SEQ ID NO:31	CP4EPSPS_Syn_p2	DNA primer molecule diagnostic for the CP4EPSPS_Syn polynucleotide.
SEQ ID NO:32	ZmAdh1 primer1	Control primer 1 diagnostic for endogenous corn Adh1 gene.
SEQ ID NO:33	ZmAdh1 primer2	Control primer 2 diagnostic for endogenous corn Adh1 gene.
SEQ ID NO:34	GNAGIAMKS	Motif providing glyphosate resistance to a plant EPSPS.
SEQ ID NO:35	CTPEPSPSCP4_G M	Polynucleotide sequence of an artificial polynucleotide encoding the CP4EPSPS protein using the <i>Glycine max</i> codon usage table.

SUMMARY OF THE INVENTION

The present invention provides methods and compositions to design an artificial polynucleotide sequence that encodes a protein of interest, wherein the artificial polynucleotide is substantially divergent from a polynucleotide naturally occurring in a plant or a polynucleotide that has been introduced as a transgene into a plant and the artificial polynucleotide and polynucleotide encode a substantially identical polypeptide.

The artificial polynucleotides of the present invention that encodes proteins that provide agronomically useful phenotypes to a transgenic plant containing a DNA construct comprising the artificial polynucleotide. The agronomically useful phenotypes include, but are not limited to: drought tolerance, increased yield, cold tolerance, disease resistance, insect resistance and herbicide tolerance.

Another aspect of the present invention are artificial polynucleotides that encode a herbicide resistant EPSPS protein, a phosphinothricin acetyltransferase protein, a chloroplast transit peptide protein. In preferred embodiments of the present invention, the artificial polynucleotide molecule is selected from the group consisting of SEQ ID NO:3, SEQ ID NO:4,

-9-

SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:35.

The present invention provides DNA constructs comprising: a promoter molecule that functions in plants, operably linked to an artificial polynucleotide molecule of the present invention, wherein the artificial polynucleotide molecule is selected from the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:35, operably linked to a transcription termination region.

The present invention further provides DNA constructs comprising: a promoter molecule that functions in plants, operably linked to an artificial polynucleotide molecule that encodes a chloroplast transit peptide, operably linked to a heterologous glyphosate resistant EPSPS, operably linked to a transcription termination signal region, wherein the artificial polynucleotide is substantially divergent in polynucleotide sequence from known polynucleotides encoding an identical chloroplast transit peptide.

The present invention provides DNA constructs comprising at least two expression cassettes, the first expression cassette comprising a promoter molecule that functions in plants, operably linked to an artificial polynucleotide molecule of the present invention, operably linked to a transcription termination signal region, and the second expression cassette comprising a promoter molecule that functions in plants, operably linked to a polynucleotide molecule that encodes a substantially identical polypeptide as said artificial polynucleotide and is less than eight-five percent similar in polynucleotide sequence to said artificial polynucleotide, operably linked to a transcription termination signal region.

The present invention provides plant cells, plants or progeny thereof comprising a DNA construct of the present invention. Of particular interest are plants of progeny thereof selected from the group consisting of wheat, corn, rice, soybean, cotton, potato, canola, turf grass, forest trees, grain sorghum, vegetable crops, ornamental plants, forage crops, and fruit crops.

A method of the present invention reduces gene silencing during breeding of transgenic plants comprising the steps of:

- a) constructing an artificial polynucleotide that is substantially divergent from known polynucleotides that encode a substantially identical protein, and
- b) constructing a DNA construct containing said artificial polynucleotide molecule; and
- c) transforming said DNA construct into a plant cell; and

-10-

d) regenerating said plant cell into a transgenic plant; and

e) crossing said transgenic plant with a fertile plant, wherein said fertile plant contains a polynucleotide molecule that encodes a protein substantially identical to a protein encoded by said artificial polynucleotide molecule and wherein said artificial polynucleotide molecule and said polynucleotide molecule are substantially divergent.

Another aspect of the invention is a transgenic plant cell comprising two polynucleotides, wherein at least one of the polynucleotides is a transgene and the two polynucleotides encode a substantially identical protein and are less than eight-five percent similar in polynucleotide sequence.

Another aspect of the present invention is a method to reduce gene silencing during production of transgenic plants comprises the steps of:

a) constructing an artificial polynucleotide that is substantially divergent from known polynucleotides that encode a substantially identical protein, and

b) constructing a first DNA construct containing said artificial polynucleotide molecule;

and

c) transforming said DNA construct into a plant cell; and

d) regenerating said plant cell into a transgenic plant; and

e) retransforming a cell from said transgenic plant with a second DNA construct comprising a polynucleotide molecule that encodes a substantially identical protein to said artificial polynucleotide and said polynucleotide and artificial polynucleotide are substantially divergent in polynucleotide sequence; and

f) regenerating said cell of step d into a transgenic plant comprising both said artificial polynucleotide and said polynucleotide.

Further provided by the present invention are methods for selection of a plants transformed with a DNA construct of the invention comprising the steps of:

a) transforming a plant cell with a DNA construct of the present invention; and

b) culturing said plant cell in a selective medium containing a herbicide selected from the group consisting of: glyphosate and glufosinate, to selectively kill cells which have not been transformed with said DNA constructs; and

c) regenerating said plant cell into a fertile plant.

Another aspect of the invention is a method of detecting an artificial polynucleotide in a transgenic plant cell, plant or progeny thereof comprising the steps:

-11-

a) contacting a DNA sample isolated from said plant cell, plant or progeny thereof with a DNA molecule, wherein said DNA molecule comprises at least one DNA molecule of a pair of DNA molecules that when used in a nucleic-acid amplification reaction produces an amplicon that is diagnostic for said artificial polynucleotide molecule selected from the group consisting of: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:35.

- (a) performing a nucleic acid amplification reaction, thereby producing the amplicon; and
- (b) detecting the amplicon.

Reagents provided for performing the detection method above include, but are not limited to: DNA molecules that specifically hybridize to an artificial polynucleotide molecule selected from the group consisting of: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, and SEQ ID NO:22; and isolated DNA molecules selected from the group consisting of: SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27.

The present invention provides plants, and progeny comprising a DNA molecule selected from the group consisting of: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27.

The present invention provides pairs of DNA molecules selected from the group comprising: a first DNA molecule and a second DNA molecule, wherein the first DNA molecule is SEQ ID NO:24 or its complement and the second DNA molecule is SEQ ID NO:25 or its complement and the pair of DNA molecules when used in a DNA amplification method produce an amplicon, and a first DNA molecule and a second DNA molecule, wherein the first DNA molecule is SEQ ID NO:26 or its complement and the second DNA molecule is SEQ ID NO:27 or its complement and the pair of DNA molecules when used in a DNA amplification method produce an amplicon, wherein the amplicon is diagnostic for the presence of an artificial polynucleotide of the present invention in the genome of a transgenic plant.

The present invention provides for a plant and progeny thereof identified by a DNA amplification method to contain in its genome a DNA molecule selected from the group consisting of: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ

-12-

ID NO:13, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21 SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27.

The present invention provides and contemplates DNA detection kits comprising: at least one DNA molecule of sufficient length to be specifically homologous or complementary to an artificial polynucleotide selected from the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, and SEQ ID NO:22, wherein said DNA molecule is useful as a DNA probe or DNA primer; or at least one DNA molecule homologous or complementary to a DNA primer molecule selected from the group consisting of: SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27.

The present invention further provides a method of detecting the presence of an artificial polynucleotide encoding a glyphosate resistant EPSPS in a DNA sample, the method comprising:

(a) extracting a DNA sample from a plant; and

(b) contacting the DNA sample with a labeled DNA molecule of sufficient length to be specifically homologous or complementary to an artificial polynucleotide selected from the group consisting of: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:17, and SEQ ID NO:18, wherein said labeled DNA molecule is a DNA probe; and

(c) subjecting the sample and DNA probe to stringent hybridization conditions; and

(d) detecting the DNA probe hybridized to the DNA sample.

The present invention provides for an isolated polynucleotide that encodes an EPSPS enzyme, the EPSPS enzyme contains the motif of SEQ ID NO:34. The present invention provides for a DNA construct containing a polynucleotide that encodes for the EPSPS enzyme with the motif of SEQ ID NO:34. A plant cell, plant or progeny thereof that is tolerant to glyphosate as a result of expressing an EPSPS enzyme that contains the motif of SEQ ID NO:34 is an aspect of the invention.

DETAILED DESCRIPTION OF THE INVENTION

The following definitions are provided to better define the present invention and to guide those of ordinary skill in the art in the practice of the present invention. Unless otherwise noted, terms are to be understood according to conventional usage by those of ordinary skill in the

-13-

relevant art. Definitions of common terms in molecular biology may also be found in Rieger *et al.*, Glossary of Genetics: Classical and Molecular, 5th edition, Springer-Verlag: New York, (1991); and Lewin, Genes V, Oxford University Press: New York, (1994). The nomenclature for DNA bases as set forth at 37 CFR § 1.822 is used. The standard one- and three-letter
5 nomenclature for amino acid residues is used.

“Amino-acid substitutions”, “Amino-acid variants”, are preferably substitutions of single amino-acid residue for another amino-acid residue at any position within the protein. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a final construct.

10 An “artificial polynucleotide” as used in the present invention is a DNA sequence designed according to the methods of the present invention and created as an isolated DNA molecule for use in a DNA construct that provides expression of a protein in host cells, and for the purposes of cloning into appropriate constructs or other uses known to those skilled in the art. Computer programs are available for these purposes, including but not limited to the “BestFit” or
15 “Gap” programs of the Sequence Analysis Software Package, Genetics Computer Group (GCG), Inc., University of Wisconsin Biotechnology Center, Madison, WI 53711. The artificial polynucleotide may be created by a one or more methods known in the art, that include, but are not limited to: overlapping PCR. An artificial polynucleotide of the present invention is substantially divergent from other polynucleotides that code for the identical or nearly identical
20 protein.

The term “chimeric” refers to a fusion nucleic acid or protein sequence. A chimeric nucleic acid coding sequence is comprised of two or more sequences joined in-frame that encode a chimeric protein. A chimeric gene refers to the multiple genetic elements derived from heterologous sources comprising a gene.

25 The phrases “coding sequence”, “open reading frame”, and “structural sequence” refer to the region of continuous sequential nucleic acid triplets encoding a protein, polypeptide, or peptide sequence.

“Codon” refers to a sequence of three nucleotides that specify a particular amino acid.

“Codon usage” or “codon bias” refers to the frequency of use of codons encoding amino
30 acids in the coding sequences of organisms.

-14-

“Complementarity” and “complement” when referring to nucleic acid sequences, refers to the specific binding of adenine to thymine (uracil in RNA) and cytosine to guanine on opposite strands of DNA or RNA.

“Construct” refers to the heterologous genetic elements operably linked to each other making up a recombinant DNA molecule and may comprise elements that provide expression of a DNA polynucleotide molecule in a host cell and elements that provide maintenance of the construct.

“C-terminal region” refers to the region of a peptide, polypeptide, or protein chain from the middle thereof to the end that carries the amino acid having a free carboxyl group.

The term “divergent”, as used herein, refers to the comparison of polynucleotide molecules that encode the same or nearly the same protein or polypeptide. The four letter genetic code (A, G, C, and T/U) comprises three letter codons that direct t-RNA molecules to assemble amino acids into a polypeptide from an mRNA template. Having more than one codon that may code for the same amino acid is referred to as degenerate. Degenerate codons are used to construct substantially divergent polynucleotide molecules that encode the same polypeptide where these molecules have a sequence of nucleotides of their entire length in which they are less than 85% identical, and there are no lengths of polynucleotide sequence greater than 23 nucleotides that are identical.

The term “encoding DNA” refers to chromosomal DNA, plasmid DNA, cDNA, or artificial DNA polynucleotide that encodes any of the proteins discussed herein.

The term “endogenous” refers to materials originating from within an organism or cell.

“Endonuclease” refers to an enzyme that hydrolyzes double stranded DNA at internal locations.

“Exogenous” refers to materials originating from outside of an organism or cell. This typically applies to nucleic acid molecules used in producing transformed or transgenic host cells and plants.

“Exon” refers to the portion of a gene that is actually translated into protein, *i.e.*, a coding sequence.

The term “expression” refers to the transcription or translation of a polynucleotide to produce a corresponding gene product, a RNA or protein.

-15-

“Fragments”. A fragment of a gene is a portion of a full-length polynucleic acid molecule that is of at least a minimum length capable of transcription into a RNA, translation into a peptide, or useful as a probe or primer in a DNA detection method.

The term “gene” refers to chromosomal DNA, plasmid DNA, cDNA, artificial DNA polynucleotide, or other DNA that encodes a peptide, polypeptide, protein, or RNA molecule, and the genetic elements flanking the coding sequence that are involved in the regulation of expression.

The term “genome” as it applies to viruses encompasses all of the nucleic acid sequence contained within the capsid of the virus. The term “genome” as it applies to bacteria encompasses both the chromosome and plasmids within a bacterial host cell. Encoding nucleic acids of the present invention introduced into bacterial host cells can therefore be either chromosomally-integrated or plasmid-localized. The term “genome” as it applies to plant cells encompasses not only chromosomal DNA found within the nucleus, but organelle DNA found within subcellular components of the cell. Nucleic acids of the present invention introduced into plant cells can therefore be either chromosomally-integrated or organelle-localized.

“Glyphosate” refers to N-phosphonomethylglycine and its’ salts, Glyphosate is the active ingredient of Roundup® herbicide (Monsanto Co.). Plant treatments with “glyphosate” refer to treatments with the Roundup® or Roundup Ultra® herbicide formulation, unless otherwise stated. Glyphosate as N-phosphonomethylglycine and its’ salts (not formulated Roundup® herbicide) are components of synthetic culture media used for the selection of bacteria and plant tolerance to glyphosate or used to determine enzyme resistance in in vitro biochemical assays.

“Heterologous DNA” sequence refers to a polynucleotide sequence that originates from a foreign source or species or, if from the same source, is modified from its original form.

“Homologous DNA” refers to DNA from the same source as that of the recipient cell.

“Hybridization” refers to the ability of a strand of nucleic acid to join with a complementary strand via base pairing. Hybridization occurs when complementary sequences in the two nucleic acid strands bind to one another. The nucleic acid probes and primers of the present invention hybridize under stringent conditions to a target DNA sequence. Any conventional nucleic acid hybridization or amplification method can be used to identify the presence of DNA from a transgenic event in a sample. Nucleic acid molecules or fragments thereof are capable of specifically hybridizing to other nucleic acid molecules under certain circumstances. As used herein, two nucleic acid molecules are said to be capable of specifically

-16-

hybridizing to one another if the two molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure. A nucleic acid molecule is said to be the "complement" of another nucleic acid molecule if they exhibit complete complementarity. As used herein, molecules are said to exhibit "complete complementarity" when every nucleotide of one of the
5 molecules is complementary to a nucleotide of the other. Two molecules are said to be "minimally complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under at least conventional "low-stringency" conditions. Similarly, the molecules are said to be "complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under
10 conventional "high-stringency" conditions. Conventional stringency conditions are described by Sambrook *et al.*, 1989, and by Haymes *et al.*, In: *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985).

Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form a double-stranded structure. In
15 order for a nucleic acid molecule to serve as a primer or probe it need only be sufficiently complementary in sequence to be able to form a stable double-stranded structure under the particular solvent and salt concentrations employed.

As used herein, a substantially homologous sequence is a nucleic acid sequence that will specifically hybridize to the complement of the nucleic acid sequence to which it is being
20 compared under high stringency conditions. The term "stringent conditions" is functionally defined with regard to the hybridization of a nucleic-acid probe to a target nucleic acid (*i.e.*, to a particular nucleic-acid sequence of interest) by the specific hybridization procedure discussed in Sambrook *et al.*, 1989, at 9.52-9.55. *See also*, Sambrook *et al.*, 1989 at 9.47-9.52, 9.56-9.58; Kanehisa, (Nucl. Acids Res. 12:203-213, 1984); and Wetmur and
25 Davidson, (J. Mol. Biol. 31:349-370, 1988). Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of DNA fragments. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degrees of
30 selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to

-17-

about 0.15 M NaCl at temperatures of about 50°C to about 70°C. A stringent conditions, for example, is to wash the hybridization filter at least twice with high-stringency wash buffer (0.2X SSC, 0.1% SDS, 65° C). Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Both temperature and salt may be varied, or either the temperature or the salt concentration may be held constant while the other variable is changed. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand. Detection of DNA sequences via hybridization is well-known to those of skill in the art, and the teachings of U.S. Pat. Nos. 4,965,188 and 5,176,995 are exemplary of the methods of hybridization analyses.

“Identity” refers to the degree of similarity between two polynucleic acid or protein sequences. An alignment of the two sequences is performed by a suitable computer program. A widely used and accepted computer program for performing sequence alignments is CLUSTALW v1.6 (Thompson, *et al.* Nucl. Acids Res., 22: 4673-4680, 1994). The number of matching bases or amino acids is divided by the total number of bases or amino acids, and multiplied by 100 to obtain a percent identity. For example, if two 580 base pair sequences had 145 matched bases, they would be 25 percent identical. If the two compared sequences are of different lengths, the number of matches is divided by the shorter of the two lengths. For example, if there are 100 matched amino acids between a 200 and a 400 amino acid protein, they are 50 percent identical with respect to the shorter sequence. If the shorter sequence is less than 150 bases or 50 amino acids in length, the number of matches are divided by 150 (for nucleic acid bases) or 50 (for amino acids), and multiplied by 100 to obtain a percent identity.

As described herein a protein can be “substantially identical” to related proteins. These proteins with substantial identity generally comprise at least one polypeptide sequence that has at least ninety-eight sequence percent identity compared to its related other polypeptide sequence. The Gap program in the WISCONSIN PACKAGE version 10.0-UNIX from Genetics Computer Group, Inc. is based on the method of Needleman and Wunsch (J. Mol. Biol. 48:443-453, 1970)

-18-

using the set of default parameters for pairwise comparison (for amino acid sequence comparison: Gap Creation Penalty = 8, Gap Extension Penalty = 20); or using the TBLASTN program in the BLAST 2.2.1 software suite (Altschul et al., Nucleic Acids Res. 25:3389-3402), using BLOSUM62 matrix (Henikoff and Henikoff, Proc. Natl. Acad. Sci. U.S.A. 89:10915-10919, 1992) and the set of default parameters for pair-wise comparison (gap creation cost = 11, gap extension cost = 1.). In BLAST, the E-value, or expectation value, represents the number of different alignments with scores equivalent to or better than the raw alignment score, S, that are expected to occur in a database search by chance. The lower the E value, the more significant the match. Because database size is an element in E-value calculations, E-values obtained by "BLASTing" against public databases, such as GenBank, have generally increased over time for any given query/entry match. Percent identity refers to the percentage of identically matched amino acid residues that exist along the length of that portion of the sequences which is aligned by the BLAST algorithm.

"Intron" refers to a portion of a gene not translated into protein, even though it is transcribed into RNA.

An "isolated" nucleic acid sequence is substantially separated or purified away from other nucleic acid sequences with which the nucleic acid is normally associated in the cell of the organism in which the nucleic acid naturally occurs, *i.e.*, other chromosomal or extrachromosomal DNA. The term embraces nucleic acids that are biochemically purified so as to substantially remove contaminating nucleic acids and other cellular components. The term also embraces recombinant nucleic acids and chemically synthesized nucleic acids.

"Isolated," "Purified," "Homogeneous" polypeptides. A polypeptide is "isolated" if it has been separated from the cellular components (nucleic acids, lipids, carbohydrates, and other polypeptides) that naturally accompany it or that is chemically synthesized or recombinant. A monomeric polypeptide is isolated when at least 60% by weight of a sample is composed of the polypeptide, preferably 90% or more, more preferably 95% or more, and most preferably more than 99%. Protein purity or homogeneity is indicated, for example, by polyacrylamide gel electrophoresis of a protein sample, followed by visualization of a single polypeptide band upon staining the polyacrylamide gel; high pressure liquid chromatography; or other conventional methods. Proteins can be purified by any of the means known in the art, for example as described in *Guide to Protein Purification*, ed. Deutscher, Meth. Enzymol. 185, Academic Press, San

-19-

Diego, 1990; and *Scopes, Protein Purification: Principles and Practice*, Springer Verlag, New York, 1982.

“Labeling” or “labeled”. There are a variety of conventional methods and reagents for labeling polynucleotides and polypeptides and fragments thereof. Typical labels include radioactive isotopes, ligands or ligand receptors, fluorophores, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, *e.g.*, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989) and *Current Protocols in Molecular Biology*, ed. Ausubel *et al.*, Greene Publishing and Wiley-Interscience, New York, (1992).

“Mature protein coding region”, this term refers to the sequence of a processed protein product, *i.e.*, a mature EPSP synthase remaining after the chloroplast transit peptide has been removed.

“Native”, the term “native” generally refers to a naturally-occurring (“wild-type”) polynucleic acid or polypeptide. However, in the context of the present invention, some modification of an isolated polynucleotide and polypeptide may have occurred to provide a polypeptide with a particular phenotype, *e.g.*, amino acid substitution in glyphosate sensitive EPSPS to provide a glyphosate resistant EPSPS. For comparative purposes in the present invention, the isolated polynucleotide that contains a few substituted nucleotides to provide amino acid modification for herbicide tolerance is referred to as the “native” polynucleotide when compared to the substantially divergent polynucleotide created by the methods of the present invention. However, the “native” polynucleotide modified in this manner is nonnative with respect to the genetic elements normally found linked to a naturally occurring unmodified polynucleotide.

“N-terminal region” refers to a region of a peptide, polypeptide, or protein chain from the amino acid having a free amino group to the middle of the chain.

“Nucleic acid” refers to deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

– Nucleic acid codes: A = adenosine; C = cytosine; G = guanosine; T = thymidine. Codes used for synthesis of oligonucleotides: N = equimolar A, C, G, and T; I = deoxyinosine; K = equimolar G and T; R = equimolar A and G; S = equimolar C and G; W = equimolar A and T; Y = equimolar C and T.

A “nucleic acid segment” or a “nucleic acid molecule segment” is a nucleic acid molecule that has been isolated free of total genomic DNA of a particular species, or that has

-20-

been synthesized. Included with the term "nucleic acid segment" are DNA segments, recombinant vectors, plasmids, cosmids, phagemids, phage, viruses, et cetera.

"Nucleotide Sequence Variants", using well-known methods, the skilled artisan can readily produce nucleotide and amino acid sequence variants of genes and proteins, respectively.

5 For example, "variant" DNA molecules of the present invention are DNA molecules containing changes in an EPSPS gene sequence, i.e., changes that include one or more nucleotides of the EPSPS gene sequence is deleted, added, and/or substituted, such that the variant EPSPS gene encodes a protein that retains EPSPS activity. Variant DNA molecules can be produced, for example, by standard DNA mutagenesis techniques or by chemically synthesizing the variant
10 DNA molecule or a portion thereof. Methods for chemical synthesis of nucleic acids are discussed, for example, in Beaucage *et al.*, Tetra. Letts. 22:1859-1862 (1981), and Matteucci *et al.*, J. Am. Chem. Soc. 103:3185- (1981). Chemical synthesis of nucleic acids can be performed, for example, on automated oligonucleotide synthesizers. Such variants preferably do not change the reading frame of the protein-coding region of the nucleic acid and preferably encode a
15 protein having no change, or only a minor reduction.

"Open reading frame (ORF)" refers to a region of DNA or RNA encoding a peptide, polypeptide, or protein.

"Operably Linked". A first nucleic-acid sequence is "operably" linked with a second nucleic-acid sequence when the first nucleic-acid sequence is placed in a functional relationship
20 with the second nucleic-acid sequence. For example, a promoter is operably linked to a protein-coding sequence if the promoter effects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in reading frame.

"Overexpression" refers to the expression of a RNA or polypeptide or protein encoded
25 by a DNA introduced into a host cell, wherein the RNA or polypeptide or protein is either not normally present in the host cell, or wherein the RNA or polypeptide or protein is present in said host cell at a higher level than that normally expressed from the endogenous gene encoding the RNA or polypeptide or protein.

The term "plant" encompasses any higher plant and progeny thereof, including monocots
30 (*e.g.*, corn, rice, wheat, barley, *etc.*), dicots (*e.g.*, soybean, cotton, canola, tomato, potato, *Arabidopsis*, tobacco, *etc.*), gymnosperms (pines, firs, cedars, *etc.*) and includes parts of plants,

-21-

including reproductive units of a plant (e.g., seeds, bulbs, tubers, fruit, flowers, *etc.*) or other parts or tissues from that the plant can be reproduced.

“Plant expression cassette” refers to chimeric DNA segments comprising the regulatory elements that are operably linked to provide the expression of a transgene product in plants

5 “Plasmid” refers to a circular, extrachromosomal, self-replicating piece of DNA.

“Polyadenylation signal” or “polyA signal” refers to a nucleic acid sequence located 3’ to a coding region that causes the addition of adenylate nucleotides to the 3’ end of the mRNA transcribed from the coding region.

10 “Polymerase chain reaction (PCR)” refers to a DNA amplification method that uses an enzymatic technique to create multiple copies of one sequence of nucleic acid (amplicon). Copies of a DNA molecule are prepared by shuttling a DNA polymerase between two amplimers. The basis of this amplification method is multiple cycles of temperature changes to denature, then re-anneal amplimers (DNA primer molecules), followed by extension to synthesize new DNA strands in the region located between the flanking amplimers. Nucleic-acid
15 amplification can be accomplished by any of the various nucleic-acid amplification methods known in the art, including the polymerase chain reaction (PCR). A variety of amplification methods are known in the art and are described, *inter alia*, in U.S. Patent Nos. 4,683,195 and 4,683,202 and in *PCR Protocols: A Guide to Methods and Applications*, ed. Innis *et al.*, Academic Press, San Diego, 1990. PCR amplification methods have been developed to amplify
20 up to 22 kb of genomic DNA and up to 42 kb of bacteriophage DNA (Cheng *et al.*, Proc. Natl. Acad. Sci. USA 91:5695-5699, 1994). These methods as well as other methods known in the art of DNA amplification may be used in the practice of the present invention.

Polynucleotide refers to a length of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) molecules greater than two, which are connected to form a larger molecule.

25 Polypeptide fragments. The present invention also encompasses fragments of a protein that lacks at least one residue of a native full-length protein, but that substantially maintains activity of the protein.

The term “promoter” or “promoter region” refers to a polynucleic acid molecule that functions as a regulatory element, usually found upstream (5’) to a coding sequence, that controls
30 expression of the coding sequence by controlling production of messenger RNA (mRNA) by providing the recognition site for RNA polymerase and/or other factors necessary for start of transcription at the correct site. As contemplated herein, a promoter or promoter region includes

-22-

variations of promoters derived by means of ligation to various regulatory sequences, random or controlled mutagenesis, and addition or duplication of enhancer sequences. The promoter region disclosed herein, and biologically functional equivalents thereof, are responsible for driving the transcription of coding sequences under their control when introduced into a host as part of a suitable recombinant vector, as demonstrated by its ability to produce mRNA.

"Recombinant". A "recombinant" nucleic acid is made by a combination of two otherwise separated segments of sequence, *e.g.*, by chemical synthesis or by the manipulation of isolated segments of nucleic acids by genetic engineering techniques.

The term "recombinant DNA construct" or "recombinant vector" refers to any agent such as a plasmid, cosmid, virus, autonomously replicating sequence, phage, or linear or circular single-stranded or double-stranded DNA or RNA nucleotide sequence, derived from any source, capable of genomic integration or autonomous replication, comprising a DNA molecule that one or more DNA sequences have been linked in a functionally operative manner. Such recombinant DNA constructs or vectors are capable of introducing a 5' regulatory sequence or promoter region and a DNA sequence for a selected gene product into a cell in such a manner that the DNA sequence is transcribed into a functional mRNA that is translated and therefore expressed. Recombinant DNA constructs or recombinant vectors may be constructed to be capable of expressing antisense RNAs, in order to inhibit translation of a specific RNA of interest.

"Regeneration" refers to the process of growing a plant from a plant cell (*e.g.*, plant protoplast or explant).

"Reporter" refers to a gene and corresponding gene product that when expressed in transgenic organisms produces a product detectable by chemical or molecular methods or produces an observable phenotype.

"Resistance" refers to an enzyme that is able to function in the presence of a toxin, for example, glyphosate resistant class II EPSP synthases. An enzyme that has resistance to a toxin may have the function of detoxifying the toxin, *e.g.*, the phosphinothricin acetyltransferase, glyphosate oxidoreductase, or may be a mutant enzyme having catalytic activity which is unaffected by an herbicide which disrupts the same activity in the wild type enzyme, *e.g.*, acetolactate synthase, mutant class I EPSP synthases.

"Restriction enzyme" refers to an enzyme that recognizes a specific palindromic sequence of nucleotides in double stranded DNA and cleaves both strands; also called a restriction endonuclease. Cleavage typically occurs within the restriction site.

-23-

“Selectable marker” refers to a polynucleic acid molecule that encodes a protein, which confers a phenotype facilitating identification of cells containing the polynucleic acid molecule. Selectable markers include those genes that confer resistance to antibiotics (e.g., ampicillin, kanamycin), complement a nutritional deficiency (e.g., uracil, histidine, leucine), or impart a visually distinguishing characteristic (e.g., color changes or fluorescence). Useful dominant selectable marker genes include genes encoding antibiotic resistance genes (e.g., neomycin phosphotransferase, *aad*); and herbicide resistance genes (e.g., phosphinothricin acetyltransferase, class II EPSP synthase, modified class I EPSP synthase). A useful strategy for selection of transformants for herbicide resistance is described, e.g., in Vasil, *Cell Culture and Somatic Cell Genetics of Plants*, Vols. I-III, Laboratory Procedures and Their Applications Academic Press, New York (1984).

The term “specific for (a target sequence)” indicates that a DNA probe or DNA primer hybridizes under given hybridization conditions only to the target sequence in a sample comprising the target sequence.

The term “substantially purified”, as used herein, refers to a molecule separated from other molecules normally associated with it in its native state. More preferably, a substantially purified molecule is the predominant species present in a preparation. A substantially purified molecule may be greater than 60% free, preferably 75% free, more preferably 90% free from the other molecules (exclusive of solvent) present in the natural mixture. The term “substantially purified” is not intended to encompass molecules present in their native state.

“Tolerant” or “tolerance” refers to a reduced effect of a biotic or abiotic agent on the growth and development of organisms and plants, e.g. a pest or a herbicide.

“Transcription” refers to the process of producing an RNA copy from a DNA template.

“Transformation” refers to a process of introducing an exogenous polynucleic acid molecule (e.g., a DNA construct, a recombinant polynucleic acid molecule) into a cell or protoplast and that exogenous polynucleic acid molecule is incorporated into a chromosome or is capable of autonomous replication.

“Transformed” or “transgenic” refers to a cell, tissue, organ, or organism into which a foreign polynucleic acid, such as a DNA vector or recombinant polynucleic acid molecule. A “transgenic” or “transformed” cell or organism also includes progeny of the cell or organism and progeny produced from a breeding program employing such a “transgenic” plant as a parent in a

-24-

cross and exhibiting an altered phenotype resulting from the presence of the foreign polynucleic acid molecule.

The term "transgene" refers to any polynucleic acid molecule nonnative to a cell or organism transformed into the cell or organism. "Transgene" also encompasses the component parts of a native plant gene modified by insertion of a nonnative polynucleic acid molecule by directed recombination or site specific mutation.

"Transit peptide" or "targeting peptide" molecules, these terms generally refer to peptide molecules that when linked to a protein of interest directs the protein to a particular tissue, cell, subcellular location, or cell organelle. Examples include, but are not limited to, chloroplast transit peptides, nuclear targeting signals, and vacuolar signals. The chloroplast transit peptide is of particular utility in the present invention to direct expression of the EPSPS enzyme to the chloroplast.

The term "translation" refers to the production the corresponding gene product, *i.e.*, a peptide, polypeptide, or protein from a mRNA.

"Vector" refers to a plasmid, cosmid, bacteriophage, or virus that carries foreign DNA into a host organism.

Polynucleotides

Methods of the present invention include designing genes that confer a trait of interest to the plant into which they are introduced. The transgenes of agronomic interest that provide beneficial agronomic traits to crop plants, for example, including, but not limited to genetic elements comprising herbicide resistance (US Patent No. 5,633,435; US Patent No. 5,463,175), increased yield (US Patent No. 5,716,837), insect control (US Patent No. 6,063,597; US Patent No. 6,063,756; US Patent No. 6,093,695; US Patent No. 5,942,664; US Patent No. 6,110,464), fungal disease resistance (US Patent No. 5,516,671; US Patent No. 5,773,696; US Patent No. 6,121,436; and US Patent No. 6,316,407, and US Patent No. 6,506,962), virus resistance (US Patent No. 5,304,730 and US Patent No. 6,013,864), nematode resistance (US Patent No. 6,228,992), bacterial disease resistance (US Patent No. 5,516,671), starch production (US Patent No. 5,750,876 and US Patent No. 6,476,295), modified oils production (US Patent No. 6,444,876), high oil production (US Patent No. 5,608,149 and US Patent No. 6,476,295), modified fatty acid content (US Patent No. 6,537,750), high protein production (US Patent No. 6,380,466), fruit ripening (US Patent No. 5,512,466), enhanced animal and human nutrition (US

-25-

Patent No. 5,985,605 and US Patent No. 6,171,640), biopolymers (US Patent No. 5,958,745 and US Patent Publication No. US20030028917), environmental stress resistance (US Patent No. 6,072,103), pharmaceutical peptides (US Patent No. 6,080,560), improved processing traits (US Patent No. 6,476,295), improved digestibility (US Patent No. 6,531,648) low raffinose (US Patent No. 6,166,292), industrial enzyme production (US Patent No. 5,543,576), improved flavor (US Patent No. 6,011,199), nitrogen fixation (US Patent No. 5,229,114), hybrid seed production (US Patent No. 5,689,041), and biofuel production (US Patent No. 5,998,700).

Herbicides for which transgenic plant tolerance has been demonstrated and the method of the present invention can be applied, include but are not limited to: glyphosate, glufosinate, sulfonylureas, imidazolinones, bromoxynil, delapon, cyclohezanedione, protoporphyrionogen oxidase inhibitors, and isoxasflutole herbicides. Polynucleotide molecules encoding proteins involved in herbicide tolerance are known in the art, and include, but are not limited to a polynucleotide molecule encoding 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS, described in U.S. Patent Nos. 5,627,061, 5,633,435, 6,040,497; Padgett *et al.* *Herbicide Resistant Crops*, Lewis Publishers, 53-85, 1996; and Penaloza-Vazquez, *et al.* *Plant Cell Reports* 14:482-487, 1995; and *aroA* (U.S. Patent No. 5,094,945) for glyphosate tolerance; bromoxynil nitrilase (*Bxn*) for Bromoxynil tolerance (U.S. Patent No. 4,810,648); phytoene desaturase (*crtI*, Misawa *et al.*, (1993) *Plant J.* 4:833-840, and (1994) *Plant J.* 6:481-489); for tolerance to norflurazon, acetohydroxyacid synthase (AHAS, *aka* ALS, Sathasiivan *et al.* *Nucl. Acids Res.* 18:2188-2193, 1990); and the *bar* gene for tolerance to glufosinate and bialaphos (DeBlock, *et al.* *EMBO J.* 6:2513-2519, 1987).

Herbicide tolerance is a desirable phenotype for crop plants. N-phosphonomethylglycine, also known as glyphosate, is a well known herbicide that has activity on a broad spectrum of plant species. Glyphosate is the active ingredient of Roundup® (Monsanto Co.), a safe herbicide having a desirably short half life in the environment. When applied onto a plant surface, glyphosate moves systemically through the plant. Glyphosate is toxic to plants by inhibiting the shikimic acid pathway, which provides a precursor for the synthesis of aromatic amino acids. Specifically, glyphosate affects the conversion of phosphoenolpyruvate and 3-phosphoshikimic acid to 5-enolpyruvyl-3-phosphoshikimic acid by inhibiting the enzyme 5-enolpyruvyl-3-phosphoshikimate synthase (hereinafter referred to as EPSP synthase or EPSPS). For purposes

-26-

of the present invention, the term "glyphosate" should be considered to include any herbicidally effective form of N-phosphonomethylglycine (including any salt thereof) and other forms which result in the production of the glyphosate anion in planta.

Through plant genetic engineering methods, it is possible to produce glyphosate tolerant plants by inserting into the plant genome a DNA molecule that causes the production of higher levels of wild-type EPSPS (Shah *et al.*, Science 233:478-481, 1986). Glyphosate tolerance can also be achieved by the expression of EPSPS variants that have lower affinity for glyphosate and therefore retain their catalytic activity in the presence of glyphosate (U.S. Patent No. 5,633,435). Enzymes that degrade glyphosate in the plant tissues (U.S. Patent No. 5,463,175) are also capable of conferring cellular tolerance to glyphosate. Such genes, therefore, allow for the production of transgenic crops that are tolerant to glyphosate, thereby allowing glyphosate to be used for effective weed control with minimal concern of crop damage. For example, glyphosate tolerance has been genetically engineered into corn (U.S. Patent No. 5,554,798, 6,040,497), wheat (Zhou *et al.* Plant Cell Rep. 15:159-163,1995), soybean (WO 9200377) and canola (WO 9204449).

Variants of the wild-type EPSPS enzyme have been isolated that are glyphosate-resistant as a result of alterations in the EPSPS amino acid coding sequence (Kishore *et al.*, Annu. Rev. Biochem. 57:627-663,1988; Schulz *et al.*, Arch. Microbiol. 137:121-123, 1984; Sost *et al.*, FEBS Lett. 173:238-241, 1984; Kishore *et al.*, In "Biotechnology for Crop Protection" ACS Symposium Series No. 379. eds. Hedlin *et al.*, 37-48,1988). These variants typically have a higher K_i for glyphosate than the wild-type EPSPS enzyme that confers the glyphosate-tolerant phenotype, but these variants are also characterized by a high K_m for PEP that makes the enzyme kinetically less efficient. For example, the apparent K_m for PEP and the apparent K_i for glyphosate for the native EPSPS from *E. coli* are 10 μ M and 0.5 μ M while for a glyphosate-resistant isolate having a single amino acid substitution of an alanine for the glycine at position 96 these values are 220 μ M and 4.0 mM, respectively. US Patent No. 6,040,497 reports that the mutation known as the TIPS mutation (a substitution of isoleucine for threonine at amino acid position 102 and a substitution of serine for proline at amino acid position 106) comprises two mutations that when introduced into the polypeptide sequence of *Zea mays* EPSPS confers glyphosate resistance to the enzyme. Transgenic plants containing this mutant enzyme are tolerant to glyphosate. Identical mutations may be made in glyphosate sensitive EPSPS enzymes from other plant sources to create glyphosate resistant enzymes.

-27-

A variety of native and variant EPSPS enzymes have been expressed in transgenic plants in order to confer glyphosate tolerance (Singh, *et al.*, In "*Biosynthesis and Molecular Regulation of Amino Acids in Plants*", Amer Soc Plant Phys. Pubs., 1992). Examples of some of these EPSPSs include those described and/or isolated in accordance with U.S. Patent No. 4,940,835, U.S. Patent No. 4,971,908, U.S. Patent No. 5,145,783, U.S. Patent No. 5,188,642, U.S. Patent No. 5,310,667, and U.S. Patent No. 5,312,910. They can also be derived from a structurally distinct class of non-homologous EPSPS genes, such as the class II EPSPS genes isolated from *Agrobacterium* sp. strain CP4 as described in U.S. Patent No. 5,633,435 and U.S. Patent No. 5,627,061.

Chloroplast transit peptides (CTPs) are engineered to be fused to the N terminus of the bacterial EPSPS to direct the glyphosate resistant enzymes into the plant chloroplast. In the native plant EPSPS, chloroplast transit peptide regions are contained in the native coding sequence (*e.g.*, CTP2, Klee *et al.*, Mol. Gen. Genet. 210:47-442, 1987). The native CTP may be substituted with a heterologous CTP during construction of a transgene plant expression cassette. Many chloroplast-localized proteins, including EPSPS, are expressed from nuclear genes as precursors and are targeted to the chloroplast by a chloroplast transit peptide (CTP) that is removed during the import steps. Examples of other such chloroplast proteins include the small subunit (SSU) of Ribulose-1,5-bisphosphate carboxylase, Ferredoxin, Ferredoxin oxidoreductase, the light-harvesting complex protein I and protein II, and Thioredoxin F. It has been demonstrated in vivo and in vitro that non-chloroplast proteins may be targeted to the chloroplast by use of protein fusions with a CTP and that a CTP sequence is sufficient to target a protein to the chloroplast. Incorporation of a suitable chloroplast transit peptide, such as, the *Arabidopsis thaliana* EPSPS CTP (Klee *et al.*, Mol. Gen. Genet. 210:437-442 (1987), and the *Petunia hybrida* EPSPS CTP (della-Cioppa *et al.*, Proc. Natl. Acad. Sci. USA 83:6873-6877 (1986) has been shown to target heterologous EPSPS protein sequences to chloroplasts in transgenic plants. The production of glyphosate tolerant plants by expression of a fusion protein comprising an amino-terminal CTP with a glyphosate resistant EPSPS enzyme is well known by those skilled in the art, (U.S. Patent No. 5,627,061, U.S. Patent No. 5,633,435, U.S. Patent No. 5,312,910, EP 0218571, EP 189707, EP 508909, and EP 924299). Those skilled in the art will recognize that various chimeric constructs can be made that utilize the functionality of a particular CTP to import glyphosate resistant EPSPS enzymes into the plant cell chloroplast.

-28-

Modification and changes may be made in the structure of the polynucleotides of the invention and still obtain a molecule that encodes a functional protein or peptide with desirable characteristics. The following is a method based upon substituting the codon(s) of a first polynucleotide to create an equivalent, or even an improved, second-generation artificial polynucleotide, where this new artificial polynucleotide is useful in methods of transgene gene stacking and enhanced expression. It is contemplated that the codon substitutions in the second-generation polynucleotide can in certain instances result in at least one amino acid different from that of the first polynucleotide. The amino acid substitution may provide an improved characteristic to the protein, *e.g.*, a glyphosate resistant EPSP synthase, or it may be a conserved change that does not substantially affect the characteristics of the protein. The method provides for an artificial polynucleotide created by the backtranslation of a polypeptide sequence into a polynucleotide using a codon usage table, followed by steps to enhance characteristics of the artificial polypeptide that make it particularly useful in transgenic plants.

In particular embodiments of the invention, modified polypeptides encoding herbicide resistant proteins are contemplated to be useful for at least one of the following: to confer herbicide tolerance in a transformed or transgenic plant, to improve expression of herbicide resistance genes in plants, for use as selectable markers for introduction of other traits of interest into a plant, and to prevent recombination with a similar endogenous plant gene or existing transgene further allowing gene stacking without gene silencing.

It is known that the genetic code is degenerate. The amino acids and their RNA codon(s) are listed below in Table 1.

TABLE 1. Amino acids and the RNA codons that encode them.

<u>Amino Acid</u>	<u>Codons</u>
Full name; 3 letter code; 1 letter code	
Alanine; Ala; A	GCA GCC GCG GCU
Cysteine; Cys ; C	UGC UGU
Aspartic acid; Asp; D	GAC GAU
Glutamic acid; Glu; E	GAA GAG
Phenylalanine; Phe; F	UUC UUU
Glycine; Gly; G	GGA GGC GGG GGU

-29-

	Histidine; His; H	CAC CAU
	Isoleucine; Ile; I	AUA AUC AUU
	Lysine; Lys; K	AAA AAG
	Leucine ; Leu ; L	UUA UUG CUA CUC CUG CUU
5	Methionine; Met; M	AUG
	Asparagine; Asn; N	AAC AAU
	Proline; Pro; P	CCA CCC CCG CCU
	Glutamine; Gln; Q	CAA CAG
	Arginine; Arg; R	AGA AGG CGA CGC CGG CGU
10	Serine; Ser; S	AGC AGU UCA UCC UCG UCU
	Threonine; Thr; T	ACA ACC ACG ACU
	Valine; Val; V	GUA GUC GUG GUU
	Tryptophan; Trp; W	UGG
	Tyrosine; Tyr; Y	UAC UAU

15

The codons are described in terms of RNA bases, *e.g.* adenine, uracil, guanine and cytosine, it is the mRNA that is directly translated into polypeptides. It is understood that when designing a DNA polynucleotide for use in a construct, the DNA bases would be substituted, *e.g.* thymine instead of uracil.

20

25

30

It is desirable to provide transgenic plants that have multiple agronomically improved phenotypes. Often herbicide tolerance is used as a selectable marker to assist in the production of transgenic plants that may possess additional genes of agronomic importance. The stacking of the transgenes by traditional breeding methods or by retransformation of a first transgenic plant with an additional plant expression cassette may include the introduction of genes or genetic elements that have identical or nearly identical polynucleotide sequence. The progeny containing these stacked genes may be susceptible to loss of gene expression due to gene silencing. The method of the present invention provides a modified polynucleotide molecule that encodes a herbicide resistant protein. The polynucleotide molecules are designed to be sufficiently divergent in polynucleotide sequence from other polynucleotide molecules that encode the same herbicide resistance protein. These molecules can then coexist in the same plant cell without the concern of gene silencing.

-30-

The divergent polynucleotide sequence is created by using a codon usage table built from the known coding sequences of various plant species. For example, codon usage tables for *Arabidopsis thaliana*, *Zea mays*, and *Glycine max* can be used in the method to design the polynucleotides of the present invention. Other codon usage tables from other plants can also be
5 used by those of ordinary skill in the art.

The first step in the method for designing a new artificial polynucleotide molecule that encodes a herbicide tolerance protein is the use of a codon usage table to determine the percent codon usage in a plant species for each amino acid of the herbicide tolerance protein, followed by replacing at least one of every eight contiguous codons with a different codon selected from
10 the codon usage table and adjusting the percent codon usage for each amino acid encoded by the polynucleotide to substantially the same percent codon usage found in the codon usage table. Additional steps can include introducing a translational stop codon in the second and third open reading frame of the new polynucleotide sequence; eliminating some translational start codons in the second and third open reading frames; adjusting the local GC:AT ratio to about 2:1 over a
15 range of about 50 nucleotides; disrupting potential polyadenylation signals or potential intron splice sites; removing at least one restriction enzyme site of six contiguous nucleotides or greater; and comparing the sequence identity of the new artificial polynucleotide to an existing polynucleotide that encodes the same or similar protein so that the sequence identity between the two polynucleotides is not more than 85 percent.

20 A back translation of a protein sequence to a nucleotide sequence maybe performed using a codon usage table, such as those found on Genetics Computer Group (GCG) SeqLab or other DNA analysis programs known to those skilled in the art of DNA analysis or as provided in Tables 2, 3 and 4 of the present invention. The codon usage table for *Arabidopsis thaliana* (Table 2), *Zea mays* (Table 3) and *Glycine max* (Table 4) are examples of tables that can be
25 constructed for plant species, codon usage tables can also be constructed that represent monocot or dicot codon usage.

Table 2. *Arabidopsis thaliana* codon usage table.

Amino Acid	Codon	Number	/1000	Fraction
Gly	GGG	188335.00	10.18	0.16
Gly	GGA	443469.00	23.98	0.37
Gly	GGT	409478.00	22.14	0.34
Gly	GGC	167099.00	9.03	0.14
Glu	GAG	596506.00	32.25	0.48
Glu	GAA	639579.00	34.58	0.52
Asp	GAT	683652.00	36.96	0.68
Asp	GAC	318211.00	17.20	0.32
Val	GTG	320636.00	17.34	0.26
Val	GTA	185889.00	10.05	0.15
Val	GTT	505487.00	27.33	0.41
Val	GTC	235004.00	12.71	0.19
Ala	GCG	162272.00	8.77	0.14
Ala	GCA	323871.00	17.51	0.27
Ala	GCT	521181.00	28.18	0.44
Ala	GCC	189049.00	10.22	0.16
Arg	AGG	202204.00	10.93	0.20
Arg	AGA	348508.00	18.84	0.35
Ser	AGT	260896.00	14.11	0.16
Ser	AGC	206774.00	11.18	0.13
Lys	AAG	605882.00	32.76	0.51
Lys	AAA	573121.00	30.99	0.49
Asn	AAT	418805.00	22.64	0.52
Asn	AAC	385650.00	20.85	0.48
Met	ATG	452482.00	24.46	1.00
Ile	ATA	235528.00	12.73	0.24
Ile	ATT	404070.00	21.85	0.41
Ile	ATC	341584.00	18.47	0.35
Thr	ACG	140880.00	7.62	0.15
Thr	ACA	291436.00	15.76	0.31
Thr	ACT	326366.00	17.65	0.34
Thr	ACC	190135.00	10.28	0.20

-32-

Trp	TGG	231618.00	12.52	1.00
End	TGA	19037.00	1.03	0.43
Cys	TGT	196601.00	10.63	0.60
Cys	TGC	131390.00	7.10	0.40
End	TAG	9034.00	0.49	0.20
End	TAA	16317.00	0.88	0.37
Tyr	TAT	276714.00	14.96	0.52
Tyr	TAC	254890.00	13.78	0.48
Leu	TTG	389368.00	21.05	0.22
Leu	TTA	237547.00	12.84	0.14
Phe	TTT	410976.00	22.22	0.52
Phe	TTC	380505.00	20.57	0.48
Ser	TCG	167804.00	9.07	0.10
Ser	TCA	334881.00	18.11	0.20
Ser	TCT	461774.00	24.97	0.28
Ser	TCC	203174.00	10.99	0.12
Arg	CGG	88712.00	4.80	0.09
Arg	CGA	115857.00	6.26	0.12
Arg	CGT	165276.00	8.94	0.17
Arg	CGC	69006.00	3.73	0.07
Gln	CAG	280077.00	15.14	0.44
Gln	CAA	359922.00	19.46	0.56
His	CAT	256758.00	13.88	0.62
His	CAC	160485.00	8.68	0.38
Leu	CTG	183128.00	9.90	0.11
Leu	CTA	184587.00	9.98	0.11
Leu	CTT	447606.00	24.20	0.26
Leu	CTC	294275.00	15.91	0.17
Pro	CCG	155222.00	8.39	0.17
Pro	CCA	298880.00	16.16	0.33
Pro	CCT	342406.00	18.51	0.38
Pro	CCC	97639.00	5.28	0.11

Table 3. *Zea mays* codon usage table

Amino Acid	Codon	Number	/1000	Fraction
------------	-------	--------	-------	----------

-33-

Amino Acid	Codon	Number	/1000	Fraction
Gly	GGG	8069.00	15.19	0.21
Gly	GGA	7100.00	13.37	0.18
Gly	GGT	7871.00	14.82	0.20
Gly	GGC	15904.00	29.94	0.41
Glu	GAG	22129.00	41.67	0.68
Glu	GAA	10298.00	19.39	0.32
Asp	GAT	11996.00	22.59	0.41
Asp	GAC	17045.00	32.09	0.59
Val	GTG	13873.00	26.12	0.38
Val	GTA	3230.00	6.08	0.09
Val	GTT	8261.00	15.55	0.23
Val	GTC	11330.00	21.33	0.31
Ala	GCG	11778.00	22.18	0.24
Ala	GCA	8640.00	16.27	0.18
Ala	GCT	11940.00	22.48	0.24
Ala	GCC	16768.00	31.57	0.34
Arg	AGG	7937.00	14.94	0.27
Arg	AGA	4356.00	8.20	0.15
Ser	AGT	3877.00	7.30	0.10
Ser	AGC	8653.00	16.29	0.23
Lys	AAG	22367.00	42.11	0.74
Lys	AAA	7708.00	14.51	0.26
Asn	AAT	6997.00	13.17	0.36
Asn	AAC	12236.00	23.04	0.64
Met	ATG	12841.00	24.18	1.00
Ile	ATA	3997.00	7.53	0.16
Ile	ATT	7457.00	14.04	0.31
Ile	ATC	12925.00	24.34	0.53
Thr	ACG	5665.00	10.67	0.22
Thr	ACA	5408.00	10.18	0.21
Thr	ACT	5774.00	10.87	0.22
Thr	ACC	9256.00	17.43	0.35
Trp	TGG	6695.00	12.61	1.00

-34-

Amino Acid	Codon	Number	/1000	Fraction
End	TGA	591.00	1.11	0.45
Cys	TGT	2762.00	5.20	0.30
Cys	TGC	6378.00	12.01	0.70
End	TAG	411.00	0.77	0.32
End	TAA	299.00	0.56	0.23
Tyr	TAT	4822.00	9.08	0.31
Tyr	TAC	10546.00	19.86	0.69
Leu	TTG	6677.00	12.57	0.14
Leu	TTA	2784.00	5.24	0.06
Phe	TTT	6316.00	11.89	0.32
Phe	TTC	13362.00	25.16	0.68
Ser	TCG	5556.00	10.46	0.14
Ser	TCA	5569.00	10.49	0.15
Ser	TCT	6149.00	11.58	0.16
Ser	TCC	8589.00	16.17	0.22
Arg	CGG	4746.00	8.94	0.16
Arg	CGA	2195.00	4.13	0.07
Arg	CGT	3113.00	5.86	0.10
Arg	CGC	7374.00	13.88	0.25
Gln	CAG	13284.00	25.01	0.64
Gln	CAA	7632.00	14.37	0.36
His	CAT	5003.00	9.42	0.39
His	CAC	7669.00	14.44	0.61
Leu	CTG	13327.00	25.09	0.28
Leu	CTA	3785.00	7.13	0.08
Leu	CTT	8238.00	15.51	0.17
Leu	CTC	12942.00	24.37	0.27
Pro	CCG	8274.00	15.58	0.27
Pro	CCA	7845.00	14.77	0.26
Pro	CCT	7129.00	13.42	0.23
Pro	CCC	7364.00	13.87	0.24

Table 4. *Glycine max* codon usage table

Amino Acid	Codon	Number	/1000	Fraction
------------	-------	--------	-------	----------

-35-

Amino Acid	Codon	Number	/1000	Fraction
Gly	GGG	3097.00	12.82	0.18
Gly	GGA	5434.00	22.49	0.32
Gly	GGT	5248.00	21.72	0.31
Gly	GGC	3339.00	13.82	0.20
Glu	GAG	8296.00	34.33	0.50
Glu	GAA	8194.00	33.91	0.50
Asp	GAT	7955.00	32.92	0.62
Asp	GAC	4931.00	20.40	0.38
Val	GTG	5342.00	22.11	0.32
Val	GTA	1768.00	7.32	0.11
Val	GTT	6455.00	26.71	0.39
Val	GTC	2971.00	12.29	0.18
Ala	GCG	1470.00	6.08	0.08
Ala	GCA	5421.00	22.43	0.31
Ala	GCT	6796.00	28.12	0.38
Ala	GCC	4042.00	16.73	0.23
Arg	AGG	3218.00	13.32	0.28
Arg	AGA	3459.00	14.31	0.30
Ser	AGT	2935.00	12.15	0.17
Ser	AGC	2640.00	10.92	0.15
Lys	AAG	9052.00	37.46	0.59
Lys	AAA	6370.00	26.36	0.41
Asn	AAT	5132.00	21.24	0.48
Asn	AAC	5524.00	22.86	0.52
Met	ATG	5404.00	22.36	1.00
Ile	ATA	3086.00	12.77	0.23
Ile	ATT	6275.00	25.97	0.47
Ile	ATC	3981.00	16.47	0.30
Thr	ACG	1006.00	4.16	0.08
Thr	ACA	3601.00	14.90	0.29
Thr	ACT	4231.00	17.51	0.34
Thr	ACC	3562.00	14.74	0.29
Trp	TGG	2866.00	11.86	1.00

-36-

Amino Acid	Codon	Number	/1000	Fraction
End	TGA	221.00	0.91	0.36
Cys	TGT	1748.00	7.23	0.49
Cys	TGC	1821.00	7.54	0.51
End	TAG	143.00	0.59	0.23
End	TAA	256.00	1.06	0.41
Tyr	TAT	3808.00	15.76	0.51
Tyr	TAC	3667.00	15.17	0.49
Leu	TTG	5343.00	22.11	0.24
Leu	TTA	2030.00	8.40	0.09
Phe	TTT	4964.00	20.54	0.49
Phe	TTC	5067.00	20.97	0.51
Ser	TCG	1107.00	4.58	0.06
Ser	TCA	3590.00	14.86	0.21
Ser	TCT	4238.00	17.54	0.24
Ser	TCC	2949.00	12.20	0.17
Arg	CGG	683.00	2.83	0.06
Arg	CGA	964.00	3.99	0.08
Arg	CGT	1697.00	7.02	0.15
Arg	CGC	1538.00	6.36	0.13
Gln	CAG	4147.00	17.16	0.46
Gln	CAA	4964.00	20.54	0.54
His	CAT	3254.00	13.47	0.55
His	CAC	2630.00	10.88	0.45
Leu	CTG	2900.00	12.00	0.13
Leu	CTA	1962.00	8.12	0.09
Leu	CTT	5676.00	23.49	0.26
Leu	CTC	4053.00	16.77	0.18
Pro	CCG	1022.00	4.23	0.08
Pro	CCA	4875.00	20.17	0.37
Pro	CCT	4794.00	19.84	0.36
Pro	CCC	2445.00	10.12	0.19

Codon usage tables are well known in the art and can be found in gene databases *e.g.*, Genbank database. The Codon Usage Database is an extended WWW version of CUTG (Codon Usage Tabulated from Genbank). The frequency of codon usage in each organism is made

-37-

searchable through this World Wide Web site (Nakamura *et al.* Nucleic Acids Res.28:292, 2000).

In various embodiments of the invention, the steps may be performed in any order or simultaneously. Any or all of the steps may be performed in the design of an artificial polynucleotide of the invention. Each step is described in detail below.

Different codons for a particular amino acid should be distributed throughout the polynucleotide based on approximate percentage codon usage for particular species from a codon usage table. Local cluster of identical codons should be avoided. At least one codon is substituted for every eight contiguous codons to provide sufficient divergence of polynucleotide sequences that encode identical or similar proteins. Except where specifically desired, *e.g.* to provide a herbicide tolerant enzyme, the encoded protein remains unchanged by substituting one codon for another codon that is translated to the same amino acid as listed in Table 1.

In embodiments of the present invention, corrections are made to the local GC:AT ratio of a polynucleotide by adjusting local GC:AT ratio to be about the same ratio as the full length polynucleotide, but not higher than 2X over a range of about 50 contiguous nucleotides of the polynucleotide molecule. The range of GC:AT ratios of a polynucleotide using codon usage tables from dicot plants should be from about 0.9 to about 1.3, and for monocot plants from about 1.2 to about 1.7. The local GC:AT ratio may be important in maintenance of appropriate secondary structure of RNA. Regions comprising many consecutive A+T bases or G+C bases are predicted to have a higher likelihood to form hairpin structures due to self-complementarity. Therefore, replacement with a different codon would reduce the likelihood of self-complementary secondary structure formation, which is known to reduce transcription and/or translation in some organisms. In most cases, the adverse effects may be minimized by using polynucleotide molecules that do not contain more than five consecutive A+T or G+C. The maximum length of local GC track (without any AT nucleotide) should be no longer than 10 nucleotides. Therefore codons encoding Gly, Ala, Arg, Ser, and Pro rich proteins can be substituted to prevent long clusters of GC nucleotides. The listed GC rich codons may be used in combination with the AT rich codons for amino acids Lys, Asn, Ile, Tyr, Leu, Phe and vice versa to correct local GC:AT ratio.

A sequence identity check using nucleotide sequence alignment tools such as GAP program (GCG, Madison, WI) can be done immediately after back translation to insure that the generated sequence has appropriate degree of sequence diversity. Contiguous polynucleotide

-38-

sequence longer than 23 nucleotides having one hundred percent sequence identity should be eliminated by making codon substitutions in these lengths of sequence.

The translational start codons (ATG from the DNA, AUG in the mRNA) present in the second reading frame (frame "b"), the third reading frame (frame "c"), and the reverse reading frames (frame "d", "e", "f"). The second and third frame start codons may initiate translation, however much less efficiently than the first. Therefore, if one or two AUG are found near the 5' end of an mRNA molecule reside in frame "b" or "c" it would be beneficial to eliminate them in a polynucleotide region that contains at least the first three Met codons in frame "a". Also, if protein sequence does not have more than one Met in frame "a", then eliminate as many as possible from the "b" or "c" forward frames. To perform this, for example, the codons for amino acids, Asp, Asn, Tyr, His in the protein of interest followed by any of the amino acids: Gly, Glu, Asp, Val, or Ala, can be substituted to eliminate a start codon in the second frame. The sequence GATGGG encodes the amino acids Asp-Gly and forms an ATG in the reading frame "b". When the sequence is modified to GACGGG, the ATG start is eliminated and the sequence still encodes Asp-Gly. A similar strategy is used to eliminate start codons in the reading frame "c". The combination of an amino acid selected from the group of Gly, Glu, Val, Ala, Arg, Lys, Ile, Thr, Cys, Tyr, Leu, Ser, His or Pro followed by Trp can result in formation on ATG in third reading frame of the gene. In this situation, the first codon can be changed to have a nucleotide other than A in third position.

The elimination of ATG codon in the complementary DNA strand of the gene in alternate frames ("d", "e", and/or "f") without changing amino acid sequence of the protein can be accomplished in a similar manner. This modification reduces the probability of translation even if the transgene is integrated into a plant genome in an orientation that allows transcription of the reverse complement mRNA from a native plant promoter. Translation from any reverse reading frame can be minimized by introduction of a stop codon in all three reverse reading frames as described below.

The creation of stop codons to all three frames of the complementary DNA can be accomplished as follows. The Leu (TTA and CTA) and Ser codon (TCA) produce three different stop codons in reverse complement strand. If those amino acids can be found at the C terminus of the protein of interest, their codons may used to generate stops in the complementary strand in the reading frame "d". To generate a stop codon in the reading frame "e" of complementary strand, find amino acids Ala, Arg, Asn, Asp, Cys, Gly, His, Ile, Leu Phe, Pro, Ser, Thr, Tyr, or

-39-

Val followed by amino acids Gln, His or Tyr the protein of interest. For example, polynucleotide sequence of GCCCAG that encode for amino acids Ala-His can be modified to GCTCAC. The complementary sequence, *GTGAGC*, will have now TGA stop codon shown in italics. When the protein of interest has a Ala, Ile, Leu, Phe, Pro, Ser, Thr or Val followed by an Arg, Asn, Ile, Lys, Met, or Ser the reading frame in the complementary strand can be modified to have a stop codon in the reading frame “f” of the complementary strand. The polynucleotide sequence ATATCT for Ile and Ser can be modified to ATCAGT to generate stop codon in complementary strand as shown in italics, *ACTGAT*. The combination of codons for Phe followed by any of the codons for amino acids Asn, Ile, Lys, Met or Thr will always generate stop codon in complementary strand frame “e” or “f”.

To create a stop codon in the forward reading frame “b”, the reading frame a must end on nucleotides TA or TG. Search the protein of interest for the amino acids Ile, Leu, Met or Val in combination with any of the following amino acids: Ala, Arg, Asn Asp, Glu, Gly, Ile, Lys, Met, Ser, Thr or Val. For example, if the polynucleotide sequence encoding the amino acids Met-Ser is ATGTCT, it can be modified to *ATGAGT* to produce a TGA stop codon in second reading frame.

To be able to create a stop codon to the reading frame “c”, the reading frame “a” must have the nucleotide T in third position and next codon must start from AA, AG or GA. To find suitable codons to modify, search the protein of interest for any of the amino acids: Ala, Asn, Asp, Arg, Cys, Gly, His, Ile, Leu Phe, Pro, Ser, Thr, Tyr or Val follow by any of the following amino acids: Arg, Asn, Asp, Glu, Lys or Ser. For example, if the nucleotide sequence for amino acids Gly-Glu is GGAGAG, the sequence can be modified to *GGTGAG* to create a TGA stop codon in the third reading frame.

Another useful modification in artificial polynucleotide design methods of the present invention is to eliminate unwanted restriction sites and other specific sequence patterns. Restriction sites may interfere with future gene cloning and manipulations. For example, some restriction sites commonly used in gene cloning include, but are not limited, to the Type II restriction enzymes with 6 or more non-N bases listed in Table 5 below which is an excerpt from the New England Biolabs, Inc. (Beverly, MA, USA) restriction endonuclease database. The search for restriction enzyme recognition sites can be done using Map function application found in GCG SeqLab or a similar application contained in other DNA analysis programs known to those skilled in the art of DNA analysis. The restriction enzymes can be also added to the

-40-

sequence to facilitate cloning. For example, The ClaI restriction site is placed in CP4EPSPS version AT (SEQ ID NO:17) and ZM (SEQ ID NO:18) to generate recombinant sequences by fragment exchange and to facilitate gene synthesis using nucleotide fragments that can be assemble to the whole gene. The transit peptide CTP2 polynucleotide sequence (SEQ ID NO:12) is connected with CP4EPSPS by SphI restriction site to facilitate substitution of CTP2 with different nucleotide versions of CTP2 (SEQ ID NO:13, SEQ ID NO:14) or polynucleotides encoding different chloroplast transit peptides. For example, in the rice EPSPS, the NgaMIV restriction site is preserved at about nucleotide position 205 in all artificial versions to facilitate chloroplast transit peptide coding region exchange. Also, for soybean EPSPS the polynucleotide sequence for the chloroplast transit peptide is separated from the mature peptide by the restriction site for SacII endonuclease.

It is understood that modification of endonuclease restriction sites is not required, but is useful for further manipulation of the DNA molecules. Table 5 provides a list of restriction endonucleases, those of particular interest to the present invention are marked with an asterisk. Other endonuclease restriction sites desirable for elimination or addition to an artificial polynucleotide of the present invention will be apparent to those of ordinary skill in the art and are not limited to those listed in Table 5.

TABLE 5. Restriction enzymes recognition sequences

Enzymes	Recognition Sequence	Enzymes	Recognition Sequence
AatII	G_ACGT^C	AsuII	TT^CG_AA
*AccI	GT^MK_AC	AvaI	C^YCGR_G
Acc65I	G^GTAC_C	AvaIII	ATGCAT
AceII	G_CTAG^C -	AvrII	C^CTAG_G
AclI	AA^CG_TT	BalI	TGG^CCA
AcyI	GR^CG_YC	*BamHI	G^GATC_C
AfeI	AGC^GCT	BanI	G^GYRC_C
AflII	C^TTAA_G	BanII	G_RGCY^C
*AflIII	A^CRYG_T	*BbeI	G_GCGC^C
AgeI	A^CCGG_T	BbvCI	CC^TCA_GC
AhaIII	TTT^AAA	*BclI	T^GATC_A
ApaI	G_GGCC^C	BetI	W^CCGG_W
ApaLI	G^TGCA_C	BfrBI	ATG^CAT
ApoI	R^AATT_Y	*BglII	A^GATC_T
AscI	GG^CGCG_CC	BloHII	CTGCA^G
AseI	AT^TA_AT	BlpI	GC^TNA_GC
AsiSI	GCG_AT^CGC	Bme1580I	G_KGCM^C

-41-

Enzymes	Recognition Sequence	Enzymes	Recognition Sequence
BmgI	GKGCCC	Eco47III	AGC^GCT
Bpu10I	CC^TNA_GC	EcoNI	CCTNN^N_NNAGG
BsaI	GGTCTCN^NNNN_	EcoO109I	RG^GNC_CY
BsaAI	YAC^GTR	*EcoRI	G^AATT_C
BsaHI	GR^CG_YC	*EcoRV	GAT^ATC
BsaWI	W^CCGG_W	EspI	GC^TNA_GC
BsbI	CAACAC	*FseI	GG_CCGG^CC
BsePI	G^CGCG_C	*FspI	TGC^GCA
BseSI	G_KGCM^C	FspAI	RTGC^GCAY
BsiI	C^ACGA_G	GdiII	C^GGCC_R
BsiEI	CG_RY^CG	HaeI	WGG^CCW
BsiWI	C^GTAC_G	HaeII	R_GCGC^Y
BsmI	GAATG_CN^	HgiAI	G_WGCW^C
Bsp1286I	G_DGCH^C	HgiCI	G^GYRC_C
Bsp1407I	T^GTAC_A	HgiJII	G_RGCY^C
BspEI	T^CCGG_A	*HincII	GTY^RAC
BspGI	CTGGAC	HindII	GTY^RAC
BspHI	T^CATG_A	*HindIII	A^AGCT_T
BspLU11I	A^CATG_T	*HpaI	GTT^AAC
BspMII	T^CCGG_A	KasI	G^GCGC_C
BsrBI	CCG^CTC	*KpnI	G_GTAC^C
BsrDI	GCAATG_NN^	LpnI	RGC^GCY
BsrFI	R^CCGG_Y	McrI	CG_RY^CG
BsrGI	T^GTAC_A	MfeI	C^AATT_G
BssHII	G^CGCG_C	*MluI	A^CGCG_T
BssSI	C^ACGA_G	MscI	TGG^CCA
BstAPI	GCAN_NNN^NTGC	MspA1I	CMG^CKG
BstBI	TT^CG_AA	MstI	TGC^GCA
BstEII	G^GTNAC_C	NaeI	GCC^GGC
BstXI	CCAN_NNNN^NTGG	NarI	GG^CG_CC
BstYI	R^GATC_Y	*NcoI	C^CATG_G
BstZ17I	GTA^TAC	*NdeI	CA^TA_TG
Bsu36I	CC^TNA_GG	*NgoMIV	G^CCGG_C
BtgI	C^CRYG_G	*NheI	G^CTAG_C
BtrI	CAC^GTC	Nli3877I	C_YCGR^G
BtsI	GCAGTG_NN^	*NotI	GC^GGCC_GC
CfrI	Y^GGCC_R	*NruI	TCG^CGA
Cfr10I	R^CCGG_Y	*NsiI	A_TGCA^T
*ClaI	AT^CG_AT	NspI	R_CATG^Y
DraI	TTT^AAA	NspBII	CMG^CKG
DraII	RG^GNC_CY	*PacI	TTA_AT^TAA
DrdII	GAACCA	*PciI	A^CATG_T
DsaI	C^CRYG_G	Pfl1108I	TCGTAG
EaeI	Y^GGCC_R	*PflMI	CCAN_NNN^NTGG
EagI	C^GGCC_G	PmaCI	CAC^GTG
Ecl136II	GAG^CTC	PmeI	GTTT^AAAC

-42-

Enzymes	Recognition Sequence	Enzymes	Recognition Sequence
PmlI	CAC^GTG	*VspI	AT^TA_AT
Ppu10I	A^TGCA_T	*XbaI	T^CTAG_A
*PpuMI	RG^GWC_CY	*XhoI	C^TCGA_G
PshAI	GACNN^NNGTC	XhoII	R^GATC_Y
PsiI	TTA^TAA	*XmaI	C^CCGG_G
*PspOMI	G^GGCC_C	XmaIII	C^GGCC_G
PssI	RG_GNC^CY	XmnI	GAANN^NNTTC
*PstI	C_TGCA^G	ZraI	GAC^GTC
*PvuI	CG_AT^CG		
*PvuII	CAG^CTG		
RsrII	CG^GWC_CG		
*SacI	G_AGCT^C		
*SacII	CC_GC^GG		
*Sall	G^TCGA_C		
SanDI	GG^GWC_CC		
SapI	GCTCTTCN^NNN_		
SauI	CC^TNA_GG		
SbfI	CC_TGCA^GG		
*ScaI	AGT^ACT		
SciI	CTC^GAG		
SduI	G_DGCH^C		
SexAI	A^CCWGG_T		
SfcI	C^TRYA_G		
SfeI	C^TRYA_G		
SfiI	GGCCN_NNN^NGGCC		
SfoI	GGC^GCC		
Sgfi	GCG_AT^CGC		
SgrAI	CR^CCGG_YG		
*SmaI	CCC^GGG		
SmlI	C^TYRA_G		
SnaI	GTATAC		
*SnaBI	TAC^GTA		
*SpeI	A^CTAG_T		
*SphI	G_CATG^C		
SplI	C^GTAC_G		
SrfI	GCCC^GGGC		
Sse232I	CG^CCGG_CG		
Sse8387I	CC_TGCA^GG		
Sse8647I	AG^GWC_CT		
*SspI	AAT^ATT		
*StuI	AGG^CCT		
*StyI	C^CWWG_G		
*Swal	ATTT^AAAT		
TatI	W^GTAC_W		
UbaMI	TCCNGGA		
UbaPI	CGAACG		

- 43 -

A pattern search may be performed to find potential destabilizing sequences and polyadenylation sites and then disrupt or eliminate them as described in US Patent No. 5,500,365. Certain long stretches of AT rich regions, *e.g.* the sequence motif ATTTA (or AUUUA, as it appears in RNA) have been implicated as a destabilizing sequence in mammalian cell mRNA (Shaw and Kamen, Cell 46:659-667, 1986). Many short lived mRNAs have A+T rich 3' untranslated regions, and these regions often have the ATTTA sequence, sometimes present in multiple copies or as multimers (*e.g.*, ATTTATTTA . . .). Shaw and Kamen showed that the transfer of the 3' end of an unstable mRNA to a stable RNA (globin or VA1) decreased the stable RNA's halflife dramatically. They further showed that a pentamer of ATTTA had a profound destabilizing effect on a stable message, and that this signal could exert its effect whether it is located at the 3' end or within the coding sequence. However, the number of ATTTA sequences and/or the sequence context in which they occur also appear to be important in determining whether they function as destabilizing sequences. They also showed that a trimer of ATTTA had much less effect than a pentamer on mRNA stability and a dimer or a monomer had no effect on stability. Note that multimers of ATTTA such as a pentamer automatically create an A+T rich region. In other unstable mRNAs, the ATTTA sequence may be present in only a single copy, but it is often contained in an A+T rich region. A repeat of 11 AUUUA pentamers has been shown to target reporter transcripts for rapid degradation in plants (Ohme-Takagi *et al.*, Proc. Nat. Acad. Sci. USA 90, 11811-11815, 1993). ATTTA sequence can be formed by combination of codons for amino acid Ile (ATT) and Tyr (TAT) as shown *ATTTAT*. Another example could be codons that end on AT as in Asn, Asp, His or Tyr, followed by TTA codon for Leu (*e.g.* *AATTTA*). Also codon for Phe (UUU) when placed between codons that ends on A and starts on A will form ATTTA motif. To eliminate this motif usually single nucleotide change is sufficient as in example shown: *GCATTTAGC* change to *GCATTCAGC* or *GCCTTTAGC*. All three polynucleotide shown code for Ala-Phe-Arg.

More cis-acting sequences that target transcript for rapid turnover in plants and in other system has been identified (Abler and Green, Plant Mol. Biol. 32:63-78, 1997). Those include the DST element that consist three highly conserved subdomains separated by two variable regions found downstream of the stop codon of SAUR transcripts (Newmaan *et al.*, Plant Cell 5: 701-714, 1993). The DST conserved sequence consist of *GGAG(N₅)CATAGATTG(N₇)CATTTTGTAT*, where highly conserved residues are shown in italics type. The second and third subdomains of DST elements contain residues that are

-44-

invariant among DST elements and are termed ATAGAT and GTA, respectively. Both of those subdomains are necessary for DST function. New artificial polynucleotide sequences are screened for the presence of conserved motifs of DST elements GGAG, ATAGATT, CATTT and CATTTTGTAT. Those sequences are eliminated by base substitutions of codons preserving
5 protein sequence encoded by the polynucleotide. The DST sequence motifs GGAG, ATAGAT, CATTT and GTAT that appeared in clusters or patterns similar to the conserved DST sequence are also eliminated by base substitutions.

Polynucleotide sequences that may possibly function as polyadenylation sites are eliminated in the new polynucleotide design (U.S. Patent No. 5,500,365). These polyadenylation
10 signals may not act as proper polyA sites, but instead function as aberrant sites that give rise to unstable mRNAs.

The addition of a polyadenylate string to the 3' end of a mRNA is common to most eukaryotic mRNAs. Contained within this mRNA transcript are signals for polyadenylation and proper 3' end formation. This processing at the 3' end involves cleavage of the mRNA and
15 addition of polyA to the mature 3' end. By searching for consensus sequences near the polyA tract in both plant and animal mRNAs, it has been possible to identify consensus sequences that apparently are involved in polyA addition and 3' end cleavage. The same consensus sequences seem to be important to both of these processes. These signals are typically a variation on the sequence AATAAA. In animal cells, some variants of this sequence that are functional have
20 been identified; in plant cells there seems to be an extended range of functional sequences (Dean *et al.*, Nucl Acid Res., 14:2229-2240, 1986; Hunt, Annu Rev. Plant Physiol. Plant Mol. Biol 45:47-60, 1994; Rothine, Plant Mol. Biol. 32:43-61, 1996). All of these consensus sequences are variations on AATAAA, therefore, they all are A+T rich sequences.

Typically, to obtain sufficient expression of modified transgenes in plants, existing
25 structural polynucleotide coding sequence ("structural gene") that encodes for the protein of interest is modified by removal of ATTTA sequences and putative polyadenylation signals by site directed mutagenesis of the DNA comprising the structural gene. Substantially all of the known polyadenylation signals and ATTTA sequences are removed in the modified polynucleotide, although enhanced expression levels are often observed with only removal of
30 some of the above identified polyadenylation signal sequences. Alternately, if an artificial polynucleotide is prepared that encodes for the subject protein, codons are selected to avoid the

-45-

ATTTA sequence and putative polyadenylation signals. For purposes of the present invention putative polyadenylation signals include, but are not necessarily limited to, AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA and CATAAA.

5 The selected DNA sequence is scanned to identify regions with greater than four consecutive adenine (A) or thymine (T) nucleotides. The A+T regions are scanned for potential plant polyadenylation signals. Although the absence of five or more consecutive A or T nucleotides eliminates most plant polyadenylation signals, if there are more than one of the minor polyadenylation signals identified within ten nucleotides of each other, then the nucleotide
10 sequence of this region is altered to remove these signals while maintaining the original encoded amino acid sequence.

The next step is to consider the about 15 to about 30 or so nucleotide residues surrounding the A+T rich region. If the A+T content of the surrounding region is less than 80%, the region should be examined for polyadenylation signals. Alteration of the region based on
15 polyadenylation signals is dependent upon (1) the number of polyadenylation signals present and (2) presence of a major plant polyadenylation signal. The polyadenylation signals are removed by base substitution of the DNA sequence in the context of codon replacement.

Two additional patterns not identified in US Patent No. 5,500,365, are searched for and eliminated in embodiments of the present invention. The sequences AGGTAA and GCAGGT
20 are consensus sequences for intron 5' and 3' splice sites, respectively, in monocot plants and dicot plants. Only GT of the 5' splice site and the AG in the 3' splice site are required to be an exact match. However, when conducting a search for these consensus sequences, no mismatch is allowed for each base.

After each step sequence mapping is done using MAP program from GCG to determine
25 location of the open reading frames and identify sequence patterns that further need to be modified. The final step would be to perform sequence identity analysis using for example the GAP program from GCG package to determine degree of sequence divergence and percent identity.

Polypeptides

Generally, the translated protein of the artificial polynucleotide will have the same amino acid sequence as the protein translated from the unmodified coding region. However, the substitution of codons that encode for amino acids that provide a functional homologue of the protein is an aspect of the invention. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the disclosed compositions by making changes in the corresponding DNA sequences that encode the peptides in which the peptides shown no appreciable loss of their biological utility or activity.

A further aspect of the invention comprises functional homologues, which differ in one or more amino acids from those of a polypeptide provided herein as the result of one or more conservative amino acid substitutions. It is well known in the art that one or more amino acids in a native sequence can be substituted with at least one other amino acid, the charge and polarity of which are similar to that of the native amino acid, resulting in a silent change. For instance, valine is a conservative substitute for alanine and threonine is a conservative substitute for serine. Conservative substitutions for an amino acid within the native polypeptide sequence can be selected from other members of the class to which the naturally occurring amino acid belongs. Amino acids can be divided into the following four groups: (1) acidic amino acids, (2) basic amino acids, (3) neutral polar amino acids, and (4) neutral nonpolar amino acids. Representative amino acids within these various groups include, but are not limited to: (1) acidic (negatively charged) amino acids such as aspartic acid and glutamic acid; (2) basic (positively charged) amino acids such as arginine, histidine, and lysine; (3) neutral polar amino acids such as glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; and (4) neutral nonpolar (hydrophobic) amino acids such as alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine. Conserved substitutes for an amino acid within a native amino acid sequence can be selected from other members of the group to which the naturally occurring

-47-

amino acid belongs. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Naturally conservative amino acids substitution groups are: valine-leucine, valine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, aspartic acid-glutamic acid, and asparagine-glutamine.

DNA Constructs

Exogenous genetic material may be transferred into a plant by the use of a DNA construct designed for such a purpose by methods that utilize *Agrobacterium*, particle bombardment or other methods known to those skilled in the art. Design of such a DNA construct is generally within the skill of the art (*Plant Molecular Biology: A Laboratory Manual*, eds. Clark, Springer, New York (1997)). Examples of such plants in to which exogenous genetic material may be transferred, include, without limitation, alfalfa, *Arabidopsis*, barley, Brassica, broccoli, cabbage, citrus, cotton, garlic, oat, oilseed rape, onion, canola, flax, maize, an ornamental annual and ornamental perennial plant, pea, peanut, pepper, potato, rice, rye, sorghum, soybean, strawberry, sugarcane, sugar beet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, oil palm, *Phaseolus*, trees, shrubs, vines, etc. It is well known that agronomically important plants comprise genotypes, varieties and cultivars, and that the methods and compositions of the present invention can be tested in these plants by those of ordinary skill in the art of plant molecular biology and plant breeding.

A large number of isolated DNA promoter molecules that are active as a genetic element of a transgene in plant cells have been described. These include the nopaline synthase (P-nos) promoter (Ebert *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 84:5745-5749, 1987), the octopine synthase (P-ocs) promoter, which are carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*, the caulimovirus promoters, such as the cauliflower mosaic virus (CaMV) 19S promoter (Lawton *et al.*, Plant Mol. Biol.

-48-

9:315-324, 1987), and the CaMV 35S promoter (Odell *et al.*, Nature 313:810-812, 1985), the figwort mosaic virus 35S promoter (U.S. Patent No. 6,018,100), the light-inducible promoter from the small subunit of ribulose-1,5-bis-phosphate carboxylase (ssRUBISCO), the Adh promoter (Walker *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 84:6624-6628, 1987), the sucrose synthase promoter (Yang *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 87:4144-4148, 1990), the R gene complex promoter (Chandler *et al.*, Plant Cell 1:1175-1183, 1989), and the chlorophyll a/b binding protein gene promoter, etc.

A variety of promoters specifically active in vegetative tissues, such as leaves, stems, roots and tubers, can be used to express the nucleic acid molecules of the present invention.

Examples of tuber-specific promoters include, but are not limited to the class I and II patatin promoters (Bevan *et al.*, EMBO J. 8: 1899-1906, 1986); Koster-Topfer *et al.*, Mol Gen Genet. 219:390-396, 1989; Mignery *et al.*, Gene 62:27-44, 1988; Jefferson *et al.*, Plant Mol. Biol. 14: 995-1006, 1990, the promoter for the potato tuber ADPGPP genes, both the large and small subunits; the sucrose synthase promoter (Salanoubat and Belliard, Gene 60:47-56, 1987), Salanoubat and Belliard, Gene 84:181-185, 1989; and the promoter for the major tuber proteins including the 22 kd protein complexes and proteinase inhibitors (Hannapel, Plant Physiol. 101:703-704, 1993). Examples of leaf-specific promoters include but are not limited to the ribulose biphosphate carboxylase (RbcS or RuBISCO) promoters (see, *e.g.*, Matsuoka *et al.*, Plant J. 6:311-319, 1994), the light harvesting chlorophyll a/b binding protein gene promoter (see, *e.g.*, Shiina *et al.*, Plant Physiol. 115:477-483, 1997; Casal *et al.*, Plant Physiol. 116:1533-1538, 1998); and the *Arabidopsis thaliana* myb-related gene promoter (Atmyb5) (Li *et al.*, FEBS Lett. 379:117-121, 1996). Examples of root-specific promoter include but are not limited to the promoter for the acid chitinase gene (Samac *et al.*, Plant Mol. Biol. 25:587-596, 1994); the root specific subdomains of the CaMV35S promoter that have been identified (Lam *et al.*, Proc. Natl. Acad. Sci. (U.S.A.)

-49-

86:7890-7894, 1989); the ORF13 promoter from *Agrobacterium rhizogenes* which exhibits high activity in roots (Hansen *et al.*, Mol. Gen. Genet. 254:337-343, 1997); the
 5 promoter for the tobacco root-specific gene RB7 (US Patent 5,750,386; Yamamoto *et al.*, Plant Cell 3:371-382, 1991); and the root cell specific promoters reported by Conkling *et al.* (Conkling *et al.*, Plant Physiol. 93:1203-1211, 1990), and the POX1 (Pox1, pox1) promoter (Hertig, *et al.* Plant Mol. Biol. 16:171, 1991).

Another class of useful vegetative tissue-specific promoters are meristematic (root tip and
 10 shoot apex) promoters. For example, the "SHOOTMERISTEMLESS" and "SCARECROW" promoters, which are active in the developing shoot or root apical meristems (Di Laurenzio *et al.*, Cell 86:423-433, 1996; Long, Nature 379:66-69, 1996); can be used.

Another example of a useful promoter is that which controls the expression of 3-hydroxy-3-methylglutaryl coenzyme A reductase HMG2 gene, whose
 15 expression is restricted to meristematic and floral (secretory zone of the stigma, mature pollen grains, gynoecium vascular tissue, and fertilized ovules) tissues (see, e.g., Enjuto *et al.*, Plant Cell. 7:517-527, 1995). Also another example of

a useful promoter is that which controls the expression of knl-related genes from maize and other species which show meristem-specific expression (see, e.g., Granger *et al.*, Plant Mol. Biol.
 20 31:373-378, 1996; Kerstetter *et al.*, Plant Cell 6:1877-1887, 1994; Hake *et al.*, Philos. Trans. R. Soc. Lond. B. Biol. Sci. 350:45-51, 1995).

Another example of a meristematic promoter is the *Arabidopsis thaliana* KNAT1 promoter. In the shoot apex, KNAT1 transcript is localized primarily to the shoot apical meristem; the expression of KNAT1 in the shoot meristem decreases during the floral transition and is
 25 restricted to the cortex of the inflorescence stem (see, e.g., Lincoln *et al.*, Plant Cell 6:1859-1876, 1994).

Suitable seed-specific and seed enhanced promoters can be derived from the following genes: MAC1 from maize (Sheridan *et al.*, Genetics 142:1009-1020, 1996); Cat3 from maize (Genbank No. L05934, Abler *et al.*, Plant Mol. Biol. 22:10131-1038,
 30 1993); viviparous-1 from *Arabidopsis* (Genbank No. U93215); Atimycl from *Arabidopsis* (Urao *et al.*, Plant Mol. Biol.

32:571-57, 1996; Conceicao *et al.*, Plant 5:493-505, 1994; napA from *Brassica napus* (Genbank No. J02798); the napin gene family from *Brassica napus* (Sjodahl *et al.*, Planta 197:264-271, 1995).

5 The ovule-specific promoter for BEL1 gene (Reiser *et al.* Cell 83:735-742, 1995, Genbank No. U39944; Ray *et al.*, Proc. Natl. Acad. Sci. USA 91:5761-5765, 1994) can also be used. The egg and central cell specific MEA (FIS1) and FIS2 promoters are also useful reproductive tissue-specific promoters (Luo *et al.*, Proc. Natl. Acad. Sci. USA, 97:10637-10642, 2000; Vielle-Calzada, *et al.*, Genes Dev. 13:2971-2982, 1999).

10 A maize pollen-specific promoter has been identified in maize (Guerrero *et al.*, Mol. Gen. Genet. 224:161-168, 1990). Other genes specifically expressed in pollen have been described (see, *e.g.*, Wakeley *et al.*, Plant Mol. Biol. 37:187-192, 1998; Ficker *et al.*, Mol. Gen. Genet. 257:132-142, 1998; Kulikaukas *et al.*, Plant 15 Mol. Biol. 34:809-814, 1997; Treacy *et al.*, Plant Mol. Biol. 34:603-611, 1997).

Promoters derived from genes encoding embryonic storage proteins, which includes the gene encoding the 2S storage protein from *Brassica napus* (Dasgupta *et al.*, Gene 133:301-302, 1993); the 2s seed storage protein gene family

20 from *Arabidopsis*; the gene encoding oleosin 20kD from *Brassica napus* (GenBank No. M63985); the genes encoding oleosin A (Genbank No. U09118) and oleosin B (GenBank No. U09119) from soybean; the gene encoding oleosin from *Arabidopsis* (GenBank No. Z17657); the gene encoding oleosin 18kD from maize (GenBank No. J05212, Lee, Plant Mol. Biol. 26:1981-1987, 1994); and the gene encoding low molecular weight sulphur rich protein 25 from soybean (Choi *et al.*, Mol. Gen. Genet, 246:266-268, 1995), can also be used.

Promoters derived from zein encoding genes (including the 15 kD, 16 kD, 19 kD, 22 kD, 27 kD, and gamma genes; Pedersen *et al.*, Cell 29:1015-1026, 1982), can be also used.

The zeins are a group of storage proteins found in maize endosperm.

-51-

Other promoters known to function, for example, in maize, include the promoters for the following genes: *waxy*, *Brittle*, *Shrunken 2*, Branching enzymes I and II, starch synthases, debranching enzymes, oleosins, glutelins, and sucrose synthases. A particularly preferred promoter for maize endosperm expression is the promoter for the glutelin gene from rice, more particularly the Osgt-1 promoter (Zheng *et al.*, Mol. Cell Biol. 13:5829-5842, 1993). Examples of promoters suitable for expression in wheat include those promoters for the ADPGlucose pyrophosphorylase (ADPGPP) subunits, the granule bound and other starch synthases, the branching and debranching enzymes, the embryogenesis-abundant proteins, the gliadins, and the glutenins. Examples of such promoters in rice include those promoters for the ADPGPP subunits, the granule bound and other starch synthases, the branching enzymes, the debranching enzymes, sucrose synthases, and the glutelins. A particularly preferred promoter is the promoter for rice glutelin, Osgt-1. Examples of such promoters for barley include those for the ADPGPP subunits, the granule bound and other starch synthases, the branching enzymes, the debranching enzymes, sucrose synthases, the hordeins, the embryo globulins, and the aleurone specific proteins.

A tomato promoter active during fruit ripening, senescence and abscission of leaves and, to a lesser extent, of flowers can be used (Blume *et al.*, Plant J. 12:731-746, 1997).

Other exemplary promoters include the pistal specific promoter in the potato (*Solanum tuberosum* L.) SK2 gene, encoding a pistil-specific basic endochitinase (Ficker *et al.*, Plant Mol. Biol. 35:425-431, 1997); the BLec4 gene from pea (*Pisum sativum* cv. Alaska), active in epidermal tissue of

vegetative and floral shoot apices of transgenic alfalfa. This makes it a useful tool to target the expression of foreign genes to the epidermal layer of actively growing shoots. The tissue specific E8 promoter from tomato is also useful for directing gene expression in fruits (Deikman, *et al.*, Plant Physiology 100:2013-2017, 1992).

It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

Promoters that are known or are found to cause transcription of DNA in plant cells can be used in the present invention. Such promoters may be obtained from a variety of sources such as plants and plant viruses. In addition to promoters that are known to cause transcription of DNA

-52-

in plant cells, other promoter molecules may be identified for use in the current invention by screening a plant cDNA library for genes that are selectively or preferably expressed in the target tissues or cells and isolating the 5' genomic region of the identified cDNAs.

Constructs or vectors may also include with the coding region of interest a polynucleic acid that acts, in whole or in part, to terminate transcription of that region. For example, such sequences have been isolated including the Tr7 3' sequence and the *nos* 3' sequence (Ingelbrecht *et al.*, The Plant Cell 1:671-680, 1989, Bevan *et al.*, Nucleic Acids Res. 11:369-385, 1983).

A vector or construct may also include regulatory elements. Examples of such include the Adh intron 1 (Callis *et al.*, Genes and Develop. 1:1183-1200, 1987), the sucrose synthase intron (Vasil *et al.*, Plant Physiol. 91:1575-1579, 1989), and the TMV omega element (Gallie *et al.*, Plant Cell 1:301-311, 1989). These and other regulatory elements may be included when appropriate.

A vector or construct may also include a selectable marker. Selectable markers may also be used to select for plants or plant cells that contain the exogenous genetic material. Examples of such include, but are not limited to, a *neo* gene (Potrykus *et al.*, Mol. Gen. Genet. 199:183-188, 1985), which codes for kanamycin resistance and can be selected for using kanamycin, G418, etc.; a *bar* gene that provides for bialaphos resistance; a mutant EPSP synthase gene (Hinchee *et al.*, Bio/Technology 6:915-922 (1988)), that provide for glyphosate resistance; a nitrilase gene that provides for resistance to bromoxynil (Stalker *et al.*, J. Biol. Chem. 263:6310-6314 (1988)); a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea; and a methotrexate resistant DHFR gene (Thillet *et al.*, J. Biol. Chem. 263:12500-12508, 1988).

A vector or construct may also include a screenable marker. Screenable markers may be used to monitor expression. Exemplary screenable markers include a β -glucuronidase or *uidA* gene (GUS) which encodes an enzyme for which various chromogenic substrates are known

-53-

(Jefferson, Plant Mol. Biol. Rep. 5:387-405 (1987)); Jefferson *et al.*, EMBO J. 6:3901-3907 (1987)); an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al.*, Stadler Symposium 11:263-282 (1988)); a β -lactamase gene (Sutcliffe *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 75:3737-3741 (1978)), a gene which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); a luciferase gene (Ow *et al.*, Science 234:856-859 (1986)); a xylE gene (Zukowsky *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 80:1101-1105 (1983)), which encodes a catechol dioxygenase that can convert chromogenic catechols; an α -amylase gene (Ikata *et al.*, Bio/Technol. 8:241-242, 1990); a tyrosinase gene (Katz *et al.*, J. Gen. Microbiol. 129:2703-2714, 1983), which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to melanin; and an α -galactosidase.

Introduction of Polynucleotides into Plants

There are many methods for introducing transforming nucleic acid molecules into plant cells. Suitable methods are believed to include virtually any method by which nucleic acid molecules may be introduced into a cell, such as by *Agrobacterium* infection or direct delivery of nucleic acid molecules such as, for example, by PEG-mediated transformation, by electroporation or by acceleration of DNA coated particles, etc. (Potrykus, Ann. Rev. Plant Physiol. Plant Mol. Biol. 42:205-225 (1991); Vasil, Plant Mol. Biol. 25:925-937 (1994)). For example, electroporation has been used to transform *Zea mays* protoplasts (Fromm *et al.*, Nature 312:791-793, 1986).

Other vector systems suitable for introducing transforming DNA into a host plant cell include but are not limited to binary artificial chromosome (BIBAC) vectors (Hamilton *et al.*, Gene 200:107-116, 1997), and

-54-

transfection with RNA viral vectors (Della-Cioppa *et al.*, Ann. N.Y. Acad. Sci. (1996), 792 pp *Engineering Plants for Commercial Products and Applications*, pp 57-61).

Technology for introduction of DNA into cells is well known to those of skill in the art.

5 Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham and van der Eb, Virology 54:536-539, 1973); (2) physical methods such as microinjection (Capecchi, *Cell* 22:479-488 (1980)), electroporation (Wong and Neumann, Biochem. Biophys. Res. Commun. 107:584-587 (1982); Fromm *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 82:5824-5828 (1985); U.S. Patent No. 5,384,253); and the
10 gene gun (Johnston and Tang, Methods Cell Biol. 43:353-365 (1994)); (3) viral vectors (Clapp, Clin. Perinatol. 20:155-168, 1993; Lu *et al.*, J. Exp. Med. 178:2089-2096, 1993; Eglitis and Anderson, *Biotechniques* 6:608-614, 1988); and (4) receptor-mediated mechanisms (Curiel *et al.*, Humm. Gen. Ther. 3:147-154, 1992, Wagner *et al.*, Proc. Natl. Acad. Sci. USA 89:6099-6103, 1992).

15 Acceleration methods that may be used include, for example, microprojectile bombardment and the like. One example of a method for delivering transforming nucleic acid
20 molecules to plant cells is microprojectile bombardment. This method has been reviewed by Yang and Christou, *eds.*, Particle Bombardment Technology for Gene Transfer, Oxford Press, Oxford, England (1994); Non-biological particles (microprojectiles) that may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold,
25 platinum, and the like.

Agrobacterium-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for
30 example the methods described by Fraley *et al.*, Bio/Technology 3:629-635 (1985) and Rogers *et al.*, Methods Enzymol. 153:253-277 (1987),

-55-

Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al.*, Mol. Gen. Genet. 205:34 (1986)).

5 A transgenic plant resulting from *Agrobacterium* transformation methods frequently contains a single gene on one chromosome. Such transgenic plants can be referred to as being hemizygous for the added gene. More preferred is a transgenic plant that is homozygous for the added structural gene; *i.e.*, a transgenic plant that contains two added genes, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be
10 obtained by sexually mating (selfing) an independent segregant transgenic plant that contains a single added gene, germinating some of the seed produced and analyzing the resulting plants produced for the gene of interest.

It is also to be understood that two different transgenic plants can also be mated to produce offspring that contain two independently segregating added, exogenous genes. Selfing
15 of appropriate progeny can produce plants that are homozygous for both added, exogenous genes that encode a polypeptide of interest. Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated, as is vegetative propagation.

The regeneration, development, and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art (Weissbach and
20 Weissbach, In: *Methods for Plant Molecular Biology*, (eds.), Academic Press, Inc. San Diego, CA, (1988)). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic
25 rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a protein of interest is well known in the art. Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely,
30 pollen from plants of these important lines is used to pollinate regenerated plants. A transgenic

-56-

plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

The present invention also provides for parts of the plants of the present invention. Plant parts, without limitation, include seed, endosperm, ovule and pollen. In a particularly preferred embodiment of the present invention, the plant part is a seed.

Methods for transforming dicots, primarily by use of *Agrobacterium tumefaciens*, and obtaining transgenic plants have been published, e.g., cotton (U.S. Patent No. 5,004,863, U.S. Patent No. 5,159,135, U.S. Patent No. 5,518,908), soybean (U.S. Patent No. 5,569,834), and *Brassica* (U.S. Patent No. 5,463,174).

Transformation of monocotyledons using electroporation, particle bombardment, and *Agrobacterium* have also been reported. For example, transformation and plant regeneration have been achieved in asparagus, barley, *Zea mays* (Fromm *et al.*, Bio/Technology 8:833 (1990), Armstrong *et al.*, Crop Science 35:550-557 (1995)); oat; rice, rye, sugarcane; tall fescue and wheat (U.S. Patent No. 5,631,152).

In addition to the above discussed procedures, practitioners are familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of recombinant organisms and the screening and isolating of clones, (see for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); Mailga *et al.*, *Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995); Birren *et al.*, *Genome Analysis: Detecting Genes*, 1, Cold Spring Harbor, New York (1998), Birren *et al.*, *Genome Analysis: Analyzing DNA*, 2, Cold Spring Harbor, New York (1998; *Plant Molecular Biology: A Laboratory Manual*, eds. Clark, Springer, New York (1997)).

-57-

Having now generally described the invention, the same will be more readily understood through reference to the following examples, which are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

EXAMPLES

EXAMPLE 1

When an isolated native plant polynucleotide comprising a coding sequence is reconstructed as a transgene, then introduced into the plant by methods of plant transformation there is a risk that expression from the endogenous homologous plant gene will interact negatively with the transgene. To avoid these negative interactions it may be necessary to provide a transgene polynucleotide substantially divergent in sequence from the native plant gene. An artificial polynucleotide molecule can be produced by the method of the present invention and used to reduce the occurrence of transgene silencing.

This example serves to illustrate methods of the present invention that result in the production of a polynucleotide encoding a modified plant EPSP synthase. The native rice (*Oryza sativa*) EPSPS enzyme and chloroplast transit peptide is used to construct an artificial polynucleotide molecule that also includes codons that encode for substituted amino acids that do not naturally occur in the rice EPSPS enzyme. These substituted amino acids provide for a glyphosate resistant rice EPSPS enzyme (OsEPSPS_TIPS, SEQ ID NO:1).

The steps described in Table 6 are used to construct such an artificial polynucleotide sequence (OsEPSPS_AT, SEQ ID NO:3) using an *Arabidopsis* codon usage table and the parameters for construction of a substantially divergent polynucleotide molecule, which when expressed in plants encodes a modified rice EPSPS enzyme resistant to glyphosate herbicide. The comparison of the native rice EPSPS gene sequence referred to as OsEPSPS_Nat (SEQ ID NO:2) that has previously been modified to encode a glyphosate resistant enzyme to the polynucleotide molecule modified for *Arabidopsis* codon usage, OsEPSPS_AT (SEQ ID NO: 3) and to the sequence modified for *Zea mays* codon usage, OsEPSPS_ZM (SEQ ID NO: 4) by this method is shown in Figure 1. Figure 1 shows nucleotide bases changed in the modified polynucleotides compared to OsEPSPS_Nat, SEQ ID NO: 2.

Table 6. Polynucleotide design for a modified rice EPSP synthase (OsEPSPS_AT)

1. Substitute amino acids at positions 173 and 177 to provide a modified rice EPSPS enzyme resistant to glyphosate herbicide shown in SEQ ID NO:1.
2. Back translate SEQ ID NO:1 to generate an artificial polynucleotide sequence using the *Arabidopsis thaliana* codon usage table (Table 2).
3. Perform sequence alignment with native OsEPSPS polynucleotide sequence (SEQ ID NO:2) and the artificial polynucleotide sequence to determine degree of sequence identity, map open reading frames, select patterns to search and identify restriction enzymes recognition sequences.
4. Make corrections to the codons used in the artificial polynucleotide sequence to achieve desired percentage of sequence identity and to avoid clustering of identical codons. This is especially important for amino acids that occur at high frequency, *i.e.*, alanine, glycine, histidine, leucine, serine, and valine. Approximate distribution of codon usage in the polynucleotide sequence according to the *Arabidopsis* codon usage, Table 2.
5. The polynucleotide sequence is inspected for local regions that have a GC:AT ratio higher than about 2 over a range of about 50 contiguous nucleotides. The polynucleotide sequence is adjusted as necessary, by substituting codons in these regions such that the local GC:AT ratio is less than about 2 and the entire polynucleotide composition is in the range of 0.9-1.3.
6. Introduce stop codons to translation frames "b", "c", "d", "e" and "f". Translation stop codons are created in the "b", "c", "d", "e" and "f" translational frames by replacing one or more codons within about 130 base pairs (bps) of the ends of the artificial polynucleotide that creates a stop codon without changing the amino acid coding sequence of frame one.
7. Eliminate ATG codons from forward (frames "b" and "c") and reverse open reading frames (frame "d", "e", "f"). The forward and reverse reading frames are inspected for the presence of ATG codons. Any ATG codons in frame "b" and "c" found in the polynucleotide sequence before third Met in frame "a" of the polynucleotide are eliminated by replacing one or more codons that overlap the ATG changing one of the nucleotides without changing the amino acid coding sequence of frame "a". In the reverse frames, replacement of ATG or stop codon introduction may be done to interrupt potential reading frames.

-59-

8. Eliminate unwanted restriction enzyme recognition patterns and other specific patterns (polyadenylation, RNA splicing, sequence instability patterns). The polynucleotide sequence is inspected for the presence of any unwanted polynucleotide patterns and the patterns are disrupted by substituting codons in these regions.
- 5 9. Check sequence identity between a first polynucleotide and the artificial polynucleotide created by the method of the present invention. Eliminate sequence identity in a contiguous polynucleotide that is longer than 23 bps. It is desirable to eliminate sequence identity greater than about 15 bps. It is helpful to select from amino acids such as, serine, arginine, and leucine that have 6 codons or from amino acids with 4 codons to eliminate
10 sequence identity.
10. Review the artificial polynucleotide sequence resulting from anyone of steps 1 to 9 for any of the sequence features identified in steps 4-9, and if the sequence does not comply with conditions make additional codon substitutions to the sequence until the conditions of steps 4-9 are met.
- 15 11. Construct the artificial polynucleotide molecule by methods known in the art, e.g., using PCR with a mixture of overlapping primers. The primers at the ends of the gene may contain convenient restriction sites to allow easy cloning of the gene to selected vector. At the 5' end usually AlflII, BspHI, NcoI, NdeI, PciI, or SphI are most convenient in as much as their sequence contains an ATG start codon, however other enzymes can be used
20 as well if a modified polynucleotide is designed to create a fusion with another polynucleotide segment, e.g., chloroplast transit peptide and EPSPS coding sequence.
12. Perform a DNA sequence analysis of the artificial polynucleotide to confirm the synthetic construction resulted in the desired polynucleotide molecule. If errors are found, then eliminate these by site directed mutagenesis for which many methods are known to those
25 skilled in the art of DNA mutagenesis.

A *Zea mays* codon usage (*Zea mays*, Table 3) version of the glyphosate resistant rice EPSPS enzyme sequence (*Oryzae sativa* EPSPS enzyme with TIPS mutations, SEQ ID NO:1) is made. The polynucleotide that encodes this enzyme includes codons that encode for substituted
30 amino acids that do not naturally occur in the native rice EPSPS enzyme. These substituted amino acids provide for a glyphosate resistant rice EPSPS enzyme. The steps described in Table

-60-

7 are used to construct a modified artificial polynucleotide sequence (OsEPSPS_ZM, SEQ ID NO:4) based on a *Zea mays* codon usage table that encodes a modified rice EPSPS enzyme resistant to glyphosate herbicide. The comparison of the OsEPSPS_Nat polynucleotide sequence (SEQ ID NO:2) to the OsEPSPS_ZM artificial polynucleotide sequence (SEQ ID NO:4) using the *Zea mays* codon usage is shown in Figure 2.

Table 7. Polynucleotide construction for modified rice EPSP synthase (OsEPSPS_ZM)

1. Back translate SEQ ID NO:1 to generate an artificial polynucleotide sequence using the *Zea mays* codon usage table (Table 3).
2. Perform sequence alignment with the native OsEPSPS polynucleotide sequence (SEQ ID NO:2) and the artificial polynucleotide sequence to determine degree of sequence identity, map open reading frames, select patterns to search and identify restriction enzymes recognition sequences.
3. Make corrections to the codons used in the artificial polynucleotide sequence to achieve desired percentage of sequence identity and to avoid clustering of identical codons. This is especially important for amino acids that occur at high frequency, *i.e.*, alanine, glycine, histidine, leucine, serine, and valine. Approximate distribution of codon usage in the polynucleotide sequence according to the *Zea mays* codon usage, Table 3.
4. The polynucleotide sequence is inspected for local regions that have a GC:AT ratio higher than about 2 over a range of about 50 contiguous nucleotides. The polynucleotide sequence is adjusted as necessary, by substituting codons in these regions such that the local GC:AT ratio is less than about 2 and the entire polynucleotide composition is in the range of 1.2-1.7.
5. Follow steps 6-12 of Table 6.

Table 8. Sequence percent identity between OsEPSPS polynucleotides.

	OsEPSPS_ZM	OsEPSPS_AT	OsEPSPS_Nat
OsEPSPS_ZM	100.00	73.51	71.58
OsEPSPS_AT		100.00	74.03
OsEPSPS_Nat			100.00

Table 9. The nucleotide composition and GC:AT ratio of the modified polynucleotide sequences for the rice EPSPS gene sequence.

-61-

	A	C	G	T	GC:AT
OsEPSPS_AT	377	336	444	391	1.02
OsEPSPS_ZM	365	381	470	332	1.22

The two rice EPSPS artificial polynucleotide sequences (SEQ ID NO:3 and SEQ ID NO:4) are modified such that the percent identity is below 75 percent compared to SEQ ID NO:2 or relative to each other (Table 8). The nucleotide composition and GC:AT ratio of the polynucleotide sequences for the rice EPSPS gene sequence are shown in Table 9. These polynucleotides can be selected for use in plant expression constructs together with different regulatory elements or they can be combined in a single plant by retransformation with a DNA construct or by methods of plant breeding. Concerns with gene silencing and recombination are reduced when DNA constructs have reduced levels of homologous DNA.

EXAMPLE 2

Corn (*Zea mays*) has been genetically modified to have resistance to glyphosate herbicide (US Patent No. 6,040,497). These corn plants contain a transgene with a corn EPSP synthase modified for glyphosate tolerance. The methods of the present invention can be used to construct a new artificial polynucleotide encoding a corn EPSP synthase that is substantially different in percent identity to the endogenous corn EPSP synthase gene. The newly constructed corn EPSP synthase artificial polynucleotide can be used as a selectable marker during the selection of transgenic plant lines that may contain additional transgenic agronomic traits. During hybrid corn seed production, it is useful to have both parents glyphosate tolerant using non-interfering transgenes.

Table 10. Polynucleotide construction for modified corn EPSP synthase (ZmEPSPS_ZM, SEQ ID NO:10)

1. Back translate SEQ ID NO:8 to generate a polynucleotide sequence using the *Zea mays* codon usage table (Table 3).
2. Perform sequence alignment with ZmEPSPS_Nat polynucleotide sequence (SEQ ID NO:9) and the artificial polynucleotide sequence to determine degree of sequence identity, map open reading frames, select patterns to search and identify restriction enzymes recognition sequences.

-62-

3. Make corrections to the codons used in the artificial polynucleotide sequence to achieve desired percentage of sequence identity and to avoid clustering of identical codons. This is especially important for amino acids that occur at high frequency, *i.e.*, alanine, glycine, histidine, leucine, serine, and valine. Approximate distribution of codon usage in the polynucleotide sequence according to the *Zea mays* codon usage, Table 3.
4. The artificial polynucleotide sequence is inspected for local regions that have a GC:AT ratio higher than about 2 over a range of about 50 contiguous nucleotides. The polynucleotide sequence is adjusted as necessary, by substituting codons in these regions such that the local GC:AT ratio is less than about 2 and the entire polynucleotide composition is in the range of 1.2-1.7.
5. Follow steps 6-12 of Table 6.

Table 11. Sequence percent identity between ZmEPSPS polynucleotides.

	ZmEPSPS_ZM	ZmEPSPS_Nat
ZmEPSPS_ZM	100.00	74.81
ZmEPSPS_Nat		100.00

Maize EPSPS gene nucleotide sequence is also modified to reduce identity between synthetic and native gene and maintain overall GC:AT ratio typical for monocots. The GC:AT ratio for the ZmEPSPS_ZM sequence is 1.38. The sequence identity is reduced to about 75% between native (ZmEPSPS_Nat, SEQ ID NO:9) and synthetic (ZmEPSPS_ZM, SEQ ID NO:10).

The comparison of native polynucleotides encoding EPSPS indicate that the chloroplast transit peptide is the most divergent fragment of the gene. Similarity in nucleotide sequence of mature peptides is higher than 88% for maize and rice enzymes, and some conserved regions have sequence identity as long as 50 bps. Posttranscriptional gene silencing has been observed for sequences as small as 60 polynucleotides (Sijen *et al.*, Plant Cell, 8:2277-2294, 1996; Mains, Plant Mol. Biol. 43:261-273, 2000).

EXAMPLE 3

Soybean (*Glycine max*) has been genetically modified to be tolerant to glyphosate by expression of a class II EPSPS isolated from *Agrobacterium* (Padgett *et al.* Crop Sci. 35:1451-1461, 1995). A soybean native EPSPS gene sequence has been identified and an artificial polynucleotide sequence designed using the method of the present invention. The artificial polynucleotide encodes a protein sequence that is modified to produce a glyphosate resistant EPSPS enzyme (GmEPSPS_IKS, SEQ ID NO:5) by replacing amino acids T to I, R to K and P to S within the GNAGTAMRP motif, resulting in a modified soybean EPSPS enzyme with the motif GNAGIAMKS (SEQ ID NO:34), also referred to as IKS mutant. Expression of a modified EPSPS enzyme in the cells of a plant by transformation with a transgene plant expression cassette, which contains a polynucleotide encoding the modified EPSPS with the motif GNAGIAMKS will confer glyphosate tolerance to the plants. Additional amino acid substitutions for the arginine (R) in the motif can also include asparagine (N).

Table 12. Polynucleotide construction for modified soybean EPSP synthase gene (GmEPSPS_GM, SEQ ID NO:7).

1. Back translate SEQ ID NO:5 to generate an artificial polynucleotide sequence using the *Glycine max* codon usage table (Table 4).
2. Perform sequence alignment with GmEPSPS_Nat polynucleotide sequence (SEQ ID NO:6) and the artificial polynucleotide sequence to determine degree of sequence identity, map open reading frames, select patterns to search and identify restriction enzymes recognition sequences.
3. Make corrections to the codons used in the artificial polynucleotide sequence to achieve desired percentage of sequence identity and to avoid clustering of identical codons. This is especially important for amino acids that are occur at high frequency, *i.e.*, alanine, glycine, histidine, leucine, serine, and valine. Approximate distribution of codon usage in the polynucleotide sequence according to the *Glycine max* codon usage, Table 4.
4. The polynucleotide sequence is inspected for local regions that have a GC:AT ratio higher than about 2 over a range of about 50 contiguous nucleotides. The polynucleotide sequence is adjusted as necessary, by substituting codons in these regions such that the

-64-

local GC:AT ratio is less than about 2 and the entire polynucleotide composition is in the range of 0.9-1.3.

5. Follow steps 6-12 of Table 6.

5 Table 13. Comparison of the sequence percent identity of the modified GmEPSPS at polynucleotide sequence level.

	GmEPSPS_GM	GmEPSPS_Nat
GmEPSPS_GM	100.00	72.43
GmEPSPS_Nat		100.00

10 The soybean native EPSPS gene is modified using a soybean codon table (Table 4) and the conditions of the method of the present invention. The relative ratio of GC:AT is not changed in the modified gene, however the sequence identity between the two is reduced to 72%.

EXAMPLE 4

The native *aroA* polynucleotide gene isolated from *Agrobacterium* strain CP4 (U.S. Patent No. 5,633,435), that encodes a glyphosate
 15 resistant EPSP synthase (SEQ ID NO:15) can be modified by the method of the present invention to provide a polynucleotide that has the codon usage of *Arabidopsis*, *Zea mays*, or *Glycine max*. For the appropriate expression of CP4EPSPS to confer glyphosate tolerance in plants, a chloroplast transit peptide is necessarily fused to the CP4EPSPS coding sequence to target accumulation of the enzyme to the chloroplasts. The CTP2 chloroplast transit peptide is
 20 commonly used for the expression of this gene in transgenic plants (Nida *et al.*, J. Agric, Food Chem. 44:1960-1966, 1996). The sequence of CP4EPSPS together with CTP2 polynucleotide (SEQ ID NO:11) have been modified by the method of the present invention. Other chloroplast transit peptides known in the art can be fused to the CP4EPSPS to direct the enzyme to the chloroplasts.

Table 14. Polynucleotide construction for aroA:CP4 EPSP synthase coding sequence (CP4EPSPS_AT, CP4EPSPS_ZM, or CP4EPSPS_GM)

1. Place CTP2 transit peptide sequence (SEQ ID NO:11) in front of CP4EPSPS (SEQ ID NO:15) as a fusion polypeptide. Back translate the fusion polypeptide to produce an artificial polynucleotide sequence using the *Arabidopsis thaliana* codon usage table (Table 2), or the *Zea mays* codon usage table (Table 3), or the *Glycine max* codon usage table (Table 4).
2. Perform sequence alignment with native CTP2 (SEQ ID NO:12) and native CP4EPSPS polynucleotide sequence (SEQ ID NO:16) and the artificial polynucleotide sequence to determine degree of sequence identity, map open reading frames, select patterns to search and identify restriction enzymes recognition sequences.
3. Make corrections to the codons used in the artificial polynucleotide sequence to achieve desired percentage of sequence identity and to avoid clustering of identical codons. This is especially important for amino acids that occur at high frequency, *i.e.*, alanine, glycine, histidine, leucine, serine, and valine. Approximate distribution of codon usage in the polynucleotide sequence according to the *Arabidopsis thaliana* codon usage, Table 2, or the *Zea mays* codon usage table (Table 3) depending on the table in use.
4. The artificial polynucleotide sequence is inspected for local regions that have a GC:AT ratio higher than about 2 over a range of about 50 contiguous nucleotides. The polynucleotide sequence is adjusted as necessary, by substituting codons in these regions such that the local GC:AT ratio is less than about 2 and the entire polynucleotide composition is in the range of 0.9-1.3 if Table 2 is used and 1.2-1.7 if Table 3 is used.
5. Follow steps 6-12 of Table 6.

Table 15. Comparison of the sequence percent identity of the artificial CP4EPSPS polynucleotides.

	CTP2CP _GM	CTP2CP4 _AT	CTP2CP4 _ZM	CTP2CP4 _Syn	CTP2CP4 _NAT
CTP2CP4 _GM	100.00	75.66	74.12	75.15	74.37
CTP2CP4 _AT		100.00	76.13	74.56*	72.93
CTP2CP4 _ZM			100.00	77.76*	82.58
CTP2CP4 _Syn				100.00	82.70
CTP2CP4 _NAT					100.00

*Percent of identity relates to the CP4EPSPS and do not include transit peptide.

Table 16. The nucleotide composition and GC:AT ratio of the artificial polynucleotide sequences for the CP4EPSPS gene sequence.

	A	C	G	T	GC:AT
CTP2CP4_GM	382	375	442	397	1.05
CTP2CP4_AT	369	408	469	350	1.22
CTP2CP4_ZM	312	487	577	290	1.65

The polynucleotide sequence CTP2_Nat (SEQ ID NO:12) plus CP4EPSPS_Nat (SEQ ID NO:16) designated as CTP2CP4_Nat is compared in Table 15 to the artificial polynucleotide sequences designated as CTP2CP4_AT (CTP2_AT, SEQ ID NO:13 fused to CP4EPSPS_AT, SEQ ID NO:17) and CTP2CP4_ZM (CTP2_AT, SEQ ID NO:14 fused to CP4EPSPS_ZM, SEQ ID NO:18) produced by the method of the present invention. The polynucleotide sequence that is the most divergent from the native sequence CTP2CP4_NAT and CTP2CP4EPSPS_Syn is CTP2CP4_AT having about 73% and 75% sequence identity, respectively. The CTP2CP4_ZM polynucleotide sequence compared to CTP2CP4_Nat and CP4EPSPS_Syn has about 83% and 78% identity to these two sequences, respectively.

A primary criteria for the selection of transgenes to combine in a plant is the percent identity. Table 15 can be used to select a CP4EPSPS polynucleotide molecule for plant expression cassette construction when it is known that the recipient plant will contain more than one CP4EPSPS polynucleotide. The GC:AT ratio in native CP4EPSPS is about 1.7. The artificial version with the *Zea mays* codon bias is produced to have a very similar GC:AT ratio. In the *Arabidopsis* codon version, the GC:AT ratio is decreased to about 1.2.

Gene expression is also a criteria for selection of transgenes to be expressed. Expression of a transgene can vary in different crop plants, therefore having several artificial polynucleotide coding sequence available for testing in different crop plants and genotypes, varieties or cultivars is an advantage and an aspect of the invention.

EXAMPLE 5

The *bar* polynucleotide sequence (SEQ ID NO:20) encoding a phosphinothricin acetyl transferase protein (SEQ ID NO:19) has been used to genetically modify plants for resistance to glufosinate herbicide. Two new *bar* polynucleotide sequences have been designed using the method of the present invention. The alignment of BAR1_Nat with the two new artificial BAR1 polynucleotides is shown in Figure. 4.

-67-

Table 17. Polynucleotide gene construction for BAR1_AT (SEQ ID NO:21) and BAR1_ZM (SEQ ID NO:22)

1. Back translate SEQ ID NO:19 to generate a polynucleotide sequence using the *Arabidopsis thaliana* codon usage table (Table 2) or the *Zea mays* codon usage table (Table 3)
2. Perform sequence alignment with native BAR1_Nat polynucleotide sequence (SEQ ID NO:20) and the artificial polynucleotide sequence to determine degree of sequence identity, map open reading frames, select patterns to search and identify restriction enzymes recognition sequences.
3. Make corrections to the codons used in the artificial polynucleotide sequence to achieve desired percentage of sequence identity and to avoid clustering of identical codons. This is especially important for amino acids that occur at high frequency, *i.e.*, alanine, glycine, histidine, leucine, serine, and valine. Approximate distribution of codon usage in the polynucleotide sequence according to the *Arabidopsis thaliana* codon usage, Table 2, or the *Zea mays* codon usage table (Table 3) depending on the table in use.
4. The artificial polynucleotide sequence is inspected for local regions that have a GC:AT ratio higher than about 2 over a range of about 50 contiguous nucleotides. The polynucleotide sequence is adjusted as necessary, by substituting codons in these regions such that the local GC:AT ratio is less than about 2 and the entire polynucleotide composition is in the range of 0.9-1.3 if Table 2 is used and 1.2-1.7 if Table 3 is used.
5. Follow steps 6-12 of Table 6.

The sequence identity of artificial BAR polynucleotides is the range of 73-77% (Table 18). The native polynucleotide is highly GC rich. The artificial version (BAR1_ZM) with *Zea mays* codon bias has reduced the GC:AT ratio to about 1.3 and artificial version (BAR1_AT) with *Arabidopsis* codon bias the ratio is about 1.0 (Table 19).

Table 18. Sequence percent identity between *bar* genes at the polynucleotide sequence level.

	BAR1_ZM	BAR1_AT	BAR1_Nat
BAR1_ZM	100.00	77.35	76.99
BAR1_AT		100.00	73.73
BAR1_Nat			100.00

Table 19. The nucleotide composition and GC:AT ratio of the artificial polynucleotide sequences for the *bar* gene sequence.

BAR_AT	139	130	144	139	1.01
BAR_ZM	122	156	154	120	1.28

5 EXAMPLE 6

This example serves to illustrate DNA constructs for the expression of the artificial polynucleotides of the present invention in plants. A transgene DNA plant expression cassette comprises regulatory elements that control the transcription of a mRNA from the cassette. A
 10 plant expression cassette is constructed to include a promoter that functions in plants that is operably linked to a 5' leader region that is operably linked to a DNA sequence of interest operably linked to a 3' termination region. These cassettes are constructed in plasmid vectors, which can then be transferred into plants by *Agrobacterium* mediated transformation methods or other methods known to those skilled in the art of plant transformation. The following plasmid
 15 vector constructs are illustrated to provide examples of plasmids containing plant expression cassettes comprising the artificial polynucleotide molecules of the present invention and are not limited to these examples.

The artificial polynucleotide molecules of the present invention, for example, CP4EPSPS_AT and CP4EPSPS_ZM are synthesized using overlapping primers. The full length
 20 product is then amplified with gene specific primers containing overhangs with SphI (forward primer) and EcoRI (reverse primer). Genes are cloned into the vector pCRII-TOPO (Invitrogen, CA). The resulting plasmids pMON54949 (CP4EPSPS_AT, Figure 6) and pMON54950 (CP4EPSPS_ZM, Figure 7) contain the artificial polynucleotide and these polynucleotides are sequenced using DNA sequencing methods to confirm that the modifications designed by the
 25 method of the present invention are contained in the artificial polynucleotides. In the next step, the artificial polynucleotide encoding the CTP2 chloroplast transit peptide is ligated to the 5' end of the CP4EPSPS polynucleotides. The CaMV 35S promoter with a duplicated enhancer (P-CaMVe35S) and a rice actin 1 intron (I-OsAct1) derived from pMON30151 (Figure 8) by digestion with SphI and HindIII ligated to the CTP2CP4EPSPS polynucleotides to create
 30 plasmids pMON59302 (CTP2CP4EPSPS_AT, Figure 9) and pMON59307 (CTP2CP4EPSPS_ZM, Figure 10).

-69-

For the expression of the new artificial polynucleotides in monocot plants, genes are placed in plant expression cassettes containing at the 5' end of the polynucleotide, a promoter and an intron, a 5' untranslated region, and at the 3' end of the polynucleotide a transcription termination signal. For this purpose, pMON42411 (Figure 11) containing P-CaMV35S:en, I-HSP70, CTP2CP4_Nat and NOS 3' are digested with NcoI and EcoRI restriction enzymes. The
 5 pMON59302 (Figure 9) and pMON59307 (Figure 10) are digested with same restriction enzymes. Fragments are gel purified using Qiagen gel purification kit and ligated to form pMON58400 (CP4EPSPS_AT, Figure 12) and pMON58401 (CP4EPSPS_ZM, Figure 13). Additional vector pMON54964 (Figure 14), containing P-OsAct1/ I-OsAct1 is made by
 10 replacing P-e35S/I-Hsp70 from pMON58400 (Figure 12) using HindIII /NcoI fragment from pMON25455 (Figure 15). To create a monocot expression vector containing the P-FMV promoter, pMON30152 (Figure 16) is digested with NheI, the ends are blunted with T4DNA polymerase in the presence of 4 dNTP-s (200 μ M) and NcoI. The CPT2CP4_AT or CTP2CP4_ZM DNA fragments are isolated from pMON59302 (Figure 9) or pMON59307
 15 (Figure 10), respectively by digesting with EcoRI, blunting with T4 DNA polymerase and NcoI digest. Gel purified DNA fragments are ligated and new plasmids pMON54992 (CTP2CP4_AT, Figure 17) and pMON54985 (CTP2CP4_ZM, Figure 18) are created. In each case the successful plasmid construction is confirmed by restriction endonuclease digestion, using among others ClaI (introduced to both artificial polynucleotides) and Pst I (introduced to CP4EPSPS_ZM).
 20 The CP4EPSPS_Nat present in parental vectors has both ClaI and two PstI restriction sites in coding region in different location than in artificial polynucleotides.

For the expression of the artificial CP4EPSPS polynucleotides in dicot plants, two parental vectors are used: pMON20999 (P-FMV/CTP2CP4_Syn/3'E-9, Figure 19) and pMON45313 (P-e35S/CTP2CP4_Syn/3'E9, Figure 20). In each plasmid, a DNA fragment
 25 containing the CTP2CP4_Syn polynucleotide is replaced with CTP2CP4_AT or CTP2CP4_ZM. To create pMON59308 (P-CaMVe35S/CTP2CP4_AT, Figure 21) or pMON59309 (P-CaMVe35S/CTP2CP4_ZM, Figure 22), pMON45313 is digested with NcoI and EcoRI and the DNA restriction fragments derived from NcoI/EcoRI digest of pMON59302 (CTP2CP4_AT, Figure 9) or pMON59307 (CTP2CP4_ZM, Figure 10) are ligated, respectively. To create
 30 pMON59313 (P-FMV/CTP2CP4_AT/3'E9, Figure 23) and pMON59396 (P-FMV/CTP2CP4_ZM/3'E9, Figure 24) parental plasmid pMON20999 is digested with NcoI and

-70-

BamHI to remove CTP2CP4_Syn and the restriction fragments NcoI/BamHI derived from pMON59308 (CTP2CP4_AT, Figure 21) or pMON59309 (CTP2CP4_ZM, Figure 22) are ligated, respectively.

5 EXAMPLE 7

The artificial polynucleotides are tested to determine efficacy for conferring glyphosate tolerance to transgenic plants. Five different expression cassettes (Table 20) with the new artificial CP4EPSPS polynucleotides are transformed into corn and the resulting transgenic corn plants compared to the commercial standard (Roundup Ready® Corn 603, Monsanto Co.). The
10 plasmid pMON25496 (Figure 25) contained in the commercial standard has two copies of the CP4EPSPS_Nat polynucleotide, the expression driven by the P-CaMVe35S (P-CaMVe35S) and P-OsAct1 promoters, respectively. The plasmids containing the new artificial CP4EPSPS polynucleotides contain only a single copy of the polynucleotide to be tested. The expression of these polynucleotides are driven by the P-CaMVe35S promoter with the heat shock protein
15 intron I-Hsp70 or the P-FMV promoter with a rice sucrose synthase intron (I-OsSS). Plasmid pMON54964 contains rice actin 1 promoter with first native intron (U.S. Patent No. 5,641,876).

These plasmids are transformed into corn cells by an *Agrobacterium* mediated method and transgenic corn lines regenerated on glyphosate selection. Transgenic corn plants can be
20 produced by an *Agrobacterium* mediated transformation method. A disarmed *Agrobacterium* strain C58 (ABI) harboring a binary construct of the present invention is used. This is transferred into *Agrobacterium* by a triparental mating method (Ditta *et al.*, Proc. Natl. Acad. Sci. 77:7347-7351). Liquid cultures of *Agrobacterium* are initiated from glycerol stocks or from a freshly streaked plate and grown overnight at 26°C-28°C with shaking (approximately 150
25 rpm) to mid-log growth phase in liquid LB medium, pH 7.0 containing the appropriate antibiotics. The *Agrobacterium* cells are resuspended in the inoculation medium (liquid CM4C) and the density is adjusted to OD₆₆₀ of 1. Freshly isolated Type II immature HiIIxLH198 and HiII corn embryos are inoculated with *Agrobacterium* containing a construct and co-cultured several days in the dark at 23 °C. The embryos are then transferred to delay media and incubated
30 at 28 °C for several or more days. All subsequent cultures are kept at this temperature. The embryos are transferred to a first selection medium containing carbenicillin 500/0.5 mM

-71-

glyphosate). Two weeks later, surviving tissue are transferred to a second selection medium containing carbenicillin 500/1.0 mM glyphosate). Subculture surviving callus every 2 weeks until events can be identified. This may take about 3 subcultures on 1.0 mM glyphosate. Once events are identified, bulk up the tissue to regenerate. The plantlets are transferred to MSOD media in culture vessel and kept for two weeks. Then the plants with roots are transferred into soil. Those skilled in the art of corn transformation can modify this method to provide substantially identical transgenic corn plants containing the DNA compositions of the present invention.

About 30 transgenic corn lines for each plasmid construct are tested, and the transformation efficiency and expression levels of the CP4EPSPS enzyme are shown in Table 20. The transgenic corn lines are treated with glyphosate at a rate of 64oz/acre as young plants, the surviving plants are assayed by CP4EPSPS ELISA (Padgett *et al.* Crop Sci. 35:1451-1461, 1995) to determine the CP4 EPSPS protein expression levels (CP4 exp %) shown in Table 20, and the level of expression is compared to the commercially available standard glyphosate tolerant corn plant (Roundup Ready® corn 603, Monsanto Co., St. Louis, MO) as a percent of the amount of protein expression determined in the commercial standard. Generally, more than 50% of corn lines survive the spray with 64oz/acre glyphosate. The surviving plants are shown to have high level of CP4EPSPS expression that ranges from about 75 to 86% of commercial standard 603.

Table 20. Transformation efficiency (TE), CP4 expression (average %) derived from transformation of different CP4-alt constructs.

pMON	Promoter/Intron #	TE (%)	CP4 exp (%)*
58400 (CP4_AT)	P-CaMVe35S/IHsp70	5.4	75.5
58401 (CP4_ZM)	P-CaMVe35S/IHsp70	7.2	84.7
54964 (CP4_AT)	P-OsAct1	8.2	78.1
54985 (CP4_ZM)	P-FMV/IOsSS	11.5	85.7
54992 (CP4_AT)	P-FMV/IOsSS	11.5	78.2
nk603 (control)	P-OsAct1/P-e35S:	-	100

* CP4EPSPS expression is calculated as percent of control (603) done on plants that survived glyphosate spray (64oz/acre).

-72-

EXAMPLE 8

Three plasmid constructs are evaluated in transgenic cotton plants (Table 21). The control construct (pMON20999) contains P-FMV/CP4EPSPS_Syn this expression cassette is contained in the commercially available glyphosate tolerant cotton line 1445 (Roundup Ready® cotton, Monsanto Co., St. Louis, MO). The plasmid constructs, pMON59313 and pMON59396 containing the CP4EPSPS_AT and CP4EPSPS_ZM polynucleotides, respectively, are assayed for transformation efficiency and CP4EPSPS enzyme levels relative to the commercial glyphosate tolerant expression cassette. About fifty transgenic cotton lines are evaluated for each construct. The artificial CP4EPSPS_AT polynucleotide driven by the P-FMV promoter gives a higher percentage of plants with a single insert, and an increase in expression level of the CP4EPSPS enzyme relative to the pMON20999 expression cassette as measured by ELISA.

Table 21. Transformation efficiency (TE), average CP4EPSPS expression in R0 cotton lines derived from transformation of different CP4EPSPS constructs.

pMON	Promoter	TE (%)	CP4 Exp (%)
20999 (CP4EPSPS_Syn)	P-FMV	15.0	100.0
59313 (CP4EPSPS_AT)	P-FMV	15.0	116.4
59396 (CP4EPSPS_ZM)	P-FMV	16.1	52.0

EXAMPLE 9

Constructs containing the artificial CP4EPSPS polynucleotides, CP4EPSPS_AT and CP4EPSPS_ZM are evaluated in soybean (Table 22). The plasmid constructs all contain the P-FMV promoter to drive expression of the new CP4EPSPS polynucleotides and are compared to the P-FMV/CP4EPSPS_Syn contained in pMON20999. About 25 to 30 transgenic soybean plants are produced for each construct. The transformation efficiency and CP4EPSPS enzyme levels are measured. A surprisingly high expression level of CP4EPSPS protein is measured in soybean plants containing the CP4EPSPS_ZM coding sequence (Table 22).

Table 22. Transformation efficiency (TE), average CP4 expression derived from transformation of different CP4EPSPS constructs.

pMON	Promoter	TE (%)	CP4Exp (%)
20999 (CP4_Syn)	P-FMV	0.55	100.0
59313 (CP4_AT)	P-FMV	0.40	66.6
54996 (CP4_ZM)	P-FMV	0.29	242.5

-73-

EXAMPLE 10

Tobacco cells are transformed with three plasmid constructs containing different CP4EPSPS polynucleotide sequences and regenerated into plants. About twenty transgenic lines are evaluated from each construct. Expression from each of the CP4EPSPS polynucleotides is driven by the P-CaMVe35S duplicated enhancer promoter (Table 23). The transformation efficiency and CP4EPSPS enzyme expression level is measured. The different CP4EPSPS polynucleotide constructs are shown to perform about the same in transgenic tobacco for transformation efficiency and expression.

Table 23. Transformation efficiency (TE), average CP4 expression in R0 tobacco lines derived from transformation of different CP4 EPSPS constructs.

pMON	Promoter	TE (%)	CP4 exp. (%)
59308 CP4EPSPS_AT	P-CaMVe35S	35	100.0
59309 CP4EPSPS_ZM	P-CaMVe35S	35	91.0
54313 CP4EPSPS_Syn	P-CaMVe35S	35	100.0

EXAMPLE 11

Arabidopsis thaliana is transformed with four plasmid constructs by vacuum infiltration (Bechtold N, *et al.*, CR Acad Sci Paris Sciences di la vie/life sciences 316: 1194-1199, (1993) and V1 progeny evaluated to compare efficacy of the different CP4EPSPS polynucleotide sequences and different promoters for the use in selection of plants on glyphosate (Table 24). About 30 transgenic V1 plants (+) are produced for each construct. The constructs driven by P-CaMVe35S with the duplicated enhancer (pMON45313, pMON59308, and pMON59309) show no substantial difference in the level of expression in leaves as determined by ELISA. The plants are transformed with pMON26140 that contains CP4EPSPS_Syn driven by the P-FMV promoter, these plants show the highest expression level, the expression levels detected from the plants of the test constructs are compared to pMON26140.

Table 24. Evaluation of different CP4 expression cassettes in *Arabidopsis*

pMON	Promoter/	Plants produced	CP4 exp.(%)
45313(CPEPSPS4_Syn)	P-CaMVe35S	+	82.1
59308(CP4EPSPS_AT)	P-CaMVe35S	+	79.3

-74-

59309(CP4EPSPS_ZM)	P-CaMVe35S	+	77.3
26140(CP4EPSPS_Syn)	P-FMV	+	100.0

EXAMPLE 12

5 Wheat plants transformed with the new CP4EPSPS polynucleotides are compared for transformation efficiency and CP4EPSPS enzyme expression determined by ELISA (Table 25). The CP4EPSPS_ZM provides at least seven times higher CPEPSPS enzyme expression than CP4EPSPS_AT. The average expression of CP4EPSPS in leaves from wheat plants containing the CP4EPSPS_ZM polynucleotide is about 64% of that found in glyphosate resistant wheat that
 10 contains a double cassette construct, pMON30139: P-e35S/I-Hsp70/CP4EPSPS_Nat and P-OsAct1/I-OsAct1/CP4EPSPS_Nat (WO/0022704).

Table 25 . Performance of different CP4EPSPS polynucleotides in wheat

pMON	Promoter/Intron	TE (%)	CP4 Exp. (%)
58400 CP4EPSPS_AT	P-e35S/I-Hsp70	0.25	9.2
58401 CP4EPSPS_ZM	P-e35S/I-Hsp70	0.35	64.0
30139 CP4EPSPS_Nat	P-e35S:P-OsAct1	-	100.0

15 EXAMPLE 13

This example serves to illustrate detection of different artificial polynucleotides in transgenic plants, specifically CP4EPSPS_AT and CP4EPSPS_ZM. The other artificial polynucleotides, OsEPSPS_AT, OsEPSPS_ZM, GmEPSPS_GM, ZmEPSPS_ZM, CTP2_AT, CTP2_ZM, Bar1_AT and Bar1_ZM can all be specifically detected in transgenic plants by
 20 methods that provide a DNA amplicon or by hybridization of a DNA probe to a plant sample. Those skilled in the art of DNA detection can easily design primer molecules from the artificial polynucleotide sequences provided in the present invention to enable a method that will specifically detect the artificial polynucleotide in a plant sample. The use of a method or a kit that provides DNA primers or probes homologous or complementary to the artificial
 25 polynucleotides disclosed herein is an aspect of the present invention.

A DNA detection method (polymerase chain reaction, PCR) is designed to detect the artificial CP4EPSPS polynucleotides in transgenic plants. The unique sets of DNA primers shown in Table 26 are designed to amplify a specific CP4EPSPS polynucleotide and to provide distinctly sized amplicons. The amplicons differ sufficiently in polynucleotide length among the

-75-

various CP4EPSPS polynucleotides to make easy separation of the amplicons by standard agarose gel electrophoresis. The presence of more than one of the artificial polynucleotides can be detected in a plant by using a multiplex PCR method.

Table 26. Sequence of primers used for detection of different CP4 genes in transgenic plants.

Primer pair:	Gene specificity	PCR product (bps)
SEQ ID NOs: 24 and 25	CP4EPSPS_AT	938 (940)
SEQ ID NOs: 26 and 27	CP4EPSPS_ZM	595 (600)
SEQ ID NOs: 28 and 29	CP4EPSPS_Nat	712 (710)
SEQ ID NOs: 30 and 31	CP4EPSPS_Syn	443 (440)

DNA primer pairs (Table 26) are used to produce an amplicon diagnostic for a specified CP4EPSPS polynucleotide contained in a transgenic plant. These primer pairs include, but are not limited to SEQ ID NO:24 and SEQ ID NO:25 for the CP4EPSPS_AT polynucleotide; SEQ ID NO:26 and SEQ ID NO:27 for the CP4EPSPS_ZM; SEQ ID NO:28 and SEQ ID NO:29 for CP4EPSPS_Nat and SEQ ID NO:30 and SEQ ID NO:31 CP4EPSPS_Syn polynucleotide molecule. In addition to these primer pairs, any primer pair derived from SEQ ID NO:17 or SEQ ID NO:18 that when used in a DNA amplification reaction produces a DNA amplicon diagnostic for the respective CP4EPSPS polynucleotide is an aspect of the present invention.

The amplification conditions for this analysis is illustrated in Table 27 and Table 28, however, any modification of these conditions including the use of fragments of the DNA molecules of the present invention or complements thereof as primer molecules, which produce an amplicon DNA molecule diagnostic for the artificial polynucleotides described herein is within the ordinary skill of the art. The DNA molecules of the present invention include at least SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:22, and SEQ ID NO:35. DNA molecules that function as primer molecules in a DNA amplification method to detect the presence of the artificial polynucleotides include, but are not limited to SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, and SEQ ID NO:31.

-76-

In a method for determining the presence of polynucleotides of the present invention, the analysis of plant tissue DNA extract sample should include a positive control known to contain the artificial polynucleotide, and a negative DNA extract control from a plant that is not transgenic or does not contain the artificial polynucleotide, and a negative control that contains
5 no template in the DNA extract.

Additional DNA primer molecules of sufficient length can be selected from SEQ ID NO:17 and SEQ ID NO:18 and conditions optimized for the production of an amplicon that may differ from the methods shown in Table 27 and Table 28, but result in an amplicon diagnostic for the artificial polynucleotides. The use of these DNA primer sequences homologous or
10 complementary to SEQ ID NO:17 and SEQ ID NO:18 used with or without modifications to the methods of Table 27 and 28 are within the scope of the invention. The assay for the CP4EPSPS_AT and CP4EPSPS_ZM amplicon can be performed by using a Stratagene Robocycler, MJ Engine, Perkin-Elmer 9700, or Eppendorf Mastercycler Gradient thermocycler as shown in Table 28, or by methods and apparatus known to those skilled in the art.

15 Table 27. DNA amplification procedure and reaction mixture for the confirmation of artificial EPSPS polynucleotide CP4EPSPS_AT in corn plants.

Step	Reagent	Amount	Comments
1	Nuclease-free water	add to final volume of 20 μ l	-
2	10X reaction buffer (with $MgCl_2$)	2.0 μ l	1X final concentration of buffer, 1.5 mM final concentration of $MgCl_2$
3	10 mM solution of dATP, dCTP, dGTP, and dTTP	0.4 μ l	200 μ M final concentration of each dNTP
4	primer (SEQ ID NO:24) (resuspended in 1X TE buffer or nuclease-free water to a concentration of 10 μ M)	0.4 μ l	0.2 μ M final concentration
5	primer (SEQ ID NO:25) (resuspended in 1X TE buffer or nuclease-free water to a concentration of 10 μ M)	0.4 μ l	0.2 μ M final concentration
6	control primer (SEQ ID NO:32) (resuspended in 1X TE buffer or nuclease-free water to a concentration of 10 μ M)	0.2 μ l	0.1 μ M final concentration

-77-

7	control primer (SEQ ID NO:33) (resuspended in 1X TE buffer or nuclease-free water to a concentration of 10 μ M)	0.2 μ l	0.1 μ M final concentration
8	RNase, DNase free (500 ng/ μ l)	0.1 μ l	50 ng/reaction
9	<i>REDTaq</i> DNA polymerase (1 unit/ μ l)	1.0 μ l (recommended to switch pipets prior to next step)	1 unit/reaction
10	Extracted DNA (template): • Samples to be analyzed * individual leaves * pooled leaves (maximum of 50 leaves/pool) • Negative control • Negative control • Positive control	• 10-200 ng of genomic DNA • 200 ng of genomic DNA • 50 ng of nontransgenic plant genomic DNA • no template DNA • 5 ng plasmid DNA	-

Table 28. Suggested PCR parameters for different thermocyclers

Gently mix and, if needed (no hot top on thermocycler), add 1-2 drops of mineral oil on top of each reaction. Proceed with the PCR in a Stratagene Robocycler, MJ Engine, Perkin-Elmer 9700, or Eppendorf Mastercycler™ Gradient thermocycler using the following cycling parameters.

Note: The MJ Engine or Eppendorf Mastercycler™ Gradient thermocycler should be run in the calculated mode. Run the Perkin-Elmer 9700 thermocycler with the ramp speed set at maximum.

Cycle No.	Settings: Stratagene Robocycler
1	94°C 3 minutes
38	94°C 1 minute 60°C 1 minute 72°C 1 minute and 30 seconds
1	72°C 10 minutes

Cycle No.	Settings: MJ Engine or Perkin-Elmer 9700
1	94°C 3 minutes

-78-

38	94°C 10 seconds
	60°C 30 seconds
	72°C 1 minute
1	72°C 10 minutes

Cycle No.	Settings: Eppendorf Mastercycler Gradient
1	94°C 3 minutes
38	94°C 15 seconds
	60°C 15 seconds
	72°C 1 minute
1	72°C 10 minutes

All of the compositions and methods disclosed and claimed herein can be made and
 5 executed in light of the present disclosure. While the compositions and methods of this
 invention have been described, it will be apparent to those of skill in the art that variations may
 be applied to the compositions and methods and in the steps or in the sequence of steps of the
 methods described herein. More specifically, it will be apparent that certain agents
 which are both chemically and physiologically related may be substituted for the agents
 10 described herein while the same or similar results would be achieved. The scope of the
 claims should not be limited by the preferred embodiments set forth herein, but should
 be given the broadest interpretation consistent with the description as a whole.

SEQUENCE LISTING

<110> MONSANTO TECHNOLOGY LLC

<120> METHODS FOR USING ARTIFICIAL POLYNUCLEOTIDES AND COMPOSITIONS
THEREOF TO REDUCE TRANSGENE SILENCING

<130> 1987-298

<140> 2,492,407

<141> July 10, 2003

<150> PCT/US03/21551

<151> July 10, 2003

<150> US 60/396,665

<151> July 18, 2002

<160> 35

<170> PatentIn version 3.3

<210> 1

<211> 515

<212> PRT

<213> Oryza sativa

<400> 1

Met	Ala	Ala	Thr	Met	Ala	Ser	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Val	Ser
1				5				10						15	

Leu	Asp	Gln	Ala	Val	Ala	Ala	Ser	Ala	Ala	Phe	Ser	Ser	Arg	Lys	Gln
		20						25					30		

Leu	Arg	Leu	Pro	Ala	Ala	Ala	Arg	Gly	Gly	Met	Arg	Val	Arg	Val	Arg
		35					40					45			

Ala	Arg	Gly	Arg	Arg	Glu	Ala	Val	Val	Val	Ala	Ser	Ala	Ser	Ser	Ser
	50					55					60				

Ser	Val	Ala	Ala	Pro	Ala	Ala	Lys	Ala	Glu	Glu	Ile	Val	Leu	Gln	Pro
65					70					75				80	

Ile	Arg	Glu	Ile	Ser	Gly	Ala	Val	Gln	Leu	Pro	Gly	Ser	Lys	Ser	Leu
			85						90					95	

Ser	Asn	Arg	Ile	Leu	Leu	Leu	Ser	Ala	Leu	Ser	Glu	Gly	Thr	Thr	Val
			100					105					110		

Val	Asp	Asn	Leu	Leu	Asn	Ser	Glu	Asp	Val	His	Tyr	Met	Leu	Glu	Ala
		115					120					125			

Leu	Lys	Ala	Leu	Gly	Leu	Ser	Val	Glu	Ala	Asp	Lys	Val	Ala	Lys	Arg
	130					135					140				

Ala	Val	Val	Val	Gly	Cys	Gly	Gly	Lys	Phe	Pro	Val	Glu	Lys	Asp	Ala
145					150					155					160

Lys	Glu	Glu	Val	Gln	Leu	Phe	Leu	Gly	Asn	Ala	Gly	Ile	Ala	Met	Arg
			165						170					175	

Ser	Leu	Thr	Ala	Ala	Val	Thr	Ala	Ala	Gly	Gly	Asn	Ala	Thr	Tyr	Val
			180					185						190	

Leu Asp Gly Val Pro Arg Met Arg Glu Arg Pro Ile Gly Asp Leu Val
 195 200 205
 Val Gly Leu Lys Gln Leu Gly Ala Asp Val Asp Cys Phe Leu Gly Thr
 210 215 220
 Glu Cys Pro Pro Val Arg Val Lys Gly Ile Gly Gly Leu Pro Gly Gly
 225 230 235 240
 Lys Val Lys Leu Ser Gly Ser Ile Ser Ser Gln Tyr Leu Ser Ala Leu
 245 250 255
 Leu Met Ala Ala Pro Leu Ala Leu Gly Asp Val Glu Ile Glu Ile Ile
 260 265 270
 Asp Lys Leu Ile Ser Ile Pro Tyr Val Glu Met Thr Leu Arg Leu Met
 275 280 285
 Glu Arg Phe Gly Val Lys Ala Glu His Ser Asp Ser Trp Asp Arg Phe
 290 295 300
 Tyr Ile Lys Gly Gly Gln Lys Tyr Lys Ser Pro Gly Asn Ala Tyr Val
 305 310 315 320
 Glu Gly Asp Ala Ser Ser Ala Ser Tyr Phe Leu Ala Gly Ala Ala Ile
 325 330 335
 Thr Gly Gly Thr Val Thr Val Gln Gly Cys Gly Thr Thr Ser Leu Gln
 340 345 350
 Gly Asp Val Lys Phe Ala Glu Val Leu Glu Met Met Gly Ala Lys Val
 355 360 365
 Thr Trp Thr Asp Thr Ser Val Thr Val Thr Gly Pro Pro Arg Glu Pro
 370 375 380
 Tyr Gly Lys Lys His Leu Lys Ala Val Asp Val Asn Met Asn Lys Met
 385 390 395 400
 Pro Asp Val Ala Met Thr Leu Ala Val Val Ala Leu Phe Ala Asp Gly
 405 410 415
 Pro Thr Ala Ile Arg Asp Val Ala Ser Trp Arg Val Lys Glu Thr Glu
 420 425 430
 Arg Met Val Ala Ile Arg Thr Glu Leu Thr Lys Leu Gly Ala Ser Val
 435 440 445
 Glu Glu Gly Pro Asp Tyr Cys Ile Ile Thr Pro Pro Glu Lys Leu Asn
 450 455 460
 Ile Thr Ala Ile Asp Thr Tyr Asp Asp His Arg Met Ala Met Ala Phe
 465 470 475 480
 Ser Leu Ala Ala Cys Ala Asp Val Pro Val Thr Ile Arg Asp Pro Gly
 485 490 495
 Cys Thr Arg Lys Thr Phe Pro Asn Tyr Phe Asp Val Leu Ser Thr Phe
 500 505 510

Val Arg Asn
515

<210> 2
<211> 1548
<212> DNA
<213> Oryza sativa

<400> 2
atggcggcga ccatggcgctc caacgccgcg gctgcggcgg cggtgtccct ggaccaggcc 60
gtggcggcgt cggcggcgtt ctcgtcgcgg aagcagctgc ggctgcccgc cgcggcgcgc 120
ggggggatgc ggggtgcgggt gcgggcgcgg gggcggcggg aggcgggtggt ggtggcgctcc 180
gcgtcgctcgt cgtcgggtggc agcgccggcg gcgaaggcgg aggagatcgt gctccagccc 240
atcagggaga tctccggggc gggttcagctg ccaggggtcca agtcgctctc caacaggatc 300
ctcctcctct cgcctctctc cgagggcaca acagtgggtg acaacttgct gaacagtgag 360
gatgttcact acatgcttga ggccctgaaa gccctcgggc tctctgtgga agcagataaa 420
gttgcaaaaa gagctgtagt cgttggctgt ggtggcaagt ttctgttga gaaggatgcg 480
aaagaggaag tgcaactctt cttggggaac gctggaatcg caatgcgac cttgacagca 540
gccgtgactg ctgctgggtg aaatgcaact tatgtgcttg atggagtgcc acgaatgagg 600
gagagaccga ttggtgactt gggtgtcggg ttgaaacaac ttggtgcgga tgtcgactgt 660
ttccttgga ctgaatgcc acctgttcgt gtcaaggga ttggaggact tctggtggc 720
aaggttaagc tctctggttc catcagcagt cagtacttga gtgccttgct gatggctgct 780
cctttggccc ttggggatgt ggagatcgaa atcattgaca aactaatctc cattccttac 840
gttgaaatga cattgagatt gatggagcgt tttgggtgtga aggcagagca ttctgatagt 900
tgggacagat tctatattaa gggagggcag aagtacaaat ctcttgaaa tgcctatgtt 960
gaaggatgat cctcaagcgc gagctatttc ttggctggtg ctgcaatcac tggaggcact 1020
gtgacagttc aaggttgtgg tacgaccagt ttgcagggtg atgtcaaatt tgctgaggta 1080
cttgagatga tgggagcaaa gggtacatgg actgacacca gtgtaaccgt aactggtcca 1140
ccacgtgagc cttatgggaa gaaacacctg aaagctgttg atgtcaacat gaacaaaatg 1200
cctgatgttg ccatgaccct tgccgttggt gcactcttcg ctgatggtec aactgctatc 1260
agagatgtgg cttcctggag agtaaaggaa accgaaagga tggttgcaat tcggaccgag 1320
ctaacaagc tgggagcatc gggtgaagaa ggtcctgact actgcatcat caccaccg 1380
gagaagctga acatcacggc aatcgacacc tacgatgatc acaggatggc catggccttc 1440
tcctcgctg cctgcgccga cgtgcccggtg acgatcaggg accctgggtg caccgcaag 1500
accttcccca actacttcga cgttctaagc actttcgtca ggaactga 1548

<210> 3
 <211> 1548
 <212> DNA
 <213> Oryza sativ

<400> 3
 atggctgcaa ctatggctag taacgcagcg gctgccgctg ccgtttcctt agaccaagca 60
 gtagcagcga gcgctgcatt ctcatcacgt aagcaactac ggctaccagc agccgctaga 120
 ggcggcatga gagttagagt gagggctaga ggtaggcggg aggctgtagt cgtagcctcc 180
 gcttctagca gttcggtggc tgcgccggct gctaaggcag aggagattgt ttacaacct 240
 attagggaaa tatcgggggc cgtacaatta cctggaagca agagcctttc caacaggatt 300
 ctgttgcttt cagctctctc ggagggaaca acagttgtgg ataactctgtt gaatagttag 360
 gatgtgcact atatgctaga ggctctcaag gctctagggc tttctgtaga agcggataaa 420
 gtagcaaaac gcgcagtggc ttaggttgtt ggtgggaagt tcccagttga aaaggatgct 480
 aaggaagaag tacagctctt tctcggaat gccgggatcg ccatgcggag ttgactgct 540
 gcggtcacag ccgctggagg caacgcaaca tacgtcctag atggggtgcc gagaatgcgt 600
 gagcgtccta ttggtgatct tgctgtaggt ctcaagcaac tcggcgctga cgtagattgt 660
 ttctgggta ctgagtgtcc gccagtcaga gttaaaggaa tcggtgggct gccgggcgga 720
 aaggtcaagc tgctgggcag tatttcgagt cagtatcttt ctgctctcct gatggctgcg 780
 ccattagctt tgggagatgt tgagatcgag atcattgata aacttatatc tatcccgtat 840
 gtcgagatga ctttaagact tatggaacgg tttgggggta aggccgagca tagcgacagt 900
 tgggatcggt tctacataaa gggaggccag aagtataagt ctctgggaa tgcttatgta 960
 gaaggggatg cttcatctgc gtcttacttc cttgcgggag cggctataac tggagggaaca 1020
 gtcacagttc agggctgcgg tacaacaagt ttgcaagggtg acgtgaagtt tgccgaggta 1080
 cttgaaatga tgggtgccaa agtaacgtgg acagacacat cggtgacagt tactggctct 1140
 ccacgagaac cttacggcaa aaagcatctt aaggccgtgg atgttaatat gaataagatg 1200
 cctgacgttg ctatgacact tgccgttggt gccctttttg cagacggccc aacggcgata 1260
 cgcgatgttg catcatggcg cgtcaaggaa acggagagga tgggtggctat tcgaactgaa 1320
 ctcaccaaac ttggtgcctc ttagaggag ggccctgatt actgtatcat tacaccccct 1380
 gagaaactta acatcactgc tattgatata tacgacgatc atagaatggc tatggctttc 1440
 tcactggccg cttgtgcaga tgttcctgtc acaatcagag atcctggctg tactagaaag 1500
 acgttcccga actactttga tgttctttca acattcgtcc gcaattga 1548

<210> 4
 <211> 1548
 <212> DNA
 <213> Oryza sativa

<400> 4
 atggctgcaa ctatggctag taacgcagcg gctgccgctg ccgtttcctt agaccaagca 60
 gtagcagcga gcgctgcatt ctcatcacgt aagcaactac ggctaccagc agccgctaga 120
 ggcgcatga gagttagagt gagggctaga ggtaggcggg aggctgtagt cgtagcctcc 180
 gcttctagca gttcgggtggc tgcgccggct gctaaggcag aggagattgt ttacaacct 240
 attagggaaa tatcgggggc cgtacaatta cctggaagca agagcctttc caacaggatt 300
 ctgttgcttt cagctctctc ggagggaaca acagttgtgg ataactctgtt gaatagtgag 360
 gatgtgcact atatgctaga ggctctcaag gctctagggc tttctgtaga agcggataaa 420
 gtagcaaaac gcgcagtggg tgtagggtgt ggtgggaagt tcccagttga aaaggatgct 480
 aaggaagaag tacagctctt tctcgggaat gccgggatcg ccatgcggag tttgactgct 540
 gcggtcacag ccgctggagg caacgcaaca tacgtcctag atgggggtgcc gagaatgcgt 600
 gagcgtccta ttggtgatct tgtcgtaggt ctcaagcaac tcggcgctga cgtagattgt 660
 ttcttgggta ctgagtgtcc gccagtcaga gttaaaggaa tcgggtgggct gccgggcgga 720
 aagggtcaagc tgtcgggcag tatttcgagt cagtatcttt ctgctctcct gatggctgcg 780
 ccattagctt tgggagatgt tgagatcgag atcattgata aacttatatc tatcccgtat 840
 gtcgagatga ctttaagact tatggaacgg tttgggggta aggccgagca tagcgacagt 900
 tgggatcggt tctacataaa gggaggccag aagtataagt ctcttgggaa tgcttatgta 960
 gaaggggatg cttcatctgc gtcttacttc cttgcgggag cggctataac tggaggaaca 1020
 gtcacagttc agggctgcgg tacaacaagt ttgcaagggt acgtgaagtt tgccgaggta 1080
 cttgaaatga tgggtgccaa agtaacgtgg acagacacat cggtgacagt tactggctct 1140
 ccacgagaac cttacggcaa aaagcatctt aaggccgtgg atgttaatat gaataagatg 1200
 cctgacgttg ctatgacact tgccgttggt gccctttttg cagacggccc aacggcgata 1260
 cgcgatgttg catcatggcg cgtcaaggaa acggagagga tgggtggctat tcgaactgaa 1320
 ctcaccaaac ttggtgcctc tgtagaggag ggccctgatt actgtatcat tacacccct 1380
 gagaaactta acatcactgc tattgataca tacgacgatc atagaatggc tatggctttc 1440
 tcactggccg cttgtgcaga tgttcctgtc acaatcagag atcctggctg tactagaaag 1500
 acgttcccga actactttga tgttctttca acattcgtcc gcaattga 1548

<210> 5
 <211> 525
 <212> PRT
 <213> Glycine max

<400> 5

Met	Ala	Gln	Val	Ser	Arg	Val	His	Asn	Leu	Ala	Gln	Ser	Thr	Gln	Ile	1	5	10	15
Phe	Gly	His	Ser	Ser	Asn	Ser	Asn	Lys	Leu	Lys	Ser	Val	Asn	Ser	Val	20	25	30	
Ser	Leu	Arg	Pro	Arg	Leu	Trp	Gly	Ala	Ser	Lys	Ser	Arg	Ile	Pro	Met	35	40	45	
His	Lys	Asn	Gly	Ser	Phe	Met	Gly	Asn	Phe	Asn	Val	Gly	Lys	Gly	Asn	50	55	60	
Ser	Gly	Val	Phe	Lys	Val	Ser	Ala	Ser	Val	Ala	Ala	Ala	Glu	Lys	Pro	65	70	75	80
Ser	Thr	Ser	Pro	Glu	Ile	Val	Leu	Glu	Pro	Ile	Lys	Asp	Phe	Ser	Gly	85	90	95	
Thr	Ile	Thr	Leu	Pro	Gly	Ser	Lys	Ser	Leu	Ser	Asn	Arg	Ile	Leu	Leu	100	105	110	
Leu	Ala	Ala	Leu	Ser	Glu	Gly	Thr	Thr	Val	Val	Asp	Asn	Leu	Leu	Tyr	115	120	125	
Ser	Glu	Asp	Ile	His	Tyr	Met	Leu	Gly	Ala	Leu	Arg	Thr	Leu	Gly	Leu	130	135	140	
Arg	Val	Glu	Asp	Asp	Lys	Thr	Thr	Lys	Gln	Ala	Ile	Val	Glu	Gly	Cys	145	150	155	160
Gly	Gly	Leu	Phe	Pro	Thr	Ser	Lys	Glu	Ser	Lys	Asp	Glu	Ile	Asn	Leu	165	170	175	
Phe	Leu	Gly	Asn	Ala	Gly	Ile	Ala	Met	Lys	Ser	Leu	Thr	Ala	Ala	Val	180	185	190	
Val	Ala	Ala	Gly	Gly	Asn	Ala	Ser	Tyr	Val	Leu	Asp	Gly	Val	Pro	Arg	195	200	205	
Met	Arg	Glu	Arg	Pro	Ile	Gly	Asp	Leu	Val	Ala	Gly	Leu	Lys	Gln	Leu	210	215	220	
Gly	Ala	Asp	Val	Asp	Cys	Phe	Leu	Gly	Thr	Asn	Cys	Pro	Pro	Val	Arg	225	230	235	240
Val	Asn	Gly	Lys	Gly	Gly	Leu	Pro	Gly	Gly	Lys	Val	Lys	Leu	Ser	Gly	245	250	255	
Ser	Val	Ser	Ser	Gln	Tyr	Leu	Thr	Ala	Leu	Leu	Met	Ala	Ala	Pro	Leu	260	265	270	
Ala	Leu	Gly	Asp	Val	Glu	Ile	Glu	Ile	Val	Asp	Lys	Leu	Ile	Ser	Val	275	280	285	

Pro Tyr Val Glu Met Thr Leu Lys Leu Met Glu Arg Phe Gly Val Ser
290 295 300

Val Glu His Ser Gly Asn Trp Asp Arg Phe Leu Val His Gly Gly Gln
305 310 315 320

Lys Tyr Lys Ser Pro Gly Asn Ala Phe Val Glu Gly Asp Ala Ser Ser
325 330 335

Ala Ser Tyr Leu Leu Ala Gly Ala Ala Ile Thr Gly Gly Thr Ile Thr
340 345 350

Val Asn Gly Cys Gly Thr Ser Ser Leu Gln Gly Asp Val Lys Phe Ala
355 360 365

Glu Val Leu Glu Lys Met Gly Ala Lys Val Thr Trp Ser Glu Asn Ser
370 375 380

Val Thr Val Ser Gly Pro Pro Arg Asp Phe Ser Gly Arg Lys Val Leu
385 390 395 400

Arg Gly Ile Asp Val Asn Met Asn Lys Met Pro Asp Val Ala Met Thr
405 410 415

Leu Ala Val Val Ala Leu Phe Ala Asn Gly Pro Thr Ala Ile Arg Asp
420 425 430

Val Ala Ser Trp Arg Val Lys Glu Thr Glu Arg Met Ile Ala Ile Cys
435 440 445

Thr Glu Leu Arg Lys Leu Gly Ala Thr Val Glu Glu Gly Pro Asp Tyr
450 455 460

Cys Val Ile Thr Pro Pro Glu Lys Leu Asn Val Thr Ala Ile Asp Thr
465 470 475 480

Tyr Asp Asp His Arg Met Ala Met Ala Phe Ser Leu Ala Ala Cys Gly
485 490 495

Asp Val Pro Val Thr Ile Lys Asp Pro Gly Cys Thr Arg Lys Thr Phe
500 505 510

Pro Asp Tyr Phe Glu Val Leu Glu Arg Leu Thr Lys His
515 520 525

<210> 6

<211> 1578

<212> DNA

<213> Glycine max

<400> 6

atggcccaag tgagcagagt gcacaatctt gctcaaagca ctcaaatttt tggccattct 60

tccaactcca acaaactcaa atcggtgaat tcggtttcat tgaggccacg cctttggggg 120

gcctcaaaat ctgcacatccc gatgcataaa aatggaagct ttatgggaaa ttttaatgtg 180

gggaagggaa attccggcgt gtttaagggt tctgcatcgg tcgccgccgc agagaagccg 240

tcaacgtcgc cggagatcgt gttggaaccc atcaaagact tctcgggtac catcacattg 300

```

ccaggggtcca agtctctgtc caatcgaatt ttgcttcttg ctgctctctc tgaggggaaca 360
actgttgtag acaacttggt gtatagttag gatattcatt acatgcttgg tgcattaagg 420
acccttggac tgcgtgtgga agatgacaaa acaaccaaac aagcaattgt tgaaggctgt 480
gggggattgt ttcccactag taaggaatct aaagatgaaa tcaatttatt ccttggaaat 540
gctggtatcg caatgaagtc cttgacagca gctgtggttg ctgcagggtg aaatgcaagc 600
tacgtacttg atgggggtgcc ccgaatgaga gagaggccaa ttggggattt gggtgctggt 660
cttaagcaac ttgggtgcaga tgttgattgc tttcttggca caaactgtcc acctgttcgt 720
gtaaatggga agggaggact tcctggcgga aaggtgaaac tgtctggatc agttagcagt 780
caatacttga ctgctttgct tatggcagct cctttagctc ttgggtgatgt ggaaattgag 840
attgttgata aactgatttc tgttccatat gttgaaatga ctctgaagtt gatggagcgt 900
tttgagagttt ctgtggaaca cagtggtaat tgggataggt tcttgggtcca tggaggtcaa 960
aagtacaagt ctcttggaac tgcttttggt gaaggtgatg cttcaagtgc cagttattta 1020
ctagctgggtg cagcaattac tgggtgggact atcactgtta atggctgtgg cacaagcagt 1080
ttacagggag atgtaaaatt tgctgaagtt cttgaaaaga tgggagctaa gggtacatgg 1140
tcagagaaca gtgtcactgt ttctggacca ccacgagatt tttctgggtcg aaaagtcttg 1200
cgaggcattg atgtcaatat gaacaagatg ccagatgttg ccatgacact tgctgttggt 1260
gcactatttg ctaatgggtcc cactgctata agagatgtgg caagttggag agttaaagag 1320
actgagagga tgatagcaat ctgcacagaa ctcagaaagc taggagcaac agttgaagaa 1380
ggtcctgatt actgtgtgat tactccacct gagaaattga atgtcacagc tatagacaca 1440
tatgatgacc acagaatggc catggcattc tctcttgctg cttgtgggga tgttccagta 1500
accatcaagg atcctgggtg caccaggaag acatttcctg actactttga agtccttgag 1560
aggttaacaa agcactaa 1578

```

<210> 7

<211> 1578

<212> DNA

<213> Glycine max

<400> 7

```

atggctcagg tctctcgcgt tcataatctc gctcagagta cccagatatt cggacattcc 60
agtaactcaa acaaactaaa gtctgtgaat agtgtatcac ttgggcctcg gctgtgggga 120
gcaagtaaga gccgtatccc tatgcacaag aacggttcgt tcatggggaa ctttaacgtc 180
ggcaaaggaa actcagggtg cttcaaagta agcgccagcg tagctgcggc tgagaagccc 240
agtacttctc ctgaaattgt tcttgaaccg ataaaggatt tctcaggtag gattacacta 300
cctggatcaa agagtctctc taatagaatt ttgttgctcg cagctctgtc cgaaggaacc 360

```

```

actgtagtcg ataacctcct ttatagcgaa gatatacatt atatgttggg ggcgctcaga 420
actcttgggc taagagttga ggacgataag actactaaac aagctatcgt cgaaggttgt 480
ggcgggttgt tccctacttc taaagaaagt aaagatgaga taaacttggt tcttggcaac 540
gcaggaatcg caatgaagag cctcacgct gctgtcgttg cggcgggtgg taacgctagt 600
tacgtcttag acggcgtgcc tagaatgca gaaagaccta tcggtgatct agtggctggc 660
ctaaaacagc ttggagcaga cgtcgattgt ttcttgggca caaattgccc gcccgtaga 720
gtgaacggga agggaggctt gccaggcggg aagggttaaac tatccggatc ggtctcgtca 780
cagtacctaa ctgcattgct catggccgcc ccgctcgctt tgggggacgt ggagattgaa 840
atcgtcgata agttgattag cgtgccttat gtggaaatga ccctcaaatt gatggagagg 900
ttcggagttt cggtagaaca ctccgggaat tgggatcggg ttcttgtaca cggagggcaa 960
aagtacaaaa gcccaggcaa tgccttcgtc gaaggggacg cttcgagcgc ttcctatctc 1020
ctcgctggcg cagccataac cgggtggcacc ataaccgtga acggctgcgg cacctcatcc 1080
cttcaagggtg atgtaaagtt cgctgaggtc ttggagaaaa tgggcgcaaa ggtcacatgg 1140
tctgagaaca gcgtaaccgt gtccggacct cccagagact ttcgtggtag aaaggtcctt 1200
aggggaatag atgtgaatat gaataagatg ccagatgtgg ctatgacgct cgctgttgtc 1260
gccctgttcg caaacggacc taccgcaata agggatgtcg cttcatggcg tgttaaggaa 1320
accgaacgga tgatcgctat ttgcaccgag ttgcgtaagc tgggtgcaac ggtggaagaa 1380
ggaccagact attgcgtgat aacacctcct gaaaagctca atgtgaccgc tattgacact 1440
tatgacgac acagaatggc tatggcattc tcacttgctg cttgcggtga cgtgccggtt 1500
acgatcaagg acccaggggtg tactaggaag acattcccag attactttga ggtgttgaa 1560
agattgacaa agcactga 1578

```

```

<210> 8
<211> 506
<212> PRT
<213> Zea mays

```

```

<400> 8

```

```

Met Ala Ala Met Ala Thr Lys Ala Ala Ala Gly Thr Val Ser Leu Asp
1          5          10          15

Leu Ala Ala Pro Ser Arg Arg His His Arg Pro Ser Ser Ala Arg Pro
          20          25          30

Pro Phe Arg Pro Ala Val Arg Gly Leu Arg Ala Pro Gly Arg Arg Val
          35          40          45

Ile Ala Ala Pro Pro Ala Ala Ala Ala Ala Ala Val Gln Ala Gly
50          55          60

```

Ala	Glu	Glu	Ile	Val	Leu	Gln	Pro	Ile	Lys	Glu	Ile	Ser	Gly	Thr	Val	
65					70					75					80	
Lys	Leu	Pro	Gly	Ser	Lys	Ser	Leu	Ser	Asn	Arg	Ile	Leu	Leu	Leu	Ala	
				85					90					95		
Ala	Leu	Ser	Glu	Gly	Thr	Thr	Val	Val	Asp	Asn	Leu	Leu	Asn	Ser	Glu	
			100					105					110			
Asp	Val	His	Tyr	Met	Leu	Gly	Ala	Leu	Arg	Thr	Leu	Gly	Leu	Ser	Val	
		115					120					125				
Glu	Ala	Asp	Lys	Ala	Ala	Lys	Arg	Ala	Val	Val	Val	Gly	Cys	Gly	Gly	
	130					135					140					
Lys	Phe	Pro	Val	Glu	Asp	Ala	Lys	Glu	Glu	Val	Gln	Leu	Phe	Leu	Gly	
145					150					155					160	
Asn	Ala	Gly	Ile	Ala	Met	Arg	Ser	Leu	Thr	Ala	Ala	Val	Thr	Ala	Ala	
				165					170					175		
Gly	Gly	Asn	Ala	Thr	Tyr	Val	Leu	Asp	Gly	Val	Pro	Arg	Met	Arg	Glu	
			180					185					190			
Arg	Pro	Ile	Gly	Asp	Leu	Val	Val	Gly	Leu	Lys	Gln	Leu	Gly	Ala	Asp	
			195				200					205				
Val	Asp	Cys	Phe	Leu	Gly	Thr	Asp	Cys	Pro	Pro	Val	Arg	Val	Asn	Gly	
	210					215					220					
Ile	Gly	Gly	Leu	Pro	Gly	Gly	Lys	Val	Lys	Leu	Ser	Gly	Ser	Ile	Ser	
225					230					235					240	
Ser	Gln	Tyr	Leu	Ser	Ala	Leu	Leu	Met	Ala	Ala	Pro	Leu	Ala	Leu	Gly	
				245					250					255		
Asp	Val	Glu	Ile	Glu	Ile	Ile	Asp	Lys	Leu	Ile	Ser	Ile	Pro	Tyr	Val	
			260					265					270			
Glu	Met	Thr	Leu	Arg	Leu	Met	Glu	Arg	Phe	Gly	Val	Lys	Ala	Glu	His	
		275					280					285				
Ser	Asp	Ser	Trp	Asp	Arg	Phe	Tyr	Ile	Lys	Gly	Gly	Gln	Lys	Tyr	Lys	
	290					295					300					
Ser	Pro	Lys	Asn	Ala	Tyr	Val	Glu	Gly	Asp	Ala	Ser	Ser	Ala	Ser	Tyr	
305					310					315					320	
Phe	Leu	Ala	Gly	Ala	Ala	Ile	Thr	Gly	Gly	Thr	Val	Thr	Val	Glu	Gly	
			325						330					335		
Cys	Gly	Thr	Thr	Ser	Leu	Gln	Gly	Asp	Val	Lys	Phe	Ala	Glu	Val	Leu	
			340					345					350			
Glu	Met	Met	Gly	Ala	Lys	Val	Thr	Trp	Thr	Glu	Thr	Ser	Val	Thr	Val	
		355					360					365				
Thr	Gly	Pro	Pro	Arg	Glu	Pro	Phe	Gly	Arg	Lys	His	Leu	Lys	Ala	Ile	
	370					375					380					

Asp Val Asn Met Asn Lys Met Pro Asp Val Ala Met Thr Leu Ala Val
385 390 395 400

Val Ala Leu Phe Ala Asp Gly Pro Thr Ala Ile Arg Asp Val Ala Ser
405 410 415

Trp Arg Val Lys Glu Thr Glu Arg Met Val Ala Ile Arg Thr Glu Leu
420 425 430

Thr Lys Leu Gly Ala Ser Val Glu Glu Gly Pro Asp Tyr Cys Ile Ile
435 440 445

Thr Pro Pro Glu Lys Leu Asn Val Thr Ala Ile Asp Thr Tyr Asp Asp
450 455 460

His Arg Met Ala Met Ala Phe Ser Leu Ala Ala Cys Ala Glu Val Pro
465 470 475 480

Val Thr Ile Arg Asp Pro Gly Cys Thr Arg Lys Thr Phe Pro Asp Tyr
485 490 495

Phe Asp Val Leu Ser Thr Phe Val Lys Asn
500 505

<210> 9
<211> 1521
<212> DNA
<213> Zea mays

<400> 9
atggcggccca tggcgaccaa ggccgcccgcg ggcaccgtgt cgctggacct cgccgcgccg 60
tcgcgcgcc accaccgccc gagctcggcg cgcccgcct tccgccccgc cgtccgcggg 120
ctgcgggcgc ctgggcgccg cgtgatcgcc gcgccgccgg cggcggcagc ggcggcggcg 180
gtgcaggcgg gtgccgagga gatcgtgctg cagcccatca aggagatctc cggcaccgtc 240
aagctgccgg ggtccaagtc gctttccaac cggatcctcc tactcgccgc cctgtccgag 300
gggacaacag tggttgataa cctgctgaac agtgaggatg tccactacat gctcggggcc 360
ttgaggactc ttggtctctc tgtcgaagcg gacaaagctg ccaaaagagc tgtagttggt 420
ggctgtggtg gaaagttccc agttgaggat gctaaagagg aagtgcagct cttcttgggg 480
aatgctggaa tcgcaatgcg gtccttgaca gcagctgtta ctgctgctgg tggaaatgca 540
acttacgtgc ttgatggagt accaagaatg agggagagac ccattggcga cttggttgtc 600
ggattgaagc agcttggtgc agatgttgat tgtttccttg gcactgactg cccacctgtt 660
cgtgtcaatg gaatcggagg gctacctggt ggcaaggtea agctgtctgg ctccatcagc 720
agtcagtact tgagtgcctt gctgatggct gtcctttgg ctcttgggga tgtggagatt 780
gaaatcattg ataaattaat ctccattccg tacgtcgaaa tgacattgag attgatggag 840
cgttttggtg tgaaagcaga gcattctgat agctgggaca gattctacat taaggagggt 900
caaaaataca agtcccctaa aaatgcctat gttgaagggt atgcctcaag cgcaagctat 960

ttcttggctg gtgctgcaat tactggaggg actgtgactg tggaagggtg tggcaccacc	1020
agtttgcagg gtgatgtgaa gtttgctgag gtactggaga tgatgggagc gaagggtaca	1080
tggaccgaga ctagcgtaac tgttactggc ccaccgcggg agccatttgg gaggaacac	1140
ctcaaggcga ttgatgtcaa catgaacaag atgcctgatg tcgccatgac tcttgctgtg	1200
gttgccctct ttgccgatgg cccgacagcc atcagagacg tggcttcctg gagagtaaag	1260
gagaccgaga ggatggttgc gatccggacg gagctaacca agctgggagc atctgttgag	1320
gaagggccgg actactgcat catcacgccg ccggagaagc tgaacgtgac ggcgatcgac	1380
acgtacgacg accacaggat ggccatggcc ttctcccttg ccgcctgtgc cgagggtccc	1440
gtcaccatcc gggaccctgg gtgcacccgg aagaccttcc ccgactactt cgatgtgctg	1500
agcactttcg tcaagaatta a	1521

<210> 10

<211> 1521

<212> DNA

<213> Zea mays

<400> 10

atggcggtta tggccacgaa ggcagcggcc ggtacagtaa gcctcgattt ggccggcccc	60
tcccgtaggc accaccggcc aagcagtgcg aggccaccgt tcaggccagc agttcgcggt	120
cttagagcgc ctggtagaag ggttatcgca gcgccaccgg cggctgccgc tgcggcagcg	180
gtgcaggccg gcgcggaaga gatcgtccta cagcctatca aggaaatctc tggtagcgta	240
aagttaccag gcagcaaaag tcttagcaac cgaatcctgc tgttggcggc actctctgaa	300
gggaccacgg tcgtagataa tctgctcaac agcgaagacg tgcactatat gttgggtgcc	360
ctgaggacgc taggtctgtc agtggaaagg gataaggccg ccaagcgcgc tgtcgtcgtt	420
ggctgcggcg gtaagttccc cgtggaggac gcgaaagaag aggtgcagtt atttcttggg	480
aacgctggca tcgccatgcg gtcccttacc gcagccgtca ccgctgcggg aggcaacgca	540
acttacgtgc ttgacggtgt tcctcgtatg agagagcggc ccatagggga tctcgtcgtg	600
gggctcaagc agctcggggc cgacgttgat tgcttcctcg gaaccgactg cccccctgtg	660
aggggtgaacg gcacgggggg actgccagga ggcaaagtca agttgtccgg ctcaatttcc	720
tcgcagtacc tgagtgcctt gcttatggcg gccctcttgg ctctgggaga cgtcgaaatt	780
gagatcattg ataagctgat ctctatccct tatgttgaga tgacactccg tctgatggaa	840
agattcgggg tcaaagctga gcactccgat tcctgggaca ggttctatat caagggcgga	900
cagaaatata agtcaccgaa gaatgcgtac gtcgagggag acgcatcgag cgcgagttac	960
ttccttgccg gcgctgccat caccggggga accgtgacag tggaaggctg tgggacaacg	1020

agcttgcagg gcgacgtcaa atttgctgag gtgctagaaa tgatggggcg taaggtgact 1080
 tggactgaga cgtccgtgac cgttacggga ccgccccgcg aacctttcgg ccggaagcat 1140
 ctgaaagcga ttgatgtgaa catgaataag atgccggacg tcgctatgac acttgccgtg 1200
 gtggccctgt tcgctgacgg cccacccgca atcagggatg tcgctagttg gaggggtcaag 1260
 gagacagagc gtatggtggc gatccgaacg gagctgacta aactcggggc cagtgtggag 1320
 gagggcccgg attactgcat aatcacacct ccagagaagt tgaacgtcac cgctatcgac 1380
 acatacgacg atcaccggat ggcaatggcc tttagcttgg cagcgtgcgc cgaagtacct 1440
 gtgactataa gagatccagg ttgcacccgc aaaacgtttc ccgactattt cgacgtcctc 1500
 tcaaccttcg tgaagaactg a 1521

<210> 11
 <211> 76
 <212> PRT
 <213> Arabidopsis thaliana

<400> 11

Met	Ala	Gln	Val	Ser	Arg	Ile	Cys	Asn	Gly	Val	Gln	Asn	Pro	Ser	Leu
1				5					10					15	
Ile	Ser	Asn	Leu	Ser	Lys	Ser	Ser	Gln	Arg	Lys	Ser	Pro	Leu	Ser	Val
			20					25					30		
Ser	Leu	Lys	Thr	Gln	Gln	His	Pro	Arg	Ala	Tyr	Pro	Ile	Ser	Ser	Ser
		35					40					45			
Trp	Gly	Leu	Lys	Lys	Ser	Gly	Met	Thr	Leu	Ile	Gly	Ser	Glu	Leu	Arg
	50					55					60				
Pro	Leu	Lys	Val	Met	Ser	Ser	Val	Ser	Thr	Ala	Cys				
65					70					75					

<210> 12
 <211> 228
 <212> DNA
 <213> Arabidopsis thaliana

<400> 12

atggcgcaag ttagcagaat ctgcaatggg gtgcagaacc catctcttat ctccaatctc 60
 tcgaaatcca gtcaacgcaa atctccctta tcggtttctc tgaagacgca gcagcatcca 120
 cgagcttata cgatttcgtc gtcgtgggga ttgaagaaga gtgggatgac gttaattggc 180
 tctgagcttc gtctctttaa ggtcatgtct tctgtttcca cggcgtgc 228

<210> 13
 <211> 228
 <212> DNA
 <213> Arabidopsis thaliana

<400> 13

atggcccagg taagtaggat ctgtaacgga gtccaaaacc cttcactaat atcgaacctg 60

tcaaaaagct ctcaaagaaa gtcgccgctt tctgtatcgt tgaaaactca acagcaccgc 120
 agggcttata ccatctcaag ctctgggggt ctaaagaaaa gtggaatgac actgatcggt 180
 agcgaactac gaccgctgaa agtcatgtcc tcagtcagca ctgcgtgc 228

<210> 14
 <211> 228
 <212> DNA
 <213> Arabidopsis thaliana

<400> 14
 atggcgcaag taagtagaat ctgcaacggc gtgcagaacc cgtcgctgat ctccaacctc 60
 agcaagtcca gccagcggaa gtcgccgctc tcggtcagcc tcaagacca acagcaccgc 120
 agggcctacc ctatcagctc atcctggggc ctcaagaaga gtggcatgac gctgatcggc 180
 agcgagctgc ggccactcaa ggtgatgtcc tcggtctcaa cggcgtgc 228

<210> 15
 <211> 455
 <212> PRT
 <213> Agrobacterium tumefaciens

<400> 15

Met	Leu	His	Gly	Ala	Ser	Ser	Arg	Pro	Ala	Thr	Ala	Arg	Lys	Ser	Ser	1	5	10	15
Gly	Leu	Ser	Gly	Thr	Val	Arg	Ile	Pro	Gly	Asp	Lys	Ser	Ile	Ser	His	20	25	30	
Arg	Ser	Phe	Met	Phe	Gly	Gly	Leu	Ala	Ser	Gly	Glu	Thr	Arg	Ile	Thr	35	40	45	
Gly	Leu	Leu	Glu	Gly	Glu	Asp	Val	Ile	Asn	Thr	Gly	Lys	Ala	Met	Gln	50	55	60	
Ala	Met	Gly	Ala	Arg	Ile	Arg	Lys	Glu	Gly	Asp	Thr	Trp	Ile	Ile	Asp	65	70	75	80
Gly	Val	Gly	Asn	Gly	Gly	Leu	Leu	Ala	Pro	Glu	Ala	Pro	Leu	Asp	Phe	85	90	95	
Gly	Asn	Ala	Ala	Thr	Gly	Cys	Arg	Leu	Thr	Met	Gly	Leu	Val	Gly	Val	100	105	110	
Tyr	Asp	Phe	Asp	Ser	Thr	Phe	Ile	Gly	Asp	Ala	Ser	Leu	Thr	Lys	Arg	115	120	125	
Pro	Met	Gly	Arg	Val	Leu	Asn	Pro	Leu	Arg	Glu	Met	Gly	Val	Gln	Val	130	135	140	
Lys	Ser	Glu	Asp	Gly	Asp	Arg	Leu	Pro	Val	Thr	Leu	Arg	Gly	Pro	Lys	145	150	155	160
Thr	Pro	Thr	Pro	Ile	Thr	Tyr	Arg	Val	Pro	Met	Ala	Ser	Ala	Gln	Val	165	170	175	

Lys Ser Ala Val Leu Leu Ala Gly Leu Asn Thr Pro Gly Ile Thr Thr
 180 185 190
 Val Ile Glu Pro Ile Met Thr Arg Asp His Thr Glu Lys Met Leu Gln
 195 200 205
 Gly Phe Gly Ala Asn Leu Thr Val Glu Thr Asp Ala Asp Gly Val Arg
 210 215 220
 Thr Ile Arg Leu Glu Gly Arg Gly Lys Leu Thr Gly Gln Val Ile Asp
 225 230 235 240
 Val Pro Gly Asp Pro Ser Ser Thr Ala Phe Pro Leu Val Ala Ala Leu
 245 250 255
 Leu Val Pro Gly Ser Asp Val Thr Ile Leu Asn Val Leu Met Asn Pro
 260 265 270
 Thr Arg Thr Gly Leu Ile Leu Thr Leu Gln Glu Met Gly Ala Asp Ile
 275 280 285
 Glu Val Ile Asn Pro Arg Leu Ala Gly Gly Glu Asp Val Ala Asp Leu
 290 295 300
 Arg Val Arg Ser Ser Thr Leu Lys Gly Val Thr Val Pro Glu Asp Arg
 305 310 315 320
 Ala Pro Ser Met Ile Asp Glu Tyr Pro Ile Leu Ala Val Ala Ala Ala
 325 330 335
 Phe Ala Glu Gly Ala Thr Val Met Asn Gly Leu Glu Glu Leu Arg Val
 340 345 350
 Lys Glu Ser Asp Arg Leu Ser Ala Val Ala Asn Gly Leu Lys Leu Asn
 355 360 365
 Gly Val Asp Cys Asp Glu Gly Glu Thr Ser Leu Val Val Arg Gly Arg
 370 375 380
 Pro Asp Gly Lys Gly Leu Gly Asn Ala Ser Gly Ala Ala Val Ala Thr
 385 390 395 400
 His Leu Asp His Arg Ile Ala Met Ser Phe Leu Val Met Gly Leu Val
 405 410 415
 Ser Glu Asn Pro Val Thr Val Asp Asp Ala Thr Met Ile Ala Thr Ser
 420 425 430
 Phe Pro Glu Phe Met Asp Leu Met Ala Gly Leu Gly Ala Lys Ile Glu
 435 440 445
 Leu Ser Asp Thr Lys Ala Ala
 450 455

<210> 16

<211> 1368

<212> DNA

<213> Agrobacterium tumefaciens

<400> 16

atgcttcacg gtgcaagcag ccggcccgca accgcccgca aatcctctgg cctttccgga

60

```

accgtccgca ttcccggcga caagtcgata tcccaccggt ctttcattgtt cggcgggtctc 120
gcgagcgggtg aaacgcgcat caccggcctt ctggaaggcg aggacgtcat caatacgggc 180
aaggccatgc aggcgatggg cgcccgcatc cgtaaggaag gcgacacctg gatcatcgat 240
ggcgtcggca atggcggcct cctggcgctt gaggcggcgc tcgatttcgg caatgccggc 300
acgggctgcc gcctgacgat gggcctcgtc ggggtctacg atttcgacag caccttcata 360
ggcgacgcct cgctcacaaa gcgcccgatg ggccgcgtgt tgaaccgcct gcgcgaaatg 420
ggcgtgcagg tgaaatcgga agacgggtgac cgtcttcccc ttaccttgcg cgggccgaag 480
acgccgacgc cgatcaccta ccgcgtgccg atggcctccg cacagggtgaa gtccgccgtg 540
ctgctcgccg gcctcaacac gcccggcata acgacgggtca tcgagccgat catgacgcgc 600
gatcatacgg aaaagatgct gcagggcttt ggcgccaacc ttaccgtcga gacggatgcg 660
gacggcgtgc gcaccatccg cctggaaggc cgcggaagc tcaccggcca agtcatcgac 720
gtgccggggc acccgctctc gacggccttc ccgctggttg cggccctgct tgttccgggc 780
tccgacgtca ccatactcaa cgtgctgatg aaccccaccc gcaccggcct catcctgacg 840
ctgcaggaaa tgggcggcga catcgaagtc atcaaccgcg gccttgccgg cggcgaagac 900
gtggcggacc tgcgcgttcg ctctccacg ctgaaggcg tcacgggtgcc ggaagaccgc 960
gcgccttcga tgatcgacga atatccgatt ctgctgtcg ccgccgcctt cgcggaagg 1020
gcgaccgtga tgaacggtct ggaagaactc cgcgtcaagg aaagcgaccg cctctcggcc 1080
gtcgccaatg gcctcaagct caatggcgtg gattgcgatg agggcgagac gtcgctcgtc 1140
gtgcgtggcc gccctgacgg caaggggctc ggcaacgcct cgggcggcgc cgtcgccacc 1200
catctcgatc accgcacgc catgagcttc ctgctcatgg gcctcgtgtc ggaaaaccct 1260
gtcacgggtg acgatgccac gatgatgcc acgagcttcc cggagtcat ggacctgatg 1320
gccgggctgg gcgcgaagat cgaactctcc gatacgaagg ctgcctga 1368

```

<210> 17

<211> 1368

<212> DNA

<213> *Agrobacterium tumefaciens*

<400> 17

```

atgcttcatg gagcttcata taggccagct actgccagga agtctagcgg gctcagtggc 60
accgtgcgca tccctggcga taaaagtatt tcacacagga gcttcatgtt cggaggactt 120
gctagtggag agacgagaat cactgggttg cttgaggcg aagatgttat caacaccgg 180
aaggcgatgc aagcaatggg tgccagaatc cgaaaagagg gcgatacgtg gatcatcgac 240
ggtgttggtg acggaggatt gctcgctccc gaagcgccac ttgactttgg gaacgcagct 300

```

acgggggtgcc gtcttactat gggactggta ggcgtgtatg actttgactc taccttcatc 360
 ggtgacgcga gcctcactaa gagaccaatg ggacgagtgc tgaatcccct gagggagatg 420
 ggtgtccagg tgaaatctga ggatggatgat cgtcttccgg ttactctgcg aggccccaag 480
 acccccacgc caatcacgta caggggttccg atggcgctcag cacagggtcaa gtcagcggta 540
 ctcttggcgg gcctcaacac acctggaatc acaaccgtga ttgaacccat catgactaga 600
 gaccacacgg agaagatggt gcaggggtttc ggcgctaata taacgggtcga aaccgacgcc 660
 gacggcgtga ggacaatccg cttggagggc agaggtaaac tgactggcca agtcatcgat 720
 gtgcctggag atccctcgtc cacagcgttt cccctcgtag ctgcgttgct cgtccctgga 780
 tctgatgtga cgatcctgaa tgcctcatg aatccaacta gaaccggcct catcctcaca 840
 ttgcaggaga tgggtgctga catcgagggt atcaatccta gggtggcagg tggagaggat 900
 gtggccgata tgcgcgtgcg ttctagtaca ctcaaaggcg tgaccgtccc tgaggatcgc 960
 gctccatcca tgatcgacga gtaccccatc ctgcgccgtt ctgctgcgtt tgccgagggc 1020
 gcaactgtaa tgaacggcct tgaggagttg agggttaagg agagtgcag gctgtccgcg 1080
 gtggcgaatg gcctgaagct aaacggcgtg gactgcgacg aagggtgaaac gtcccttgta 1140
 gtccgtggtc gccagacgg gaaggggttg gggaatgctt cgggagctgc tgtggcgacg 1200
 caccttgatc atagaatcgc catgtcattt ctggatgatg gacttgtctc cgagaatccg 1260
 gtgaccgttg acgatgctac catgatcgcc acctcctttc ctgagttcat ggacctcatg 1320
 gcaggcttgg gggccaagat cgagctgtct gataactaagg ccgcttga 1368

<210> 18

<211> 1368

<212> DNA

<213> *Agrobacterium tumefaciens*

<400> 18

atgctacacg gtgcaagcag ccggccggca accgctcgca aatcttccgg cctttcggga 60
 acggtcagga ttccggggca taagtccata tcccaccggg cgttcatggt cggcgggtctt 120
 gccagcgggtg agacgcgcat cacgggcctg cttgaagggtg aggacgtgat caataccggg 180
 aaggccatgc aggctatggg agcgcgtatc cgcaagggaag gtgacacatg gatcattgac 240
 ggcgttggga atggcgggtc gctcgcccct gagggccctc tcgacttcgg caatgcggcg 300
 acgggctgca ggctcactat gggactggtc ggggtgtacg acttcgatag cacgttcatc 360
 ggagacgcct cgctcacaaa gcgcccgaatg ggccgcgttc tgaaccggtt gcgcgagatg 420
 ggcgtacagg tcaaataccga ggatggatgac cgtttgcccg ttacgctgcg cgggccgaag 480
 acgcctaccc cgattaccta ccgcgtgcca atggcatccg cccagggtcaa gtcagccgtg 540
 ctctcgccg gactgaacac tccgggcata accacgggtga tcgagcccat catgaccagg 600

gatcataccg aaaagatgct tcagggggttt ggcgccaacc tgacgggtcga gacggagcgt 660
 gacggcggtca ggaccatccg ccttgagggc aggggtaaac tgactggcca agtcatcgat 720
 gttccgggag acccgtcgtc cacggccttc ccgttggttg cggcgctgct cgtgccgggg 780
 agtgacgtga ccatactgaa cgtcctcatg aacccgacca ggaccggcct gatcctcacg 840
 cttcaggaga tgggagccga catcgaggtg atcaaccgcg gcctggcagg cgggtgaagac 900
 gttgcggatc tgcgcggtcg ctctctacc ctgaagggcg tgacgggtccc ggaagatcgc 960
 gcgcccgtcca tgatagacga gtatcctatt ctggccgctg ccgctgcgtt cgccgaaggg 1020
 gccacgggtca tgaacgggtct tgaggaactc cgcgtgaagg aatcggatcg cctgtcggcg 1080
 gtggccaatg gcctgaagct caacggtggt gactgcgacg agggtgagac ctcaactcgtg 1140
 gtccgtggcc ggctgatgg caagggcctc ggcaacgcca gtggagcggc cgtcgccacg 1200
 cacctcgatc atcgcatcgc gatgtccttc ttggtgatgg gtctcgtctc agagaacccg 1260
 gtgaccgtcg atgacgccac gatgatagcg acgagcttcc cagagttcat ggatctgatg 1320
 gcgggcctcg gggccaagat cgaactgtct gacacgaagg ccgcttga 1368

<210> 19

<211> 183

<212> PRT

<213> Streptomyces hygroscopicus

<400> 19

Met	Ser	Pro	Glu	Arg	Arg	Pro	Ala	Asp	Ile	Arg	Arg	Ala	Thr	Glu	Ala
1				5					10					15	
Asp	Met	Pro	Ala	Val	Cys	Thr	Ile	Val	Asn	His	Tyr	Ile	Glu	Thr	Ser
			20					25					30		
Thr	Val	Asn	Phe	Arg	Thr	Glu	Pro	Gln	Glu	Pro	Gln	Asp	Trp	Thr	Asp
		35					40					45			
Asp	Leu	Val	Arg	Leu	Arg	Glu	Arg	Tyr	Pro	Trp	Leu	Val	Ala	Glu	Val
	50					55					60				
Asp	Gly	Glu	Val	Ala	Gly	Ile	Ala	Tyr	Ala	Gly	Pro	Trp	Lys	Ala	Arg
65					70					75				80	
Asn	Ala	Tyr	Asp	Trp	Thr	Ala	Glu	Ser	Thr	Val	Tyr	Val	Ser	Pro	Arg
			85						90					95	
His	Gln	Arg	Thr	Gly	Leu	Gly	Ser	Thr	Leu	Tyr	Thr	His	Leu	Leu	Lys
			100					105					110		
Ser	Leu	Glu	Ala	Gln	Gly	Phe	Lys	Ser	Val	Val	Ala	Val	Ile	Gly	Leu
		115					120					125			
Pro	Asn	Asp	Pro	Ser	Val	Arg	Met	His	Glu	Ala	Leu	Gly	Tyr	Ala	Pro
	130					135					140				

Arg Gly Met Leu Arg Ala Ala Gly Phe Lys His Gly Asn Trp His Asp
 145 150 155 160

Val Gly Phe Trp Gln Leu Asp Phe Ser Leu Pro Val Pro Pro Arg Pro
 165 170 175

Val Leu Pro Val Thr Glu Ile
 180

<210> 20
 <211> 552
 <212> DNA
 <213> Streptomyces hygroscopicus

<400> 20
 atgagcccag aacgacgccc ggccgacatc cgccgtgcc a ccgaggcgga catgccggcg 60
 gtctgcacca tcgtcaacca ctacatcgag acaagcacgg tcaacttcg taccgagccg 120
 caggaaccgc aggactggac ggacgacctc gtccgtctgc gggagcgcta tccctggctc 180
 gtcgccgagg tggacggcga ggtcgccggc atcgccctac cgggcccctg gaaggcacgc 240
 aacgcctacg actggacggc cgagtcgacc gtgtacgtct ccccccgcca ccagcggacg 300
 ggactgggct ccacgctcta caccacctg ctgaagtccc tggaggcaca gggcttcaag 360
 agcgtggctg ctgtcatcgg gctgcccac gacccgagcg tgcgcatgca cgaggcgctc 420
 ggatatgccc cccgcggcat gctgcggggc gccggcttca agcacgggaa ctggcatgac 480
 gtgggtttct ggcagctgga cttcagcctg ccggtaccgc cccgtccggt cctgcccgtc 540
 accgagatct ga 552

<210> 21
 <211> 552
 <212> DNA
 <213> Streptomyces hygroscopicus

<400> 21
 atgagtccag aaaggagacc ggctgatatt cggagagcca ccgaagctga tatgcctgct 60
 gtttgtacaa tcgtaaacca ttatatcgag acctcgacag ttaattttcg cactgagccg 120
 caggagccac aggattggac ggacgatctg gtacgtttta gagaacgtta tccgtggcta 180
 gttgctgagg ttgacggaga agtcgctggc atagcttacg ctggaccgtg gaaagctcgt 240
 aacgcttacg actggacagc agaatccact gtctacgtca gccctcgtca tcaaagaacc 300
 ggattagggg gcacgttgta cactcatctt ttaaagtcac tggaggcaca aggcttcaag 360
 tctgttggtg cagttattgg attgccaaac gatccgagtg ttcgaatgca cgaagcgctt 420
 ggatacgctc cacgaggtat gctccgtgct gccggattca aacatggaaa ttggcacgac 480
 gtaggttttt ggcaactgga cttttcactt cccgttcccc ctagacctgt acttccagtt 540
 actgaaatct ag 552

<210> 22
 <211> 552
 <212> DNA
 <213> *Streptomyces hygroscopicus*

<400> 22
 atgtcgctg agcgccgtcc tgctgacata agacgcgcta ccgaggcaga catgcctgct 60
 gtttgcacca ttgtgaatca ctacatcgag acatctacgg taaacttccg cactgagcct 120
 caagaaccgc aggattggac cgacgatctc gtgcgtctca gagagcgta tccgtggctg 180
 gttgcagagg tggacgggta agtggctggg atcgccctacg ctggaccgtg gaaggctaga 240
 aacgcatacg attggactgc ggagtccaca gtctacgtct caccagaca tcaaagaacc 300
 gggctcggct cgaccctcta tacgcctctc ctcaagtcct tagaggcgca gggcttcaaa 360
 tctgtagtgg cggatgatcg cttgccaaac gatcccagtg tgagaatgca cgaggcactc 420
 gggtacgctc ctagaggaat gctcagggcg gctggattca agcacggtaa ttggcacgac 480
 gttggcttct ggcaactgga cttctctttg ccagttccac ctgcctctgt gctaccgctc 540
 accgaaatct ag 552

<210> 23
 <211> 1368
 <212> DNA
 <213> *Agrobacterium tumefaciens*

<400> 23
 atgcttcacg gtgcaagcag ccgtccagca actgctcgta agtcctctgg tctttctgga 60
 accgtccgta ttccagggtga caagtctatc tcccacaggt ccttcatggt tggagggtctc 120
 gctagcgggtg aaactcgtat caccggctctt ttggaagggtg aagatgttat caacactgggt 180
 aaggctatgc aagctatggg tgccagaatc cgtaaggaag gtgatacttg gatcattgat 240
 ggtgttggtg acggtggact ccttgctcct gaggctcctc tcgatttcgg taacgctgca 300
 actgggttgc gtttgactat gggctcttgtt ggtgtttacg atttcgatag cactttcatt 360
 ggtgacgctt ctctcactaa gcgtccaatg ggtcgtgtgt tgaaccact tcgcgaaatg 420
 ggtgtgcagg tgaagtctga agacgggtgat cgtcttccag ttaccttgcg tggaccaaag 480
 actccaacgc caatcaccta cagggtacct atggcttccg ctcaagtga gtccgctggt 540
 ctgcttgctg gtctcaacac cccagggtatc accactgtta tcgagccaat catgactcgt 600
 gaccacactg aaaagatgct tcaagggtttt ggtgctaacc ttaccgttga gactgatgct 660
 gacgggtgtgc gtaccatccg tcttgaagggt cgtggtaagc tcaccgggtca agtgattgat 720
 gttccagggtg atccatcctc tactgctttc ccattgggtg ctgccttgct tgttccagggt 780
 tccgacgtca ccatacctaa cgttttgatg aaccaaccc gtactgggtct catcttgact 840
 ctgcaggaaa tgggtgccga catcgaagtg atcaaccac gtcttgctgg tggagaagac 900

gtggctgact tgcgtgttcg ttcttctact ttgaagggtg ttactgttcc agaagaccgt 960
gctccttcta tgatcgacga gtatccaatt ctgctgttg cagctgcatt cgctgaagg 1020
gctaccgtta tgaacgggtt ggaagaactc cgtgttaagg aaagcgaccg tctttctgct 1080
gtcgcaaacg gtctcaagct caacgggtgtt gattgcatg aaggtgagac ttctctcgtc 1140
gtgctgtgtc gtctgacgg taagggtctc ggtaacgctt ctggagcagc tgtcgtacc 1200
cacctcgatc accgtatcgc tatgagcttc ctggttatgg gtctcgtttc tgaaaaccct 1260
gttactgttg atgatgctac tatgatcgt actagcttcc cagagttcat ggatttgatg 1320
gctgggtcttg gagctaagat cgaactctcc gacactaagg ctgcttga 1368

<210> 24
<211> 16
<212> DNA
<213> Artificial sequence

<220>
<223> DNA Primer molecule

<400> 24
catggagctt catcta 16

<210> 25
<211> 16
<212> DNA
<213> Artificial sequence

<220>
<223> DNA Primer molecule

<400> 25
gcctttgagt gtacta 16

<210> 26
<211> 16
<212> DNA
<213> Artificial sequence

<220>
<223> DNA Primer molecule

<400> 26
gggagcgcgt atccgc 16

<210> 27
<211> 16
<212> DNA
<213> Artificial sequence

<220>
<223> DNA Primer molecule

<400> 27
ggatggtcac gtcact 16

<210>	28	
<211>	16	
<212>	DNA	
<213>	Artificial sequence	
<220>		
<223>	DNA Primer molecule	
<400>	28	
	cggcatcacg acggtc	16
<210>	29	
<211>	16	
<212>	DNA	
<213>	Artificial sequence	
<220>		
<223>	DNA Primer molecule	
<400>	29	
	ggcatcgtcc accgtg	16
<210>	30	
<211>	16	
<212>	DNA	
<213>	Artificial sequence	
<220>		
<223>	DNA Primer molecule	
<400>	30	
	gcaactgggtt gccgtt	16
<210>	31	
<211>	16	
<212>	DNA	
<213>	Artificial sequence	
<220>		
<223>	DNA Primer molecule	
<400>	31	
	atcacctgga acatca	16
<210>	32	
<211>	22	
<212>	DNA	
<213>	Zea mays	
<400>	32	
	cgtcaagatc ctcttcacct cg	22
<210>	33	
<211>	22	
<212>	DNA	
<213>	Zea mays	
<400>	33	
	acaccctctc caacactctc ta	22

<210> 34
 <211> 9
 <212> PRT
 <213> Artificial sequence

<220>
 <223> Motif providing glyphosate resistance to a plant EPSPS

<400> 34

Gly Asn Ala Gly Ile Ala Met Lys Ser
 1 5

<210> 35
 <211> 1596
 <212> DNA
 <213> Agrobacterium tumefaciens

<400> 35
 atggcccaag ttagccgaat ctgcaacggt gtgcagaatc catcactaat ctccaacctg 60
 tccaaatcgt cacaacgtaa gtcgccatta tctgttagct tgaagactca gcaacatcct 120
 cgcgcatatc ctatatcaag cagttgggggt ttgaagaaat cgggtatgac cttgattggt 180
 tcggaactta ggccattgaa ggtgatgtct tcagttagta cagcttgcat gcttcacggt 240
 gcttcttcca gacccgcaac ggctagaaag agttctggct tgtctggaac cgtccgtatt 300
 ccaggagaca aaagcattag tcaccgctct ttcattgttg gtgggctggc atctggagag 360
 acgcgcatca ctggtcttct ggaaggagag gacgtcatca atacagggaa ggcaatgcag 420
 gctatgggtg cccgtattcg caaggaagggt gatacttgga tcatagacgg agttgggaac 480
 ggtggcttac ttgcaccgga ggctcctctc gactttggca acgcagccac aggggtgtaga 540
 cttactatgg gcctcgtggg tgtttacgat ttcgattcaa cctttattgg ggatgcctct 600
 ctcactaaac gcccaatggg aagagtcctt aacccggtga gggagatggg cgtacaagtt 660
 aagtccgagg acggcgacag attgcccgtc accttgccgc gccctaagac acccaccctt 720
 attacttaca gggttccaat ggcatctgct caagtgaagt ccgcagttct gctcgtgga 780
 ttgaacacac cgggtattac taccgtgatt gagccgatca tgactcgtga ccacactgag 840
 aagatgcttc agggtttcgg tgctaacctc accgttgaaa cagacgcgga cgggtgtgagg 900
 accattcgcc tggaggggaag gggaaaactc actggtcaag tcattgacgt gcccggtgat 960
 ccctccagca cggcgttccc actgggtgcc gctcttctcg taccaggctc cgatgtgaca 1020
 attctaaacg tcctcatgaa tcctactaga accggattga tacttacatt gcaggaaatg 1080
 ggtgctgata ttgaagttat caatcctaga ctagccggag gtgaggacgt agctgatttg 1140
 cgggtgaggt cttctacatt gaaagggtgt accgtacctg aagatagggc accttcaatg 1200
 attgacgagt atccaattct tgccgtcgcg gctgcctttg ctgagggcgc gaccgtgatg 1260

aatggactag aggagttgag agtgaaggaa tccgacagat tgagcgcagt cgctaacgga 1320
cttaaactca atggcgttga ttgtgatgag ggtgagacta gcttggtagt ccgtgggcga 1380
ccagacggaa agggtttggg caacgcttcg ggtgctgccg ttgcaactca cttggatcat 1440
cggatagcga tgagttttct ggtgatgggt ctcgtaagcg agaatcctgt gacagtcgac 1500
gatgcaacta tgatcgctac ttccttcctt gagtttatgg atttaatggc aggactaggt 1560
gcaaagattg aactctctga taccaaagcg gcctaa 1596

WHAT IS CLAIMED IS:

1. A polynucleotide molecule selected from the group consisting of: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:17, SEQ ID NO:18 and SEQ ID NO:35.
2. A DNA construct comprising: a promoter molecule that functions in plants, operably linked to said polynucleotide molecule of claim 1.
3. A plant cell comprising the DNA construct of claim 2.
4. The plant cell of claim 3, wherein said plant cell is selected from the group consisting of wheat, corn, rice, soybean, cotton, potato, canola, turf grass, forest tree, grain sorghum, vegetable crop, ornamental plants, forage crop, and fruit crop cells.
5. A plant cell comprising the polynucleotide molecule of claim 1.
6. A DNA molecule comprising: a polynucleotide molecule that specifically hybridizes under stringent conditions to the polynucleotide molecule of claim 1, wherein said DNA molecule is selected from the group consisting of SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27 and said stringent conditions comprise 6.0 x SSC at about 45°C followed by a wash of about 0.2-2.0 x SSC at a temperature in the range of about 22°C to 65°C.
7. A plant cell comprising the DNA molecule as recited in claim 6.
8. A pair of DNA molecules comprising a first DNA molecule and a second DNA molecule, wherein the first DNA molecule is SEQ ID NO:24 or its complement and the second DNA molecule is SEQ ID NO:25 or its complement and the pair of DNA molecules when used in a DNA amplification method produces an amplicon diagnostic for SEQ ID NO:17.
9. A plant cell wherein said amplicon of claim 8 is produced from genomic DNA extracted from said plant cell.

10. A pair of DNA molecules comprising a first DNA molecule and a second DNA molecule, wherein the first DNA molecule is SEQ ID NO:26 or its complement and the second DNA molecule is SEQ ID NO:27 or its complement and the pair of DNA molecules when used in a DNA amplification method produces an amplicon diagnostic for SEQ ID NO:18.

11. A plant cell wherein said amplicon of claim 10 is produced from genomic DNA extracted from said plant cell.

1/52

Figure 1A
Figure 1B
Figure 1C
Figure 1D
Figure 1E
Figure 1F
Figure 1G
Figure 1H

Figure 1

Figure 3A
Figure 3B
Figure 3C
Figure 3D
Figure 3E
Figure 3F
Figure 3G

Figure 3

Figure 5A
Figure 5B
Figure 5C
Figure 5D
Figure 5E
Figure 5F
Figure 5G

Figure 5

Figure 2A
Figure 2B
Figure 2C
Figure 2D
Figure 2E
Figure 2F

Figure 2

Figure 4A
Figure 4B
Figure 4C

Figure 4

1	OsEPSPS_Nat	ATGGCGGGCGA	CCATGGCGGC	CAACGCCGCG	GCTGCGGCGG	CGGTGTCCCT	50
	OsEPSPS_AT	T..A.	.T.....TAG	T.....A...	C..T.	.C..T...T.	
	OsEPSPS_ZM	C..T.	.G.....TAG	T.....G..A	.C..T.....	.T..T..G..	
51	OsEPSPS_Nat	GGACCAGGCC	GTGGCGGCGT	CGGCGGCGTT	CTCGTCGCGG	AAGCAGCTGC	100
	OsEPSPS_AT	A.....A..A	..A..A...A	GC..T..A...	...A..A..T	A..A.	
	OsEPSPS_ZM	A..T.....G	..T..T..CA	GC..T..C...	...CAGTA...	..A...T..A.	
101	OsEPSPS_Nat	GGCTGCCCGC	CGCGGCGCGC	GGGGGGATGC	GGGTGCGGGT	GCGGGCGCGG	150
	OsEPSPS_AT	A..A..	A..C..TA..A	..C..C...A	.A..TA..A..	.A...TA..A	
	OsEPSPS_ZM	.T..C..G..	G..T..C..A	..T..A...A	A.....C...	T..A..CA..A	
151	OsEPSPS_Nat	GGGCGGCGGG	AGGCGGTGGT	GGTGGCGTCC	GCGTCGTCGT	CGTCGGTGGC	200
	OsEPSPS_AT	..TA.....	T..A..	C..A..C...	..T..TAGCA	GT.....	
	OsEPSPS_ZM	..CA..A..T.	C.....	A..C..C..G	..CAGC....	.CAGT..T..	

Figure 1A

201	OsEPSPS_Nat	AGCGCCGGCG	GCGAAGGCGG	AGGAGATCGT	GCTCCAGCCC	ATCAGGGAGA	250
	OsEPSPS_AT	T.....T	A.	T..	TT.A..A..T	..T...A.	
	OsEPSPS_ZM	T.....A	C.	.A..A..T..	T..G..A..G	..AC...A.	
251	OsEPSPS_Nat	TCTCCGGGGC	GGTTCAGCTG	CCAGGGTCCA	AGTCGCTCTC	CAACAGGATC	300
	OsEPSPS_AT	.A..G.....	C..A..AT.A	..T..AAG..	..AGC..T..	T	
	OsEPSPS_ZM	.T..G..C..	...G.....C	..T..C..G.	C..G..	G..TC.T...	
301	OsEPSPS_Nat	CTCCTCCTCT	CCGCCCTCTC	CGAGGGCACA	ACAGTGGTGG	ACAACTTGCT	350
	OsEPSPS_AT	..GT.G..T.	.A..T.....	G.....A...	T....	..TC..T.	
	OsEPSPS_ZM	..T..G..A.	:G..G..G..	...A..G..C	..G..T..A.	..TC..T.	
351	OsEPSPS_Nat	GAACAGTGAG	GATGTTCACT	ACATGCTTGA	GGCCCTGAAA	GCCCTCGGGC	400
	OsEPSPS_AT	...T.....	G....	.T.....A..	...T..C..G	..T..A....	
	OsEPSPS_ZM	TCC..A	..C..G....	C..	A..G..A...	..GT.G..CT	

Figure 1B

401	OsEPSPS_Nat	TCTCTGTGGA	AGCAGATAAA	GTTGCAAAA	GAGCTGTAGT	CGTTGGCTGT	450
	OsEPSPS_AT	.T....A..	...G.....	..A.....C	.C..A..G..	T..A..T...	
	OsEPSPS_ZM	.GAGC..T..	G..T..C...	..G..C..GC	A..C..	A..G..A..C	
451	OsEPSPS_Nat	GGTGGCAAGT	TTCCTGTTGA	GAAGGATGCG	AAAGAGGAAG	TGCAACTCTT	500
	OsEPSPS_AT	G....	.C..A.....	A.....T	..G..A....	.A..G.....	
	OsEPSPS_ZM	..C..T....	.C..A..G..	..A..C..C	A..G.	.T....T.G..	
501	OsEPSPS_Nat	CTTGGGGAAC	GCTGGAATCG	CAATGCGATC	CTTGACAGCA	GCCGTGACTG	550
	OsEPSPS_AT	TC.C.....	T..C..G....	.C.....GAG	T.....T..T	..G..C..A.	
	OsEPSPS_ZM	A...	C....	A..AG	.C.....T..C	..G..C.....	
551	OsEPSPS_Nat	CTGCTGGTGG	AAATGCAACT	TATGTGCTTG	ATGGAGTGCC	ACGAATGAGG	600
	OsEPSPS_AT	.C.....A..	C..C.....A	..C..C..A.	G.....	GA.....C.T	
	OsEPSPS_ZM	.C..A..G..	T..C.....	..C..C..G.	.C..T..T..	..G....C.C	

Figure 1C

601	OsEPSPS_Nat	GAGAGACCGA	TTGGTGACTT	GGTTGTCGGG	TTGAAACAAC	TTGGTGCGGA	650
	OsEPSPS_AT	. . . C . T . . T .	 TC .	T . . C . . A . . T	C . C . . G	. C . . C . . T . .	
	OsEPSPS_ZM	. . AC C .	. C TC .	C . . G . . A . . A	C . T T	. G . . A . . C . .	
651	OsEPSPS_Nat	TGTCGACTGT	TTCCTTGGCA	CTGAATGCCC	ACCTGTTTCGT	GTC AAGGGA	700
	OsEPSPS_AT	C . . A . . T . . .	 G . . T .	 G . . T .	G . . A . . CA . A	. . T . . A	
	OsEPSPS_ZM	. . . A . . T . . .	. . . T . G . . G .	. A . . G	T . . A . . G . . G	. . A C .	
701	OsEPSPS_Nat	TTGGAGGACT	TCCTGGTGGC	AAGGTTAAGC	TCTCTGGTTC	CATCAGCAGT	750
	OsEPSPS_AT	. C . . T . . G . .	G . . G . . C . . A	 C	. G . . G . . CAG	T . . TTCG . . .	
	OsEPSPS_ZM	. C . . G . . T . .	C . . A . . C . . A	 G . . . T	. A . . A . . GAG	T . . . TC . TCG	
751	OsEPSPS_Nat	CAGTACTTGA	GTGCCCTTGCT	GATGGCTGCT	CCTTTGGCCC	TTGGGGATGT	800
	OsEPSPS_AT	 TC . TT	C TC . C . .	 G	. . A . . A . . TT	. G . . A	
	OsEPSPS_ZM	 C . CT	CC . . GC . . T .	 A . . A	. AC GT	. G . . C . . C . .	

Figure 1D

801	OsEPSPS_Nat	GGAGATCGAA	ATCATTGACA	AACTAATCTC	CATTCCCTTAC	GTTGAAATGA	850
	OsEPSPS_AT	T.....G	T.	T..A..	T..C..G..T	..C..G.....	
	OsEPSPS_ZM	T..A..T...	..T..C.....	..G..C..A..	...C..G..T	..G..G.....	
851	OsEPSPS_Nat	CATTGAGATT	GATGGAGCGT	TTTGGTGTGA	AGGCAGAGCA	TTCTGATAGT	900
	OsEPSPS_AT	.T..A...C.	T.....A..G	G..T.	C.....	..AGC..C...	
	OsEPSPS_ZM	..C..C.GC.	AA.A	A....	..A..T..A..	..AG...CTC.	
901	OsEPSPS_Nat	TGGGACAGAT	TCTATATTAA	GGGAGGGCAG	AAGTACAAAT	CTCCTGGAAA	950
	OsEPSPS_AT	TC.T.	C..A..	C...	T..G.	G..	
	OsEPSPS_ZM	TC.G.	C..C..	C..A...	..A..T..GA	G...A..C..	
951	OsEPSPS_Nat	TGCCTATGTT	GAAGGTGATG	CCTCAAGCGC	GAGCTATTTC	TTGGCTGGTG	1000
	OsEPSPS_AT	...T.....A	G....	...T...TCT..	..TCT..C...	C.T..G..A.	
	OsEPSPS_ZM	...G..C..G	..G..C..C.	TTC...	C.....C....	A..C..	

Figure 1E

OsEPSPS_Nat	1001	CTGCAATCAC	TGGAGGCACT	GTGACAGTTC	AAGTTGTGG	TACGACCAGT	1050
OsEPSPS_AT		.G..T..A..	A..A	.C.....	.G..C..C..	...A..A...	
OsEPSPS_ZM		.A..T..T..	C.....T...	.T.....G.		A..A..G..C	
OsEPSPS_Nat	1051	TTGCAGGGTG	ATGTCAAATT	TGCTGAGGTA	CTTGAGATGA	TGGGAGCAAA	1100
OsEPSPS_AT		A....	.C..G..G..	...C.....	A....	T..C..	
OsEPSPS_ZM		C.....A..C.	.C..A..G..	C..A..A..C	A....	C..T..	
OsEPSPS_Nat	1101	GGTTACATGG	ACTGACACCA	GTGTAACCGT	AACTGGTCCA	CCACGTGAGC	1150
OsEPSPS_AT		A..A..G...	..A.....AT	CG..G..A..	T.....T	A..A.	
OsEPSPS_ZM		...G..T...	..A..T..GT	CA..C.....	C.....G..T	..T..G..A.	
OsEPSPS_Nat	1151	CTTATGGGAA	GAAACACCTG	AAAGCTGTTG	ATGTCAACAT	GAACAAAATG	1200
OsEPSPS_AT		...C..C..	A..G..T..T	..G..C..G.	...T..T..	...T..G...	
OsEPSPS_ZM		.C..C..T..	T...	..G..A..G.	.C..G.....	...T..G...	

Figure 1F

OsEPSPS_Nat	1201	CCTGATGTTG	CCATGACCCCT	TGCCGTTGTT	GCACTCTTCG	CTGATGGTCC	1250
OsEPSPS_AT		 C	 A		 C T	 A C	
OsEPSPS_ZM		 A C	 G	 C A C	 T G	 A C A	
OsEPSPS_Nat	1251	AACTGCTATC	AGAGATGTGG	CTTCCTGGAG	AGTAAAGGAA	ACCGAAAGGA	1300
OsEPSPS_AT		 G G A C	 T A A C	 C C	 G C	 G G	
OsEPSPS_ZM		 T C	 T G C T	 AGT G C	 G C G T G A		
OsEPSPS_Nat	1301	TGGTTGCAAT	TCGGACCGAG	CTAACAAAGC	TGGGAGCATC	GGTTGAAGAA	1350
OsEPSPS_AT		 G T A T A	 A T A C C A	 C C A T T C	 T T C T A G	 T A G G	
OsEPSPS_ZM		 C T C	 C C A T G A	 T G A C T AG C G			
OsEPSPS_Nat	1351	GGTCCTGACT	ACTGCATCAT	CACCCACCG	GAGAAGCTGA	ACATCACGGC	1400
OsEPSPS_AT		 C T T	 T T A C T	 T A A A T	 T T A T	 T T	
OsEPSPS_ZM		 A .	 A T C T C	 A C	 A C	 A C	

Figure 1G

OsEPSPS_Nat	1401	AATCGACACC	TACGATGATC	ACAGGATGGC	CATGGCCCTTC	TCCCTCGCTG	1450
OsEPSPS_AT		T..T..T..A	C....	.T..A.....	T.....T...	..A..G..C.	
OsEPSPS_ZM		T..A.....T	C....	A.....	G.....G...	..G.....A.	
OsEPSPS_Nat	1451	CCTGCGCCGA	CGTGCCCGTG	ACGATCAGGG	ACCCTGGTTG	CACCCGCAAG	1500
OsEPSPS_AT		.T..T..A..	T..T..T..C	..A.....A.	.T.....C..	T..TA.A...	
OsEPSPS_ZM		.G.....G..	...T..G..T	T..A.	.T.....G..	...T..T..A	
OsEPSPS_Nat	1501	ACCTTCCCCA	ACTACTTCGA	CGTTCTAAGC	ACTTTCGTCA	GGAACTGA	1548
OsEPSPS_AT		..G.....G.	T..	T.....TTCA	..A.....C	.C..T...	
OsEPSPS_ZM		..T.....T.	.T.....	T..G...TCA	..C..T..AC	.C.....	

Figure 1H

1	ZmEPSPS_Nat	ATGGCGGCCA	TGGCGACCAA	GGCCGCCGCG	GGCACCGTGT	CGCTGGACCT	50
	ZmEPSPS_ZM	T.	C..G..	...A..G..C	..T..A..AA	GC..C..TT.	
51	ZmEPSPS_Nat	CGCCGCGCCG	TCGCGCCGCC	ACCACCGCCC	GAGCTCGGCG	CGCCCGCCCT	100
	ZmEPSPS_ZM	G..G..C..C	..C..TA.G.	G..	A...AGT...	A.G..A..G.	
101	ZmEPSPS_Nat	TCCGCCCCGC	CGTCCGCGGG	CTGCGGGCGC	CTGGGCGCCG	CGTGATCGCC	150
	ZmEPSPS_ZM	..A.G..A..	A..T.....T	..TA.A....	TA.AA.	G..T.....A	
151	ZmEPSPS_Nat	GCGCCGCCGG	CGGCGGCAGC	GGCGGCGGCG	GTGCAGGCGG	GTGCCGAGGA	200
	ZmEPSPS_ZM	A....	T..C..	T.....A....	C..	.C..G..A..	
201	ZmEPSPS_Nat	GATCGTGCTG	CAGCCCATCA	AGGAGATCTC	CGGCACCGTC	AAGCTGCCGG	250
	ZmEPSPS_ZM	C..A	T....	A.....	T..T..G..A	...T.A..A.	

Figure 2A

251	ZmEPSPS_Nat	GGTCCAAGTC	GCTTTCCAAC	CGGATCCTCC	TACTCGCCGC	CCTGTCCGAG	300
	ZmEPSPS_ZM	.CAG...AAG	T...AG....	..A.....G.	.GT.G..G..	A..C..T..A	
301	ZmEPSPS_Nat	GGGACAACAG	TGGTTGATAA	CCTGCTGAAC	AGTGAGGATG	TCCACTACAT	350
	ZmEPSPS_ZM	C..G.	.C..A.....	T.....C...	..C..A..C.	.G.....T..	
351	ZmEPSPS_Nat	GCTCGGGGCC	TTGAGGACTC	TTGGTCTCTC	TGTCGAAGCG	GACAAAGCTG	400
	ZmEPSPS_ZM	.T.G..T...C	G.	.A.....G..	A..G.....C	..T..G..C.	
401	ZmEPSPS_Nat	CCAAAAGAGC	TGTAGTTGTT	GGCTGTGGTG	GAAAGTTCCC	AGTTGAGGAT	450
	ZmEPSPS_ZM	GC.C..	...C.C...	C..C.	T.....C..	C..G.....C	
451	ZmEPSPS_Nat	GCTAAAGAGG	AAGTGCAGCT	CTTCTTGGGG	AATGCTGGAA	TCGCAATGCG	500
	ZmEPSPS_ZM	..G.....A.	.G.....T.	A..TC.T...	..C.....C.	C.....	

Figure 2B

501	ZmEPSPS_Nat	GTCCTTGACA	GCAGCTGTTA	CTGCTGCTGG	TGGAAATGCA	ACTTACGTGC	550
	ZmEPSPS_ZM	...C.T..C	...C..C.	.C....G..	A..C..C...		
	Consensus	...c.g..a	...c..c.	.c....g..	a..a..c...		
551	ZmEPSPS_Nat	TTGATGGAGT	ACCAAGAATG	AGGGAGAGAC	CCATTGGCGA	CTTGGTTGTC	600
	ZmEPSPS_ZM	...C..T..	T..TC.T...	.A...C.G.	A..G..	TC.C..C..G	
601	ZmEPSPS_Nat	GGATTGAAGC	AGCTTGGTGC	AGATGTTGAT	TGTTTCCTTG	GCACTGACTG	650
	ZmEPSPS_ZM	..GC.C....	C..G..	C..C.....	.C.....C.	.A..C.....	
651	ZmEPSPS_Nat	CCCACCTGTT	CGTGTCAATG	GAATCGGAGG	GCTACCTGGT	GGCAAGGTCA	700
	ZmEPSPS_ZM	...C.....G	A.G..G..C.	.C....G..	A..G..A..A	A....	
701	ZmEPSPS_Nat	AGCTGTCTGG	CTCCATCAGC	AGTCAGTACT	TGAGTGCCTT	GCTGATGGCT	750
	ZmEPSPS_ZM	..T....C..	...A..TTC.	TCG.....C	C.	T.....G	

Figure 2C

751	ZmEPSPS_Nat	GCTCCTTTGG	CTCTTGGGGA	TGTGGAGATT	GAAATCATTG	ATAAATTAAT	800
	ZmEPSPS_ZM	..C...C...	G..A..	C..C..A...	..G.....	GC.G..	
801	ZmEPSPS_Nat	CTCCATTCCG	TACGTCGAAA	TGACATTGAG	ATTGATGGAG	CGTTTTGGTG	850
	ZmEPSPS_ZM	...T..C..T	..T..T..G.	C.CC.	TC.....	A A.A..C..G.	
851	ZmEPSPS_Nat	TGAAAGCAGA	GCATTCTGAT	AGCTGGGACA	GATTCTACAT	TAAGGGAGGT	900
	ZmEPSPS_ZM	.C.....T..	...C..C...	TC.....	..G.....T..	C.....C..A	
901	ZmEPSPS_Nat	CAAAAATACA	AGTCCCCTAA	AAATGCCTAT	GTTGAAGGTG	ATGCCTCAAG	950
	ZmEPSPS_ZM	..G.....T.	A..G..	G.....G..C	..C..G..A.	.C..A..G..	
951	ZmEPSPS_Nat	CGCAAGCTAT	TTCTTGGCTG	GTGCTGCAAT	TACTGGAGGG	ACTGTGACTG	1000
	ZmEPSPS_ZM	...G..T..C	...C.T..G.	.C.....C..	C..C..G..A	..C.....A.	

Figure 2D

ZmEPSPS_Nat	1001	TGGAAGGTTG	TGGCACCACC	AGTTTGCAGG	GTGATGTGAA	GTTTGCTGAG	1050
ZmEPSPS_ZM		C..	...G..A..G	..C.....	.C..C..C..	A.....	
ZmEPSPS_Nat	1051	GTACTGGAGA	TGATGGGAGC	GAAGGTTACA	TGGACCGAGA	CTAGCGTAAC	1100
ZmEPSPS_ZM		..G..A..A.	C..	T.....G..T	T....	.GTC...G..	
ZmEPSPS_Nat	1101	TGTTACTGGC	CCACCGCGGG	AGCCATTTGG	GAGGAAACAC	CTCAAGGCGA	1150
ZmEPSPS_ZM		C.....G..A	..G..C..C.	.A..T..C..	CC....G..T	..G..A....	
ZmEPSPS_Nat	1151	TTGATGTCAA	CATGAACAAG	ATGCCTGATG	TCGCCATGAC	TCTTGCTGTG	1200
ZmEPSPS_ZM		G..	T...	G..C.	T.....	A.....C...	
ZmEPSPS_Nat	1201	GTTGCCCTCT	TTGCCGATGG	CCCGACAGCC	ATCAGAGACG	TGGCTTCCTG	1250
ZmEPSPS_ZM		..G.....G.	.C..T..C..	...C..C..A	G..T.	.C...AGT..	

Figure 2E

1251 1300
ZmEPSPS_Nat GAGAGTAAAG GAGACCGAGA GGATGGTTGC GATCCGGACG GAGCTAACCA
ZmEPSPS_ZM ...G..C... ..A...C .T.....G.. ..A... ..G..T..

1301 1350
ZmEPSPS_Nat AGCTGGGAGC ATCTGTTGAG GAAGGGCCGG ACTACTGCAT CATCAGGCCG
ZmEPSPS_ZM .A..C..G.. CAG...G... ..G..C.... .T.....A.....A..T

1351 1400
ZmEPSPS_Nat CCGGAGAAGC TGAACGTGAC GGCGATCGAC ACGTACGACG ACCACAGGAT
ZmEPSPS_ZM ..A.....T C.. C..T.....A..... .T...C....

1401 1450
ZmEPSPS_Nat GGCCATGGCC TTCTCCCTTG CCGCCTGTGC CGAGGTCCCC GTCACCATCC
ZmEPSPS_ZM ...A..... ..TAG.T.G. .A..G..C.. ..A..A..T ..G..T..AA

1451 1500
ZmEPSPS_Nat GGGACCCCTGG GTGCACCCGG AAGACCTTCC CCGACTACTT CGATGTGCTG
ZmEPSPS_ZM .A..T..A.. T.....C ..A..G..T. T.. ..C..C..C

1501 1521
ZmEPSPS_Nat AGCACTTTTCG TCAAGAATTA A
ZmEPSPS_ZM TCA..C.... .G.....C.G .

Figure 2F

1	GmEPSPS_Nat	ATGGCCCAAG	TGAGCAGAGT	GCACAATCTT	GCTCAAAGCA	CTCAAATTTT	50
	GmEPSPS_GM	T..G.	.CTCTC.C..	T..T.....C	G..T.	.C..G..A..	
51	GmEPSPS_Nat	TGGCCATTCT	TCCAACCTCCA	ACAAACTCAA	ATCGGTGAAT	TCGGTTTTCAT	100
	GmEPSPS_GM	C..A.....C	AGT.....A.	A..	G..T.....	AGT..A....C	
101	GmEPSPS_Nat	TGAGGCCACG	CCTTTGGGGG	GCCTCAAAAT	CTCGCATCCC	GATGCATAAA	150
	GmEPSPS_GM	.TC....T..	G..G.....A	..AAGT..GA	GC..T.....	T.....C..G	
151	GmEPSPS_Nat	AATGGAAGCT	TTATGGGAAA	TTTTAATGTG	GGGAAGGGAA	ATTCCGGCGT	200
	GmEPSPS_GM	..C..TTCG.	.C.:...G..	C.....C..C	..C..A.....	.C..A..T..	
201	GmEPSPS_Nat	GTTTAAGGTT	TCTGCATCGG	TCGCCGCCGC	AGAGAAGCCG	TCAACGTCGC	250
	GmEPSPS_GM	C..C..A..A	AGC..CAGC.	.A..T..G..	T.....C	AGT..T..T.	

Figure 3A

251	GmEPSPS_Nat	CGGAGATCGT	GTTGGAACCC	ATCAAAGACT	TCTCGGGTAC	CATCACATTG	300
	GmEPSPS_GM	.T..A..T..	TC.T.....G	.A..G..T.	A.....	G..T...C.A	
301	GmEPSPS_Nat	CCAGGGTCCA	AGTCTCTGTC	CAATCGAATT	TTGCTTCTTG	CTGCTCTCTC	350
	GmEPSPS_GM	..T..A..A.	..AG...C..	T...A.....	...T.G..C.	.A.....G..	
351	GmEPSPS_Nat	TGAGGGAACA	ACTGTTGTAG	ACAACTTGTT	GTATAGTGAG	GATATTCAAT	400
	GmEPSPS_GM	C..A.....C	A..C.	.T...C.CC.	T.....C..A	A....	
401	GmEPSPS_Nat	ACATGCTTGG	TGCATTAAAGG	ACCCTTGGAC	TGCGTGTGGA	AGATGACAAA	450
	GmEPSPS_GM	.T...T.G..	G..GC.C..A	.T.....G.	..AA.A..T..	G...C..T..G	
451	GmEPSPS_Nat	ACAACCAAAC	AAGCAATTGT	TGAAGGCTGT	GGGGGATTGT	TTCCCACTAG	500
	GmEPSPS_GM	..T..T.....	T..C..	C.....T...	..C..G.....	.C..T...TC	

Figure 3B

GmEPSPS_Nat	501	TAAGGAATCT	AAAGATGAAA	TCAATTTATT	CCTTGGAAT	GCTGGTATCG	550
GmEPSPS_GM		...A...AG.	G.	.A..C..G..	T.....C..C	..A..A.....	
GmEPSPS_Nat	551	CAATGAAGTC	CTTGACAGCA	GCTGTGGTTG	CTGCAGGTGG	AAATGCAAGC	600
GmEPSPS_GM		AG	.C.C..C..T	C.....	.G..G.....	T..C..T..T	
GmEPSPS_Nat	601	TACGTACTTG	ATGGGGTGCC	CCGAATGAGA	GAGAGGCCAA	TTGGGGATTT	650
GmEPSPS_GM		CT.A.	.C..C.....	TA.....C..	..A..A..T.	.C..T...C.	
GmEPSPS_Nat	651	GGTTGCTGGT	CTTAAGCAAC	TTGGTGCAGA	TGTTGATTGC	TTTCTTGGCA	700
GmEPSPS_GM		A..G.....C	..A..A..G.	A.....	C..C.....T	..CT.G.....	
GmEPSPS_Nat	701	CAAACTGTCC	ACCTGTTTCGT	GTAAATGGGA	AGGGAGGACT	TCCTGGCGGA	750
GmEPSPS_GM		T..C..	G..C..GA.A	..G..C.....	CT.	G..A.....T	

Figure 3C

GmEPSPS_Nat	AAGGTGAAAC	TGTCTGGATC	AGTTAGCAGT	CAATACTTGA	CTGCTTTTGCT	800
GmEPSPS_GM	T.....	A..C.....	G..CTCGTCA	..G...C.A.	A.....	
GmEPSPS_Nat	TATGGCAGCT	CCTTTAGCTC	TTGGTGATGT	GGAAATTGAG	ATTGTTGATA	850
GmEPSPS_GM	C.....C..C	..GC.C...T	.G..G..C..	..G.....A	..C..C.....	
GmEPSPS_Nat	AACTGATTTC	TGTTCCATAT	GTTGAAATGA	CTCTGAAGTT	GATGGAGCGT	900
GmEPSPS_GM	.GT.....AG	C..G..T...	..G.....	.C..C..A..	A.G	
GmEPSPS_Nat	TTTGGAGTTT	CTGTGGAACA	CAGTGGTAAT	TGGGATAGGT	TCTTGGTCCA	950
GmEPSPS_GM	..C.....	.G..A.....	.TCC..G...	C...	.TC.T..A...	
GmEPSPS_Nat	TGGAGGTCAA	AAGTACAAGT	CTCCTGGCAA	TGCTTTTGT	GAAGGTGATG	1000
GmEPSPS_GM	C.....G...	AA	GC..A.....	...C..C..C	G..C..	

Figure 3D

GmEPSPS_Nat	1001	CTTCAAGTGC	CAGTTATTTA	CTAGCTGGTG	CAGCAATTAC	TGGTGGGACT	1050
GmEPSPS_GM		G..C..	TTCC...C.C	.C.....C.	C..A..	C.....C..C	
GmEPSPS_Nat	1051	ATCACTGTTA	ATGGCTGTGG	CACAAGCAGT	TTACAGGGAG	ATGTAAAATT	1100
GmEPSPS_GM		.A..C..G.	.C.....C..	...CTCATCC	C.T..A..T.	G...	
GmEPSPS_Nat	1101	TGCTGAAGTT	CTTGAAAAGA	TGGGAGCTAA	GGTTACATGG	TCAGAGAACA	1150
GmEPSPS_GM		C.....G..C	T.G..G..A.	C..A..	...C.....	...T.....	
GmEPSPS_Nat	1151	GTGTCACGT	TTCTGGACCA	CCACGAGATT	TTTCTGGTCG	AAAAGTCTTG	1200
GmEPSPS_GM		.C..A..C..	G..C.....T	..CA....C.	..CG....A.	...G...C.T	
GmEPSPS_Nat	1201	CGAGGCATTG	ATGTCAATAT	GAACAAGATG	CCAGATGTTG	CCATGACACT	1250
GmEPSPS_GM		A.G..A..A.	G.....	...T.....	G.	.T.....G...	

Figure 3E

GmEPSPS_Nat	1251	TGCTGTTGTT	GCACTATTTG	CTAATGGTCC	CACTGCTATA	AGAGATGTGG	1300
GmEPSPS_GM		C.....C	..C..G..C.	.A..C..A..	T..C..A...	..G.....C.	
GmEPSPS_Nat	1301	CAAGTTGGAG	AGTTAAAGAG	ACTGAGAGGA	TGATAGCAAT	CTGCACAGAA	1350
GmEPSPS_GM		.TTCA...C.	T.....G..A	..C..AC...	C..T..	T.....C..G	
GmEPSPS_Nat	1351	CTCAGAAAGC	TAGGAGCAAC	AGTTGAAGAA	GGTCCTGATT	ACTGTGTGAT	1400
GmEPSPS_GM		T.GC.T....	.G..T.....	G..G.....	..A..A..C.	.T..C.....	
GmEPSPS_Nat	1401	TACTCCACCT	GAGAAATTGA	ATGTCACAGC	TATAGACACA	TATGATGACC	1450
GmEPSPS_GM		A..A..T...	..A..GC.C.	G..C..	...T.....T	C..T.	
GmEPSPS_Nat	1451	ACAGAAATGGC	CATGGCATTTC	TCTCTTGCTG	CTTGTGGGGA	TGTTCCAGTA	1500
GmEPSPS_GM		T	T.....	..A.....	C..T..	C..G..G..T	

Figure 3F

1501	1550
GmEPSPS_Nat	ACCATCAAGG ATCCTGGTTG CACCAGGAAG ACATTTCCCTG ACTACTTTGA
GmEPSPS_GM	..G..... .C..A..G.. T..T..... .C..A.. .T.....
1551	1578
GmEPSPS_Nat	AGTCCTTGAG AGGTTAACAA AGCACTAA
GmEPSPS_GM	G..GT..G..A ..A..G.... ..G..G..

Figure 3G

1	BAR1_Nat	ATGAGCCCGAG	AACGACGCC	GGCCGACATC	CGCCGTGCCA	CCGAGGCGGA	50
	BAR1_ZM	...TCG...T	.G..C..T..	T..T...A	A..C..T.	A..	
	BAR1_AT	T....	..A.GA.A..	...T..T..	..GA.A....	A..T..	
51	BAR1_Nat	CATGCCGGCG	GTCTGCACCA	TCGTCAACCA	CTACATCGAG	ACAAGCACGG	100
	BAR1_ZM	T..T	..T.....	T..G..T..		...TCT...	
	BAR1_AT	T.....T..T	..T..T..A.	A.....	T..T.....	..CTCG..A.	
101	BAR1_Nat	TCAACTTCCG	TACCGAGCCG	CAGGAACCGC	AGGACTGGAC	GGACGACCTC	150
	BAR1_ZM	.A.....	C..T.....	..A.....	T.....	C.....T...	
	BAR1_AT	.T..T..T..	C..T.....	G..A.	T.....	T..G	
151	BAR1_Nat	GTCCGTCTGC	GGGAGCGCTA	TCCCTGGCTC	GTCGCCGAGG	TGGACGGCGA	200
	BAR1_ZM	..G.....CA	.A.....T..	...G.....	..T..A....	T..	
	BAR1_AT	..A...T.AA	.A..A..T..	...G.....A	..T..T....	.T.....A..	

Figure 4A

201	250
BAR1_Nat	GGTCGCCGGC ATCGCCTACG CGGGCCCCCTG GAAGGCACGC AACGCCTACG
BAR1_ZM	A..G..T..G T..A..G.. TA.A A.....
BAR1_AT	A.....T..T ..A..T.....T..A..G.. ...A..T..T T.....
251	300
BAR1_Nat	ACTGGACGGC CGAGTCGACC GTGTACGTCT CCCCCCGCCA CCAGCGGACG
BAR1_ZM	.T.....T.. G.....C..A ..C..... ..A...A.A..T..AA.A..C
BAR1_AT	A.. A..A..C..T ..C.....A G...T..T..T..AA.A..C
301	350
BAR1_Nat	GGACTGGGCT CCACGCTCTA CACCACCTG CTGAAGTCCC TGGAGGCACA
BAR1_ZM	..G..C.... ..G..C....T..G..T..C ..C.....T .A.....G..
BAR1_AT	...T.A..GA G....T.G.. ...T..T..T T.A.....A.
351	400
BAR1_Nat	GGGCTTCAAG AGCGTGGTCG CTGTCATCGG GCTGCCCAAC GACCCGAGCG
BAR1_ZM	A TCT..A..G. .G..G..... CT....A.... ..T..C..T..
BAR1_AT	A..... TCT..T..G. .A..T..T.. AT....A.... ..T.....T..

Figure 4B

401	BAR1_Nat	TGCGCATGCA	CGAGGCGCTC	GGATATGCC	CCCGCGGCAT	GCTGCGGGCG	450
	BAR1_ZM	..A.A....	A...	..T.C..T.	.TA.A..A..	...CA.....	
	BAR1_AT	.T..A....	...A.....T	C..T.	.A..A..T..	...C..T..T	
451	BAR1_Nat	GCCGGCTTCA	AGCACGGGAA	CTGGCATGAC	GTGGGTTTCT	GGCAGCTGGA	500
	BAR1_ZM	..T..A....	T..	T.....C...	..T..C....	A.....	
	BAR1_AT	A....	.A..T..A..	T.....C...	.A.....T.	A.....	
501	BAR1_Nat	CTTCAGCCTG	CCGGTACCGC	CCCGTCCGGT	CCTGCCCGTC	ACCGAGATCT	550
	BAR1_ZM	...TCTT..	..A..T..A.	.T.....T..	G..A.....	A.....	
	BAR1_AT	...TTCA..T	..C..T..C.	.TA.A..T..	A..T..A..T	..T..A.....	
551	BAR1_Nat	GA					
	BAR1_ZM	AG					
	BAR1_AT	AG					

Figure 4C

CTP2CP4NAT	1	ATGGCGCAAG	TTAGCAGAAT	CTGCAATGGT	GTGCAGAAC	CATCTCTTAT	CTCCAATCTC	60
CTP2CP4_ZM			A..T.....	C..C		.G..G..G..	C....	
CTP2CP4_AT		C..G.	A..T..G..	...T..C..A	.C..A....	T..A..A..	A..G..C..G	
CTP2CP4_GM		C....	C....	C....	T..	A..A..	C..G	
CTP2CP4NAT	61	TCGAAATCCA	GTCACGCAA	ATCTCCCTTA	TCGGTTTCTC	TGAAGACGCA	GCAGCATCCA	120
CTP2CP4_ZM		AGC..G....	.C..G..G..	G..G..GC.C	CAGC.	.C.....C..	A.....C..G	
CTP2CP4_AT		..A...AG.T	C....A.A..	G..G..GC.T	..T..A..GT	...A..T..	A.....C..G	
CTP2CP4_GM		..C....GT	CA.....T..	G..G..A...	..T...AGCT	T..	...A.....T	
CTP2CP4NAT	121	CGAGCTTATC	CGATTTCGTC	GTCGTGGGA	TTGAAGAAGA	GTGGGATGAC	GTTAATTGGC	180
CTP2CP4_ZM		A.G..C..C.	.T..CAGC..	A..C.....C	C.C.....	C.....	.C.G..C....	
CTP2CP4_AT		A.G.....	.C..C..AAG	C..C.....T	C.A.....A.	A.....	AC.G..C..T	
CTP2CP4_GM		..C..A....	.T..A..AAG	CAGT.....T	AT	CG..T.....	C..G.....T	
CTP2CP4NAT	181	TCTGAGCTTC	GTCCTCTTAA	GGTCATGTCT	TCTGTTTCCA	CGGCGTGCAT	GCTTCACGGT	240
CP2CP4_ZM		AGC.....G.	.G..A..C..	...G....C	.G..C..A.		...A.....	
CTP2CP4_AT		AGC..A..A.	A..G..G..	A.....C	.A..CAG..	T.....	T..A	
CTP2CP4_GM		..G..A...A	.G..AT.G..	...G.....	.A...AGT.	A..T.....		

Figure 5A

CTP2CP4NAT	241	GCAAGCAGCC	GGCCCGCAAC	CGCCCGCAAA	TCCTCTGGCC	TTTCCGGAAC	CGTCCGCATT	300
CTP2CP4_ZM			G.....	T.....	T..C.....	G.....	G...A.G...	
CTP2CP4_AT		..TTCATCTA	...A..T..	T...A.G..G	..TAGC..G.	.CAGT..C..	...G.....C	
CTP2CP4_GM		..TTCCTTC.A	.A.....	G..TA.A..G	AGT.....T	.G..T.....	T...	
CTP2CP4NAT	301	CCCGGCGACA	AGTCGATCTC	CCACCGGTCC	TTCATGTTCG	GCGGTCTCGC	GAGCGGTGAA	360
CTP2CP4_ZM		..G.....T.	...C..A..	G		T..	C.....G	
CTP2CP4_AT		..T.....T.	.AAGT..T..	A...A..AG.		A..A..T..	T..T..A..G	
CTP2CP4_GM		..A..A....	.AAGC..TAG	T.....C..T	T.	.T..G..G..	ATCT..A..G	
CTP2CP4NAT	361	ACGCGCATCA	CCGGCCTTCT	GGAAGGCGAG	GACGTCATCA	ATACGGGCAA	GGCCATGCAG	420
CTP2CP4_ZM			..G.....G..	T...	G....	C..G..		
CTP2CP4_AT		..A.A....	..T..TT.G..	T..G.....A	..T..T...	.C..C..T..	...G.....A	
CTP2CP4_GM			..T..T.....	A...		A..G..	...A.....	
CTP2CP4NAT	421	GCGATGGGCG	CCCGCATCCG	TAAGGAAGGC	GACACCTGGA	TCATCGATGG	CGTCGGCAAT	480
CTP2CP4_ZM		..T.....A.	..G..T.....	C.....	T...	T..C..	...T..G..	
CTP2CP4_AT		..A.....T.	..A.A.....	A..A..G...	..T..G....	C..	T..T..T..	
CTP2CP4_GM		..T.....T.	T..T..	C.....T	..T..T....	A..C..	A..T..G.,	

Figure 5B

481 CTP2CP4NAT GCGGGCCTCC TGGCGCCTGA GCGCGCGCTC GATTTCGGCA ATGCCGCCAC GGGCTGCCGC 540
CTP2CP4_ZM T . . . G . . . C . . . C . . . T G . . . G A . . . G . . .
CTP2CP4_AT . . A . . AT . G . . C . . T . . C . . A . . T G T
CTP2CP4_GM . . T . . . T . A . . T . . A . . G T C . . A A . . G . . TA . A

541 CTP2CP4NAT CTGACGATGG GCCTCGTCGG GGTCTACGAT TTCGACAGCA CCTTCATCGG CGACGCCCTCG 600
CTP2CP4_ZM . . C . . T A . . G G T G A
CTP2CP4_AT . . T . . T A . . G . . A . . C . . G . . T . . C . . T . . TCT T GAGC
CTP2CP4_GM . . T . . T T G T TTCA T . . T . . G . . T T

601 CTP2CP4NAT CTCACAAAGC GCCCGATGGG CCGCGTGTTG AACCCGCTGC GCGAAATGGG CGTGCAGGTG 660
CTP2CP4_ZM T A TC T G A C
CTP2CP4_AT T A . . A C T . . C . . A . . G . . G T . . C
CTP2CP4_GM T . . A A . . A . . CC . T T . . A . . G . . G A . . A . . T

661 CTP2CP4NAT AAATCGGAAG ACGGTGACCG TCTTCCCGTT ACCTTGCGCG GGCCGAAGAC GCCGACGCCG 720
CTP2CP4_ZM C . . G . . T T . . G GC T T . . C . .
CTP2CP4_AT T . . G . . T T G TC A . . C . . C A
CTP2CP4_GM . . G . . C . . G C A . . AT . G C C . . T A . . C . . C . . T

Figure 5C

721 CTP2CP4NAT ATCACCTACC GCGTGCCGAT GGCCTCCGCA CAGGTGAAGT CCGCCGTGCT GCTCGCCGGC 780
CTP2CP4_ZM ..T..... A.....C A.....C.....A
CTP2CP4_AT G...A .G..T..... .G..A... .A..G..A.. C..G..G...
CTP2CP4_GM ..T...A .G..T..A.. ..A..T..T ..A.....A..T.. ..T...A

781 CTP2CP4NAT CTCAACACGC CCGGCATCAC GACGGTCATC GAGCCGATCA TGACGCGCGA TCATACGGAA 840
CTP2CP4_ZM ..G.....T. .G.....G... C.....G... ..CA..G... ..C...
CTP2CP4_AT A. .T..A..... .A..C..G..T ..A..C... ..TA..A.. C..C.....G
CTP2CP4_GM T.G.....A. .G..T..T.. .G..C..G..T T..T.. C..C..T..G

841 CTP2CP4NAT AAGATGCTGC AGGGCTTTGG CGCCAACCTT ACCGTCGAGA CGGATGCGGA CGGCGTGCGC 900
CTP2CP4_ZM T. G..... ..G.....C..T.. CA..G
CTP2CP4_AT T... T..C... T..T..A ..G.....A. .C..C..C... ..A..G
CTP2CP4_GM T. T..C... T..T.....C T..A. .A..C..... ..T...A..G

901 CTP2CP4NAT ACCATCCGCC TGAAGGCCG CGGCAAGCTC ACCGGCCAAG TCATCGACGT GCCGGGCGAC 960
CTP2CP4_ZM T..G...A. G..T..A..G ..T.....T... ..T.....A...
CTP2CP4_AT ..A.....T G...A. A..T..A..G ..T.....T... ..T..A...T
CTP2CP4_GM T.... G..AA. G..A..A... ..T..T.... ..T.....C..T...T

Figure 5D

CTP2CP4NAT	961	CCGTCCTCGA	CGGCCTTCCC	GCTGGTTGCG	GCCCTGCTTG	TTCCGGGCTC	CGACGTCACC	1020
CTP2CP4_ZM		...G..C.		T.....	...G...C.	...G...GAG	T.....G...	
CTP2CP4_AT		..C..G..C.	.A..G..T..	C..C..A..T	..GT...C.	..C..T..A..	T..T..G..G	
CTP2CP4_GM		..C...AGC.	...G.....	A.....C	..T..T..C.	..A..A.....	...T..G..A	
CTP2CP4NAT	1021	ATCCTCAACG	TGCTGATGAA	CCCCACCCGC	ACCGGCCCTCA	TCCTGACGCT	GCAGGAAATG	1080
CTP2CP4_ZM		...G....	.C..C....	...G...A.G	...G...G.	...C...G.	T.....G...	
CTP2CP4_AT		...G..T.	.C..C....	T..A..TA.A	...G...G.	...C..AT.	...G...G...	
CTP2CP4_GM		..T..A....	.C..C....	T..T..TA.A	...AT.G.	..A..T..AT.	...G...G...	
CTP2CP4NAT	1081	GGCGCCGACA	TCGAAGTCAT	CAACCCGCGC	CTTGCCGGCG	GCGAAGACGT	GGCGGACCTG	1140
CTP2CP4_ZM		..A.....	...G..G..	...G...G.	...G..A...	...T.....	T.....T...	
CTP2CP4_AT		..T..T....	...G..T..	...T..TA.G	T.G..A..T.	..A..G..T..	...C..T...	
CTP2CP4_GM		..T..T..T.	T.....T..	...T..TA.A	..A...A.	..T..G.....	A..T..TT..	
CTP2CP4NAT	1141	CGCGTTCGCT	CCTCCACGCT	GAAGGGCGTC	ACGGTGCCGG	AAGACCGCGC	GCCTTCGATG	1200
CTP2CP4_ZM		...G....	...T..C..	...G...G	...C...C...	...T.....	...G..C...	
CTP2CP4_AT		...G..T.	.TAGT..A..	C..A.....	..C..C..T.	..G..T.....	T..A..C...	
CTP2CP4_GM		..G..GA.G.	.T..T..AT.	...A..T..T	..C..A..T.	...TA.G..	A.....A...	

Figure 5E

CTP2CP4NAT	1201	ATCGACGAAT	ATCCGATTCT	CGCTGTCGCC	GCCGCCCTTCG	CGGAAGGGGC	GACCGTGATG	1260
CTP2CP4_ZM		.A....G.	T....	G..C....	..T.G....	.C.....	C..G..C...	
CTP2CP4_AT		G.	.C..C....	..C..T..	..T.G..T.	.C..G..C..	A..T..A...	
CTP2CP4_GM		..T.....	A....	T..C....	G..T.....	.T..G..C..		
CTP2CP4NAT	1261	AACGGTCTGG	AAGAACTCCG	CGTCAAGGAA	AGCGACCGCC	TCTCGGCCGT	CGCCAAATGGC	1320
CTP2CP4_ZM		T.	.G.....	G....	TCG..T....	.G.....	G.....	
CTP2CP4_AT		C..T.	.G..GT.GA.	G..T.....	G..T...A.G.	.G..C..G..	G..G.....	
CTP2CP4_GM		..T..A..A.	.G..GT.GA.	A..G.....	TC....A.AT	.GAGC..A..	...T..C..A	
CTP2CP4NAT	1321	CTCAAGCTCA	ATGGCGTGGA	TTGCGATGAG	GGCGAGACGT	CGCTCGTCGT	GCGTGGCCGC	1380
CTP2CP4_ZM		.G.....	.C..T..T..	C.....	..T....C.	.A.....	G..C.....	
CTP2CP4_AT		.G.....A.	.C.....	C....C..A	..T..A....	.C..T..A..	C.....T...	
CTP2CP4_GM		..T..A....	T..	..T.....	..T....TA	GCT..G..A..	C.....G..A	
CTP2CP4NAT	1381	CCTGACGGCA	AGGGGCTCGG	CAACGCCCTCG	GGCGCCGCCG	TCGCCACCCA	TCTCGATCAC	1440
CTP2CP4_ZM		T....	C....	AGT	..A..G....	G..	C.....T	
CTP2CP4_AT		..A.....	T.G..	G..T..T...	..A..T..T.	.G..G..G..	C..T.....	
CTP2CP4_GM		..A.....	TT.G..	T...	..T..T....	.T..A..T..	CT.G.....	

Figure 5F

CTP2CP4NAT	1441	CGCATCGCCA	TGAGCTTCCT	CGTCATGGGC	CTCGTGTCGG	AAAACCCCTGT	CACGGTGGAC	1500
CTP2CP4_ZM		G.	..TC....T.	G..G.....T	C..A.	..G....G..	G..C..C..T	
CTP2CP4_AT		A.A.....	..TCA..T..	G..G.....A	..T..C..C..	..G..T..G..	G..C..T...	
CTP2CP4_GM		..G..A..G.	T..T..	G..G.....T	AAGC..	..G..T....	G..A..C...	
CTP2CP4NAT	1501	GATGCCACGA	TGATCGCCAC	GAGCTTCCCG	GAGTTCATGG	ACCTGATGGC	CGGGCTGGGC	1560
CTP2CP4_ZM		..C.....	A..G..	A	T....	G..C..C..G		
CTP2CP4_AT		T..C..	T..T..	CTC...T..T	C....	A..CT....G		
CTP2CP4_GM		A..T..	T..T..	TTC.....T	T....	..TT.A....A..A..T		
CTP2CP4NAT	1561	GCGAAGATCG	AACTCTCCGA	TACGAAGGCT	GCCTGA			1596
CTP4CP4_ZM		..C.....	G..T..	C.....C	..T...			
CTP2CP4_AT		..C.....	..G..G..T..	...T.....C	..T...			
CTP2CP4_GM		..A.....T.	T..T..	...C...A..G	A..			

Figure 5G

33/52

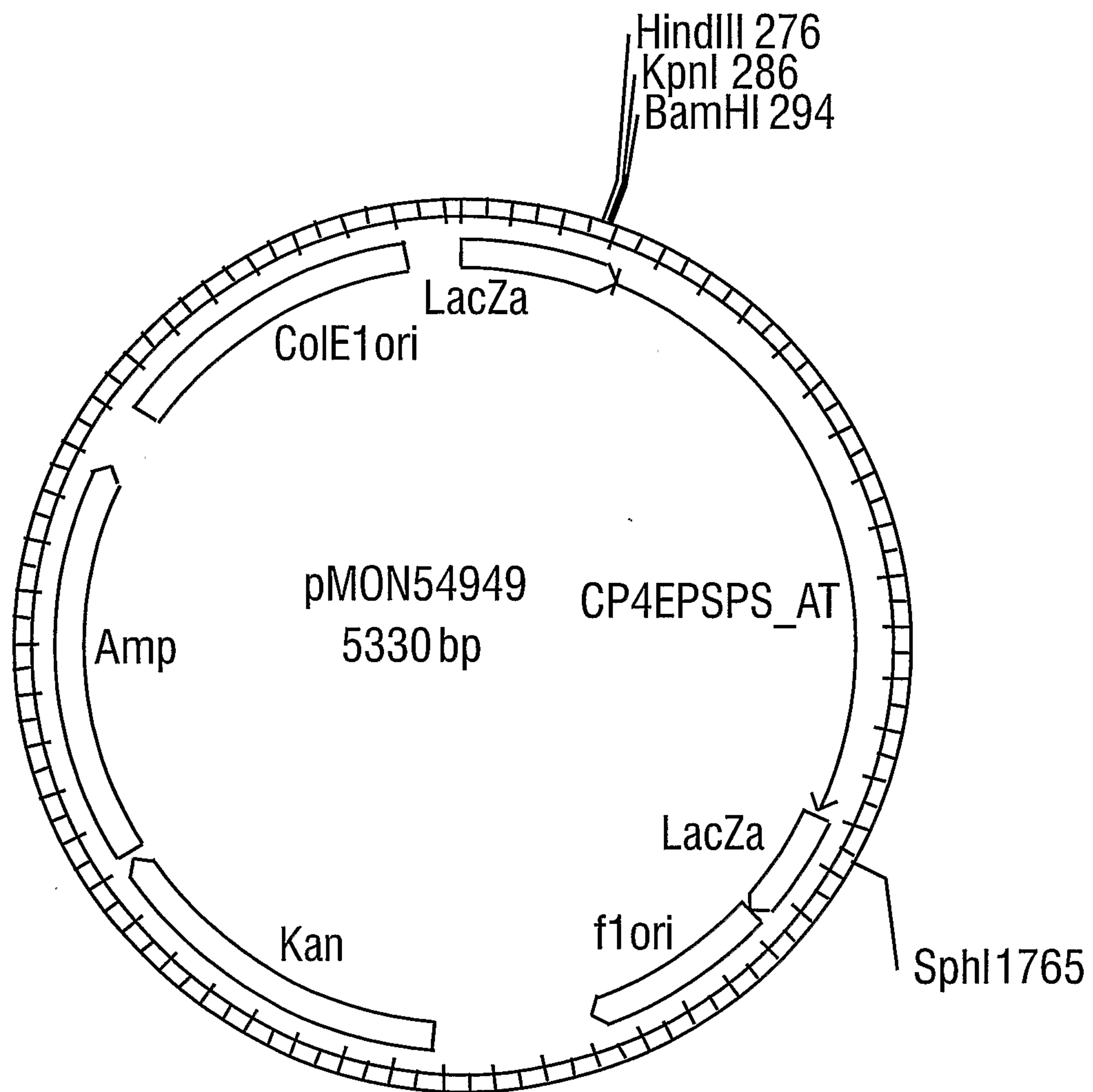


Figure 6

34/52

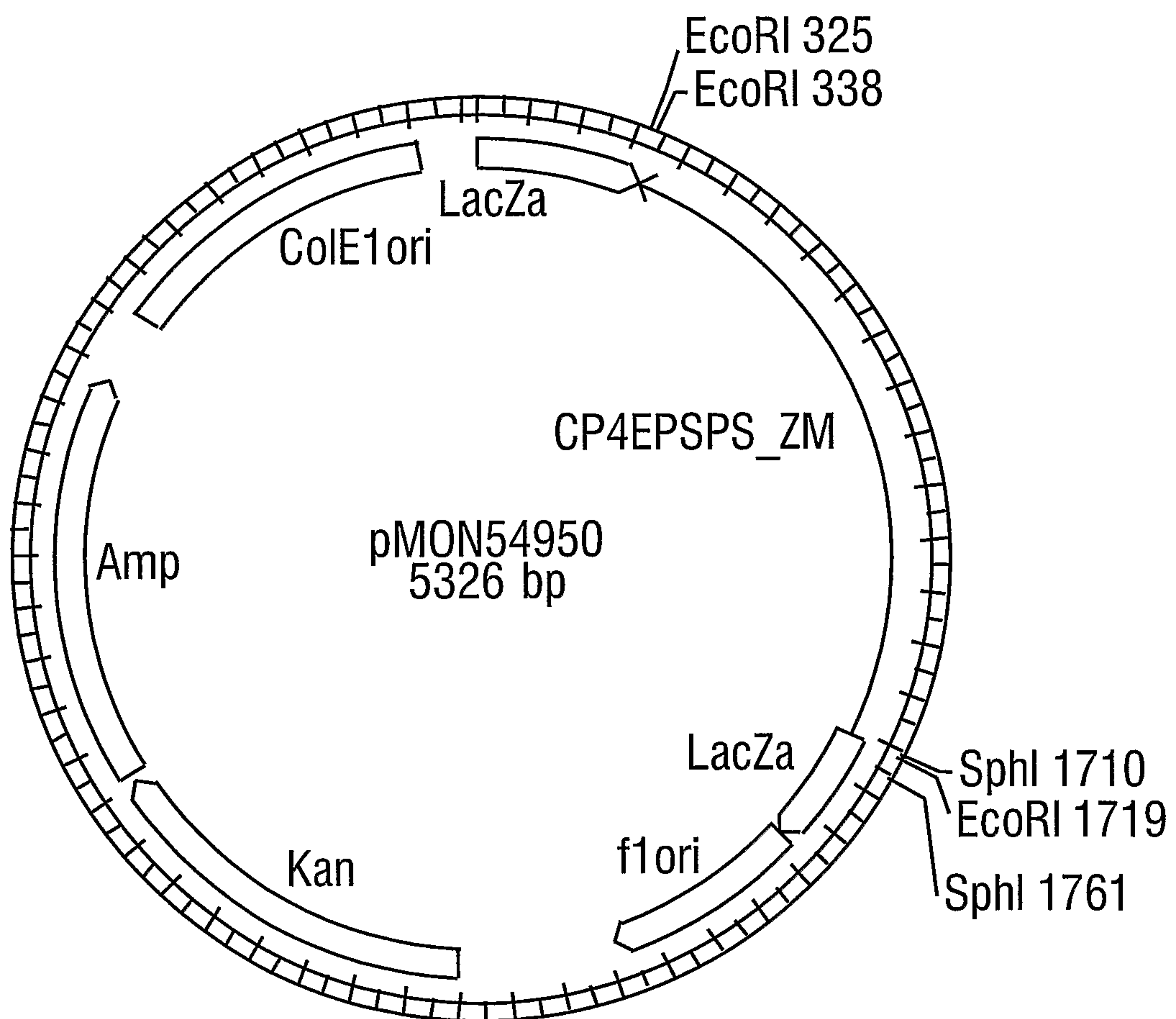


Figure 7

35/52

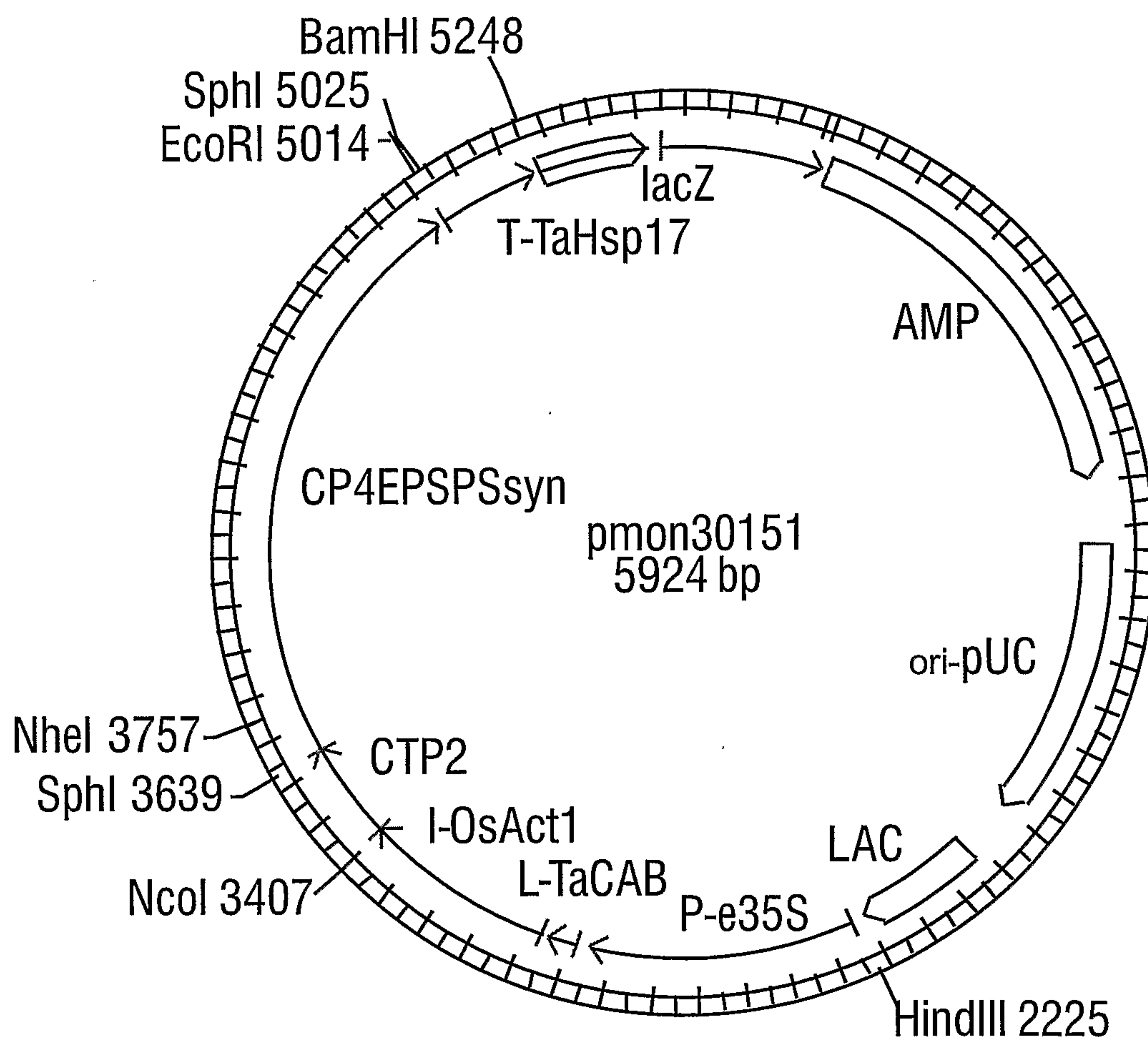


Figure 8

36/52

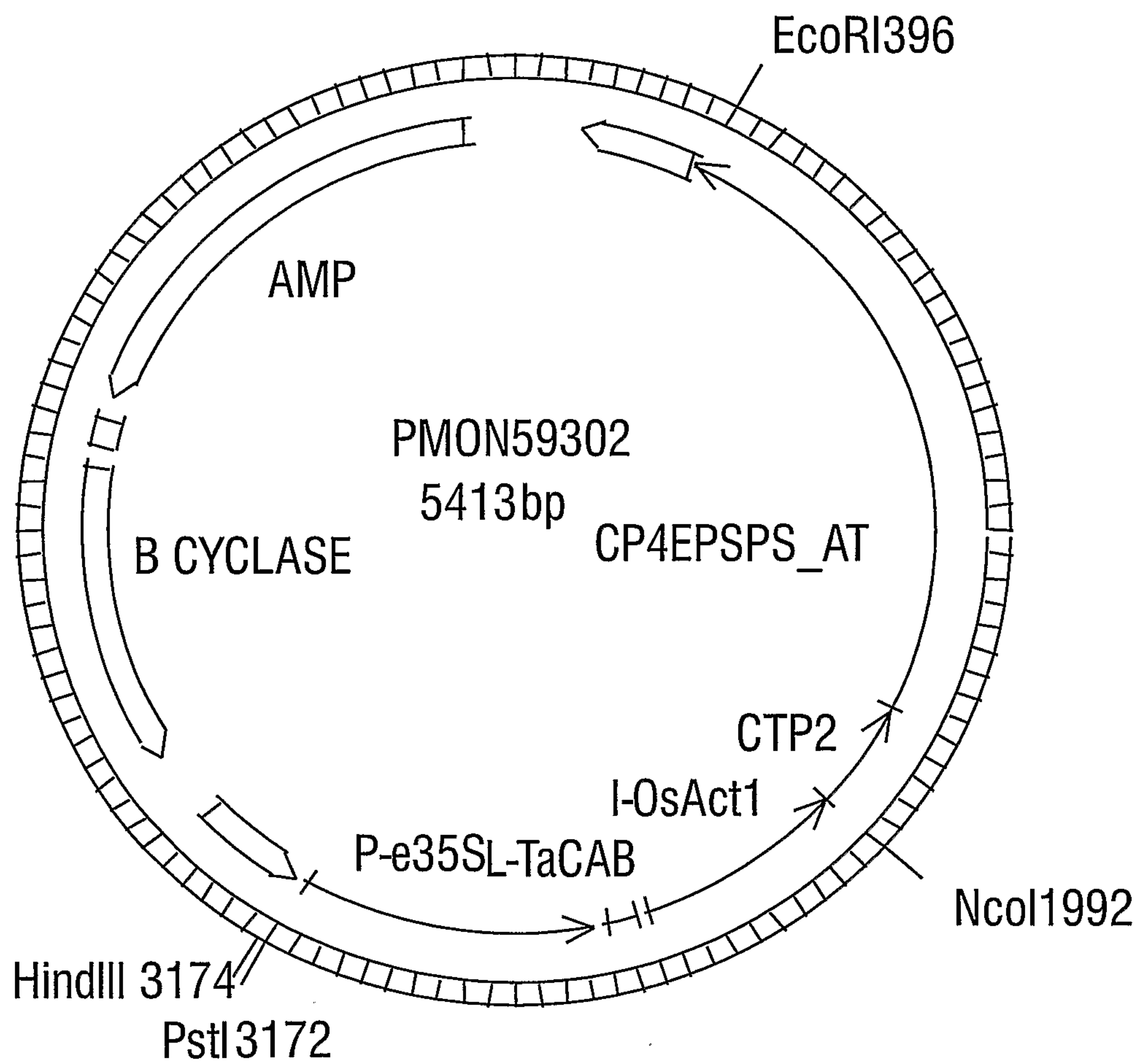


Figure 9

37/52

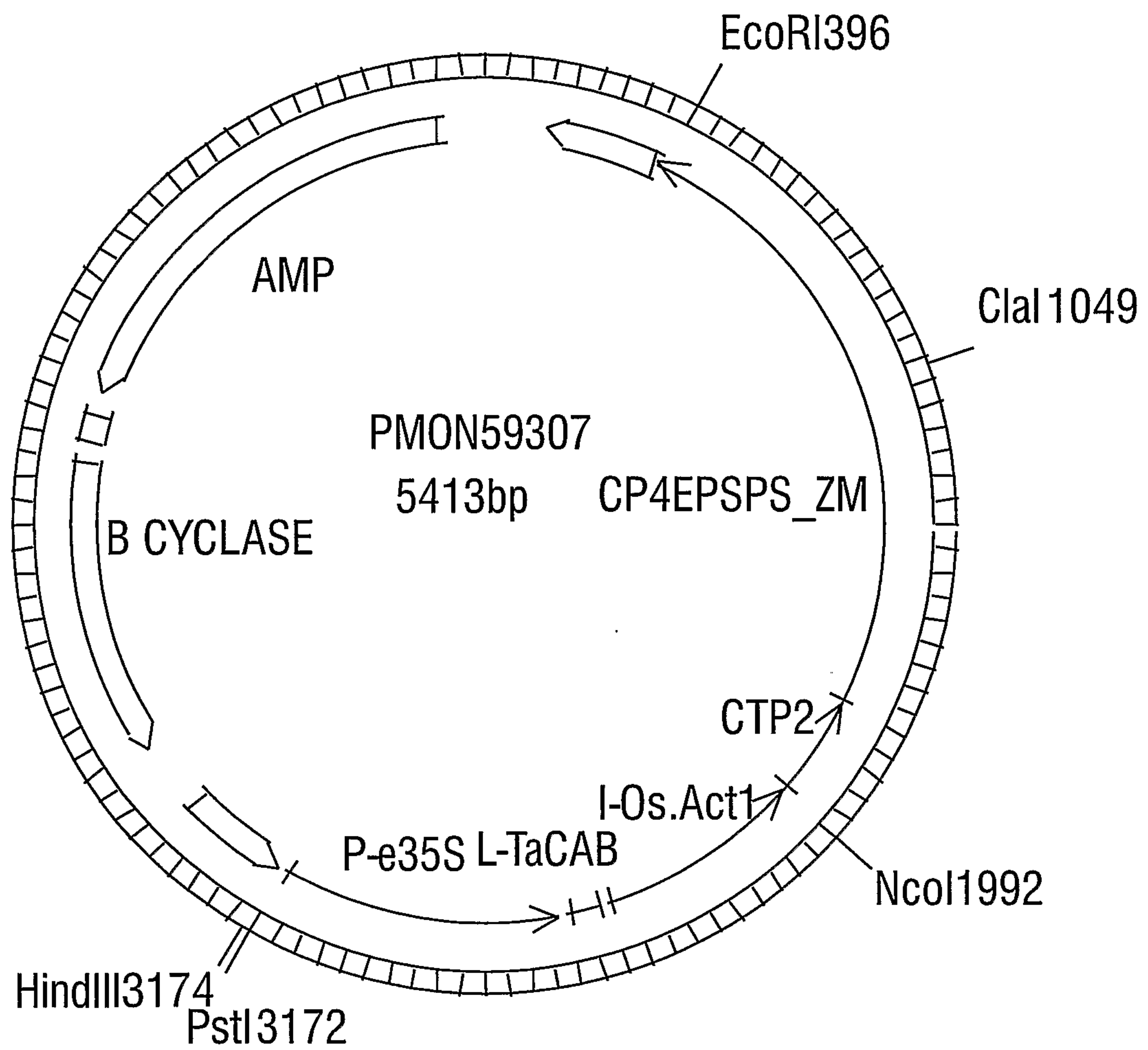


Figure 10

38/52

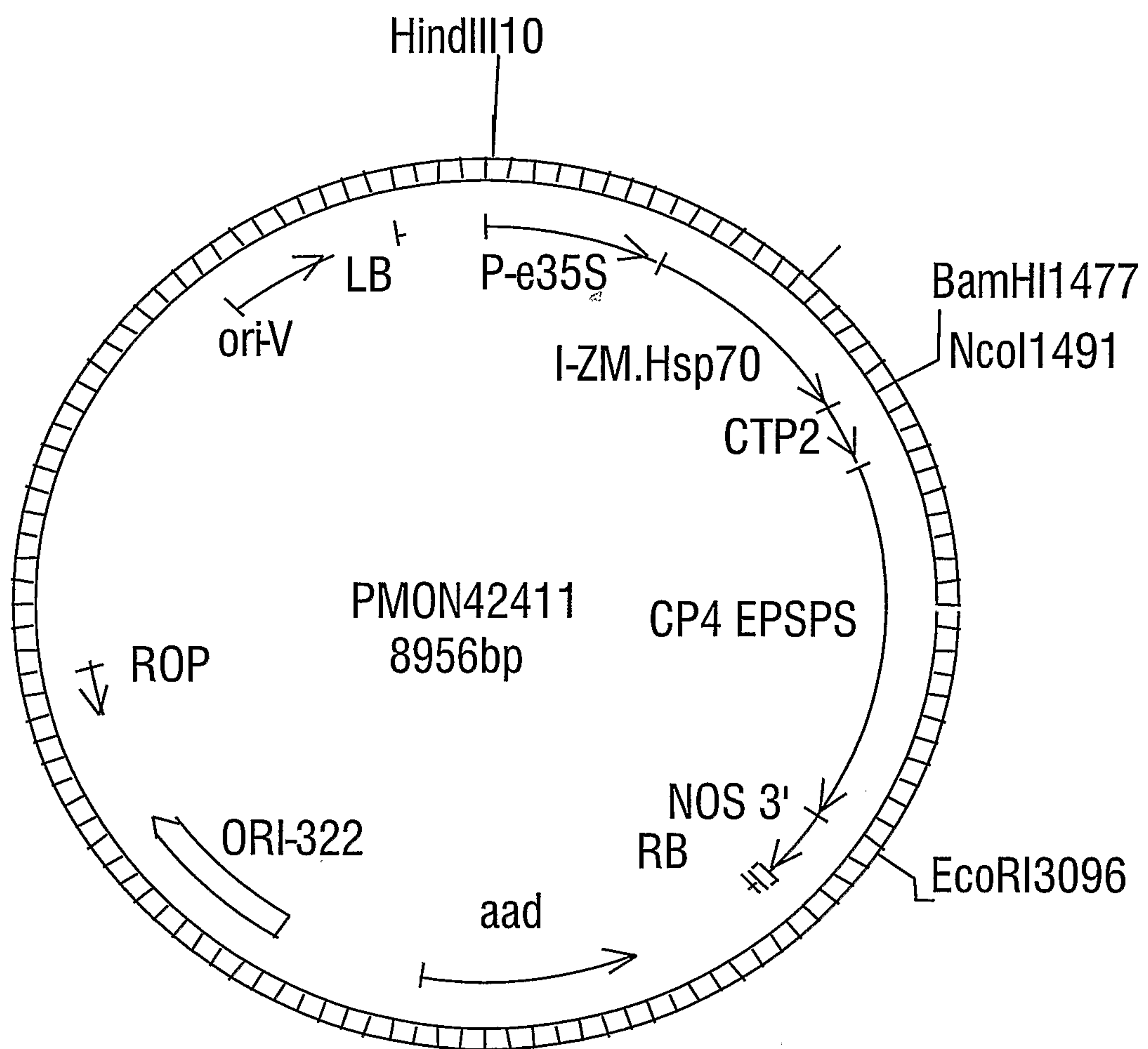


Figure 11

39/52

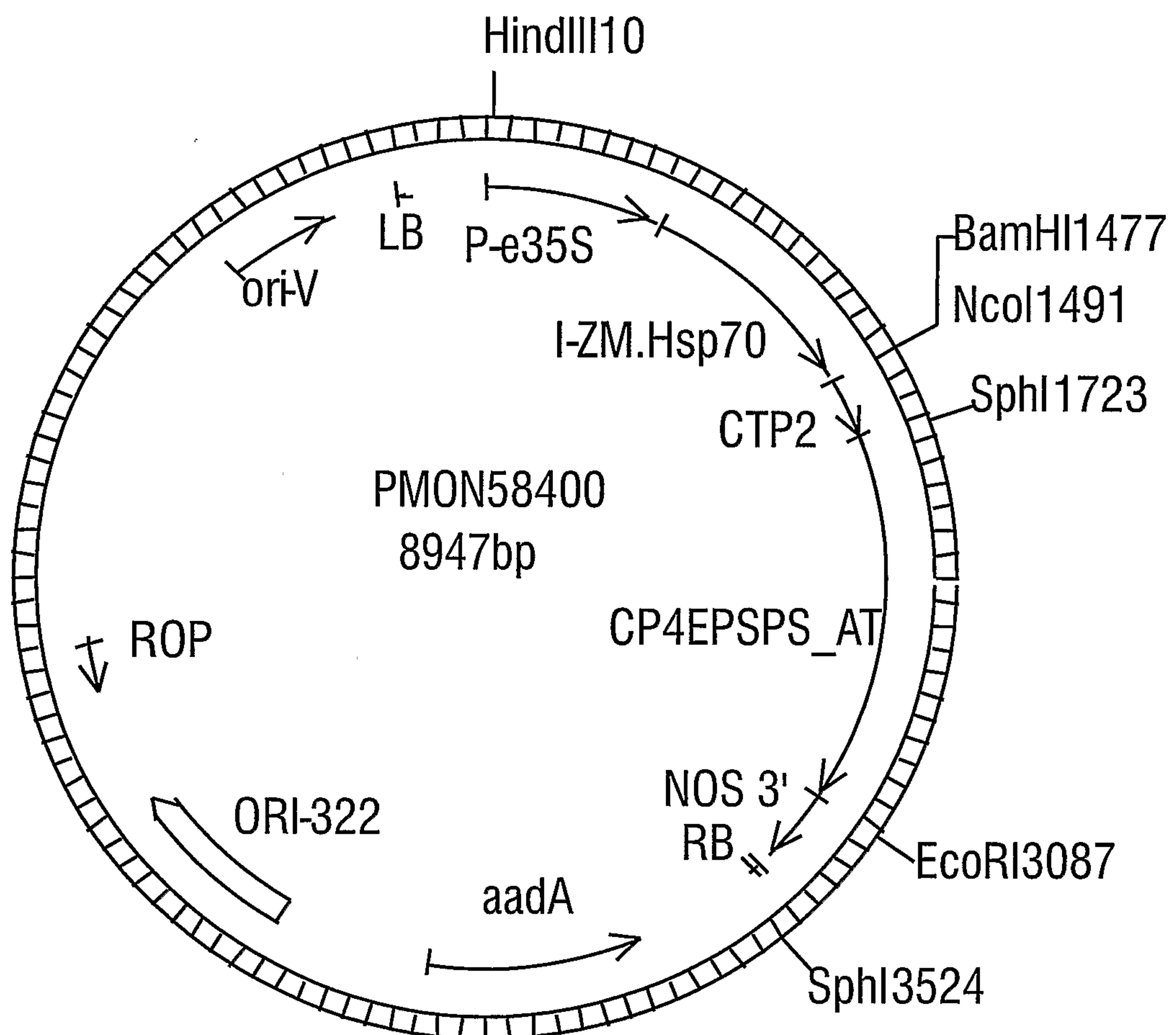


Figure 12

40/52

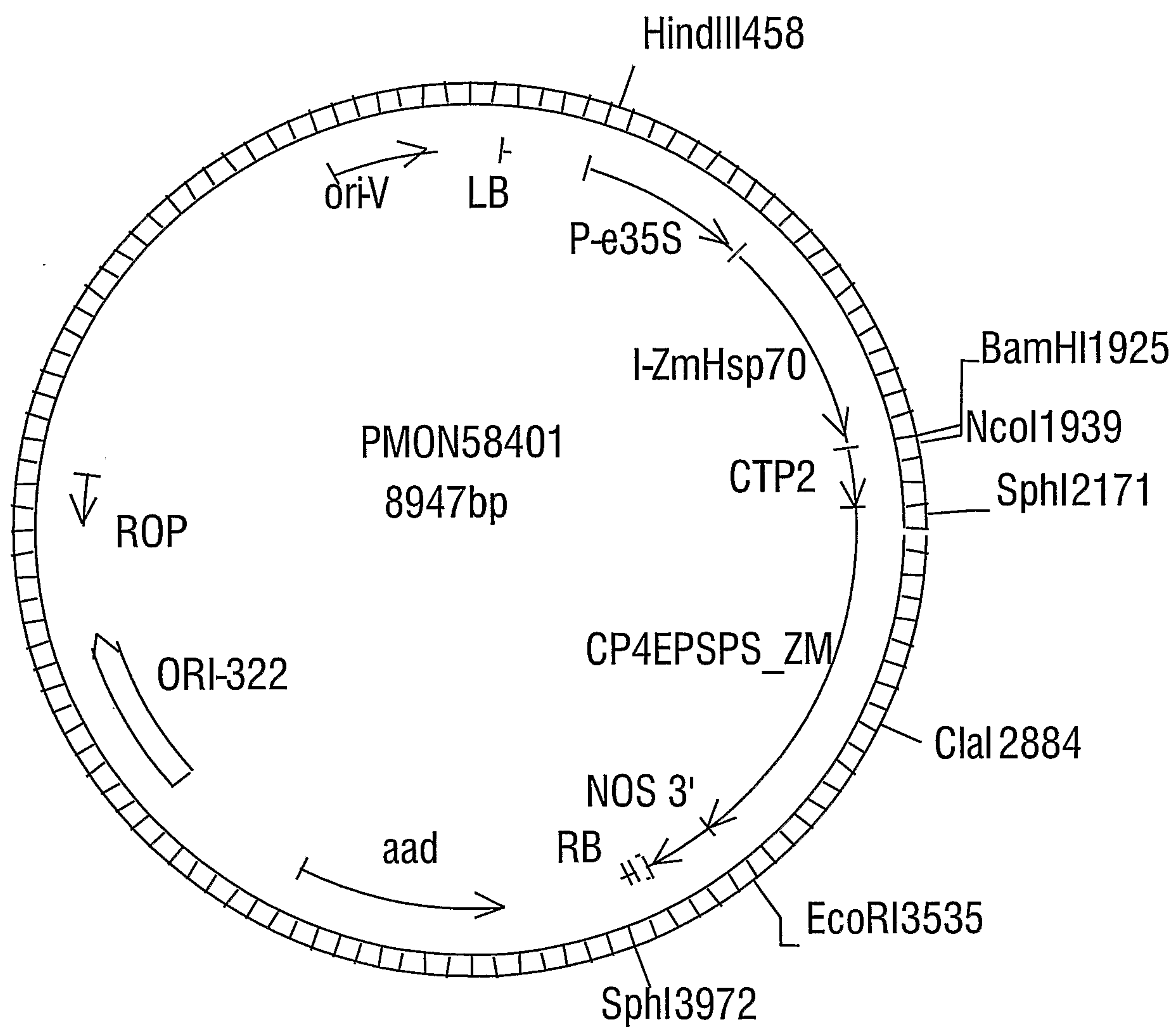


Figure 13

41/52

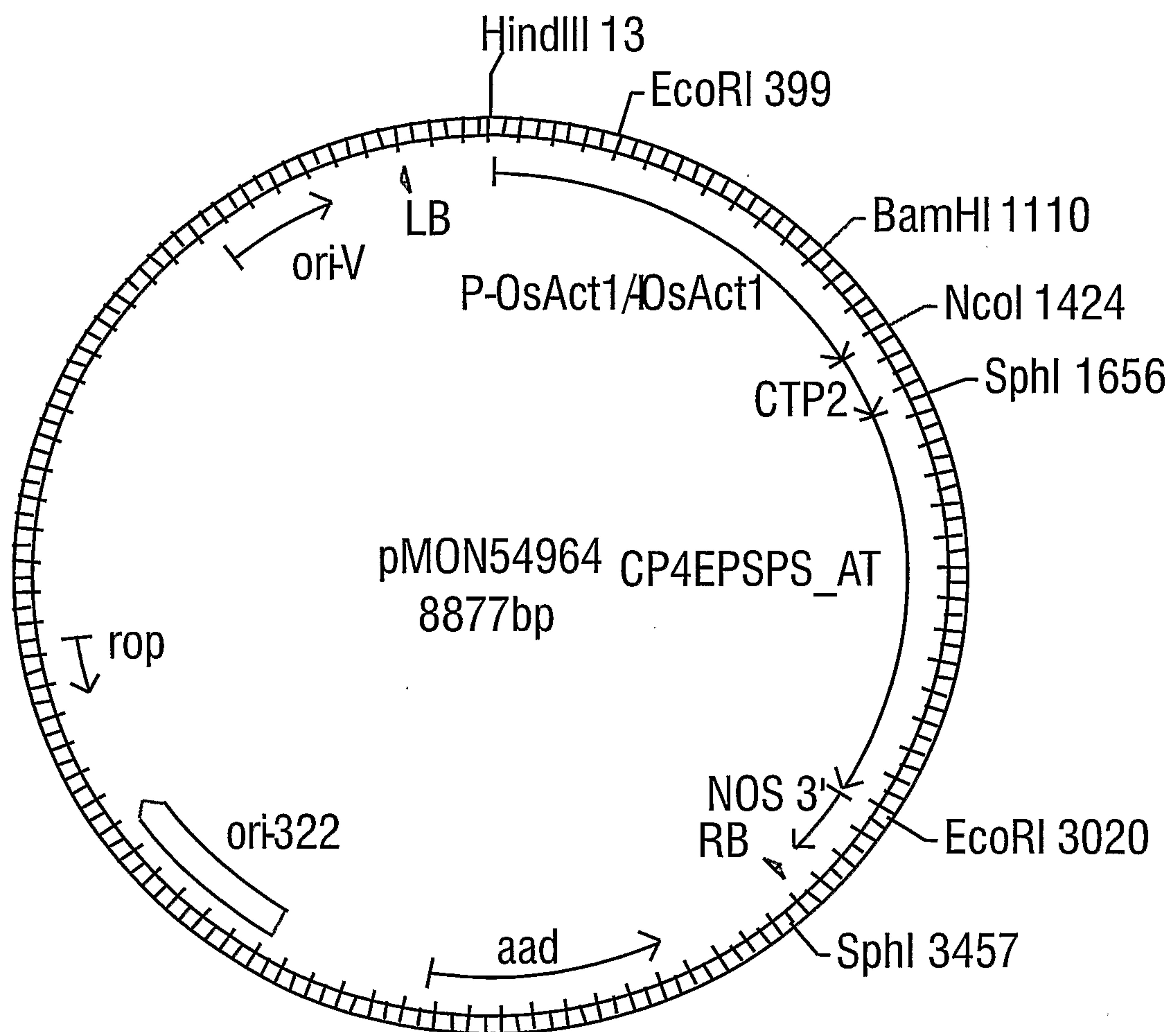


Figure 14

42/52

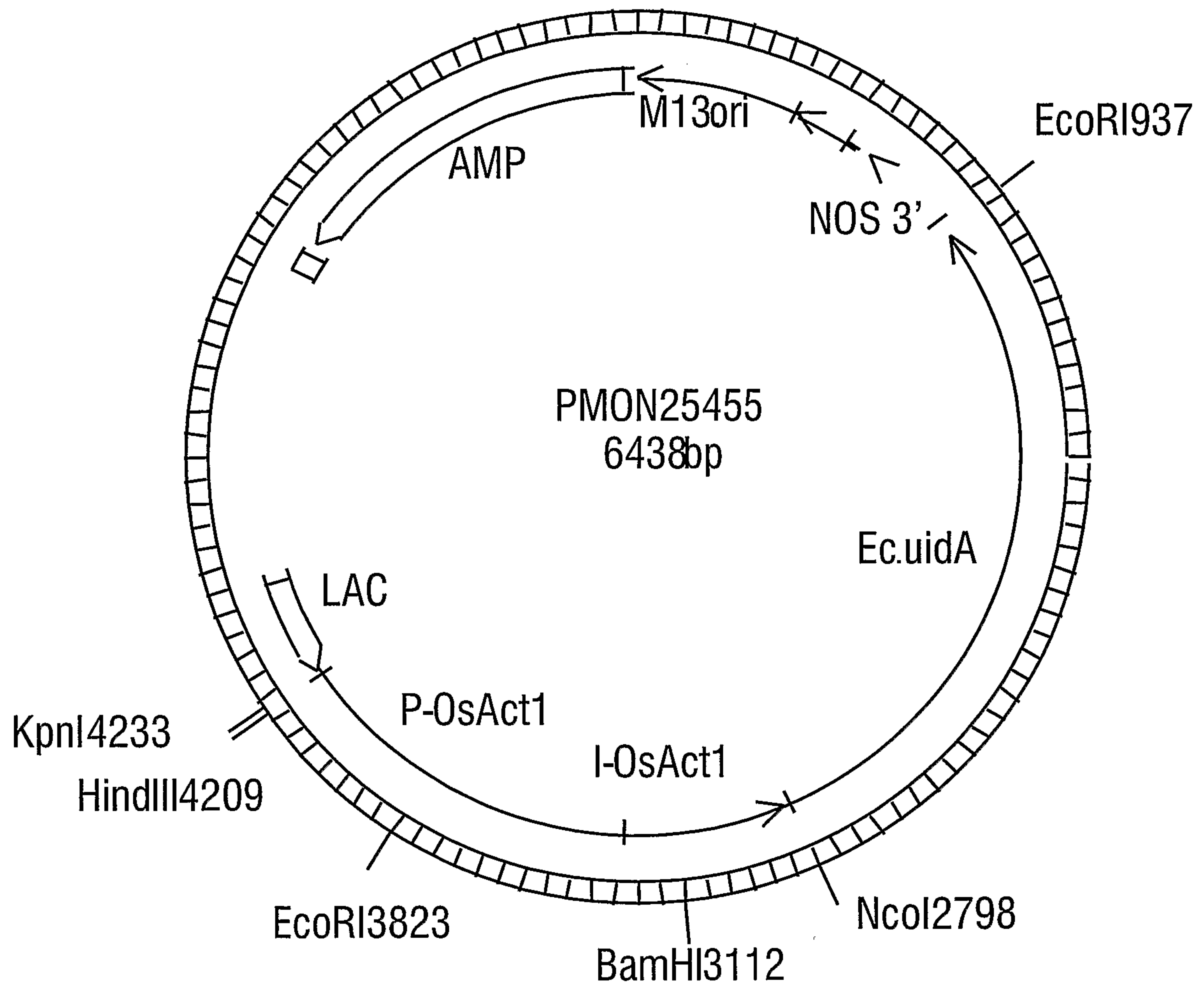


Figure 15

43/52

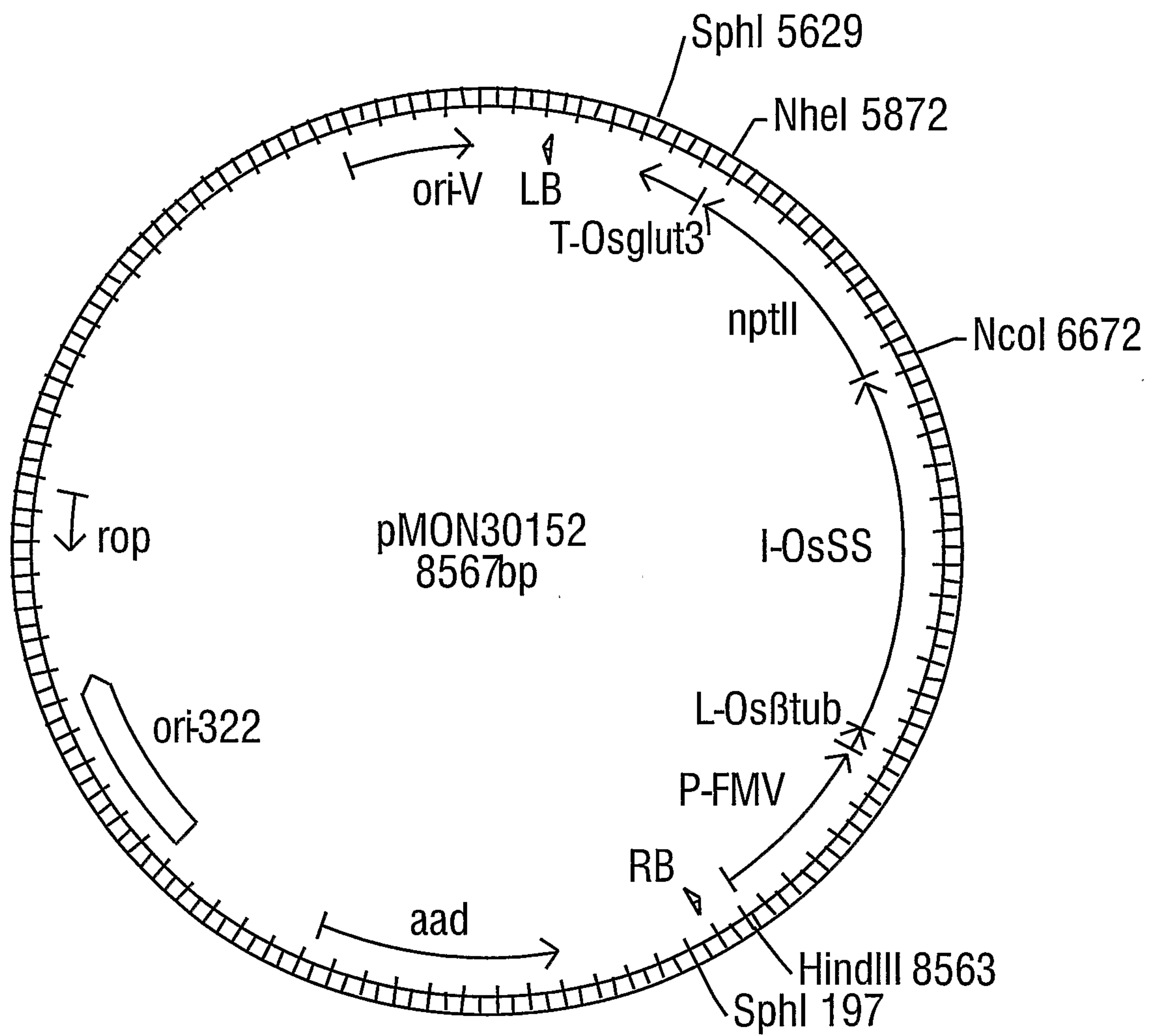


Figure 16

44/52

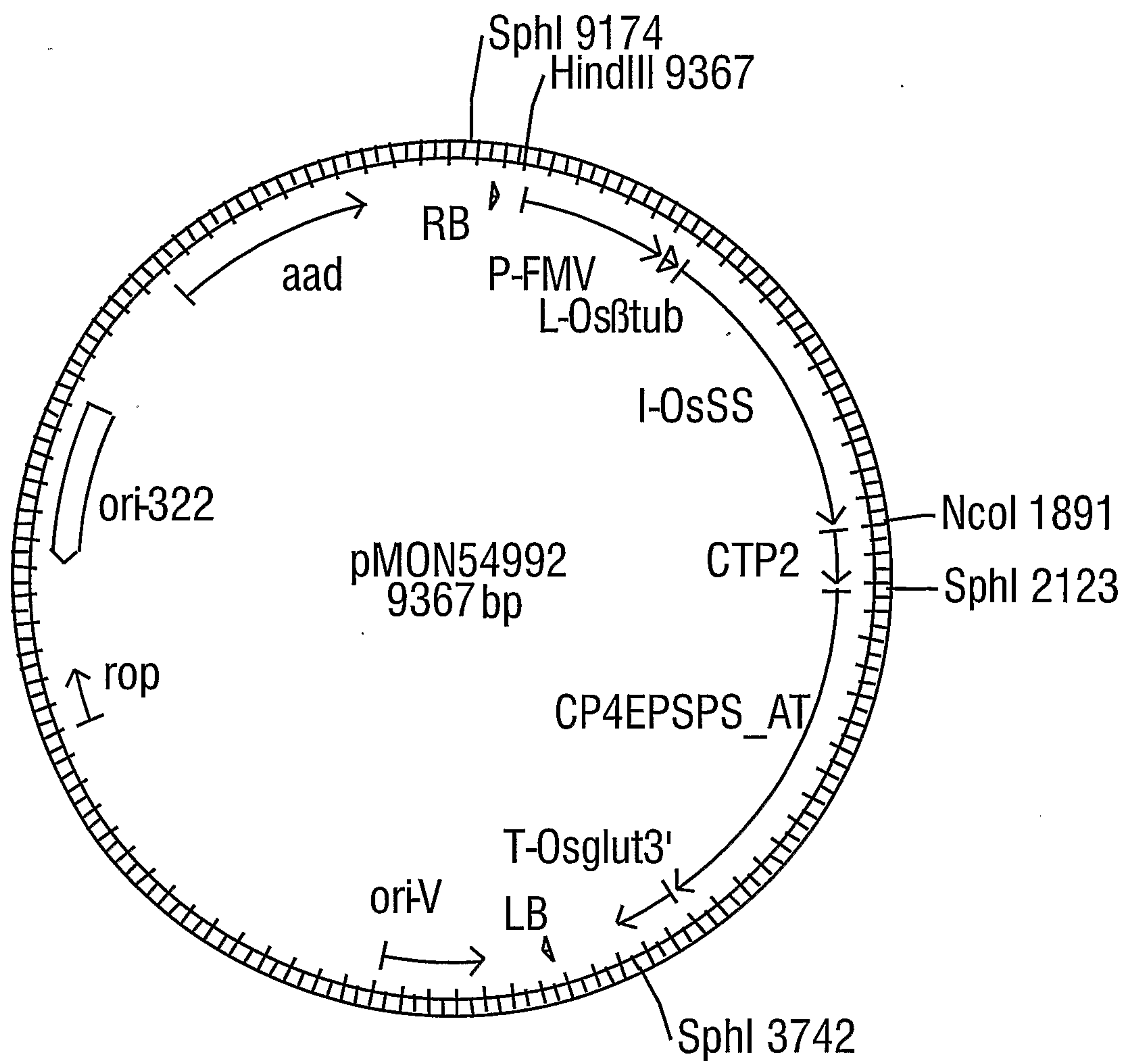


Figure 17

45/52

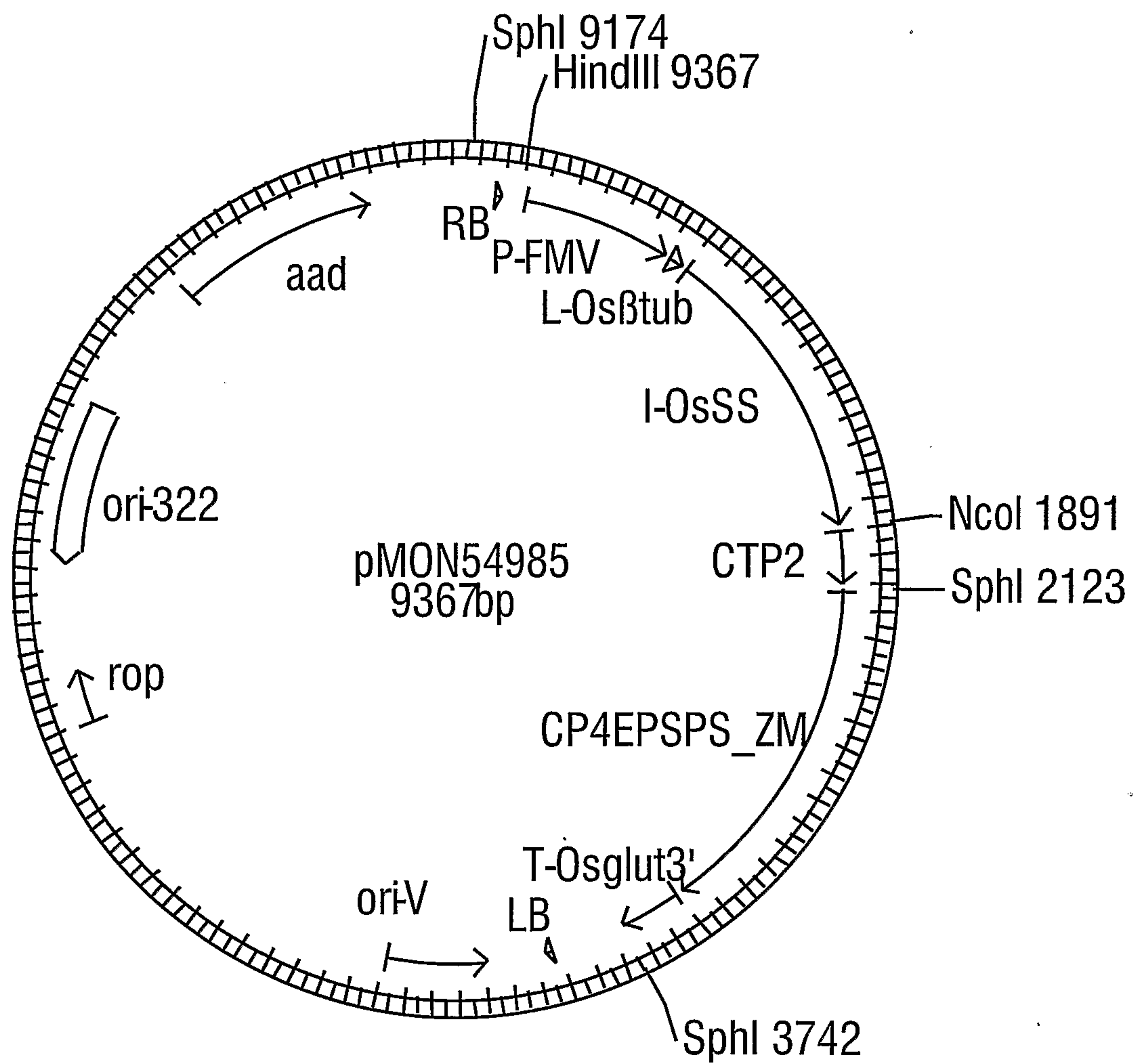


Figure 18

46/52

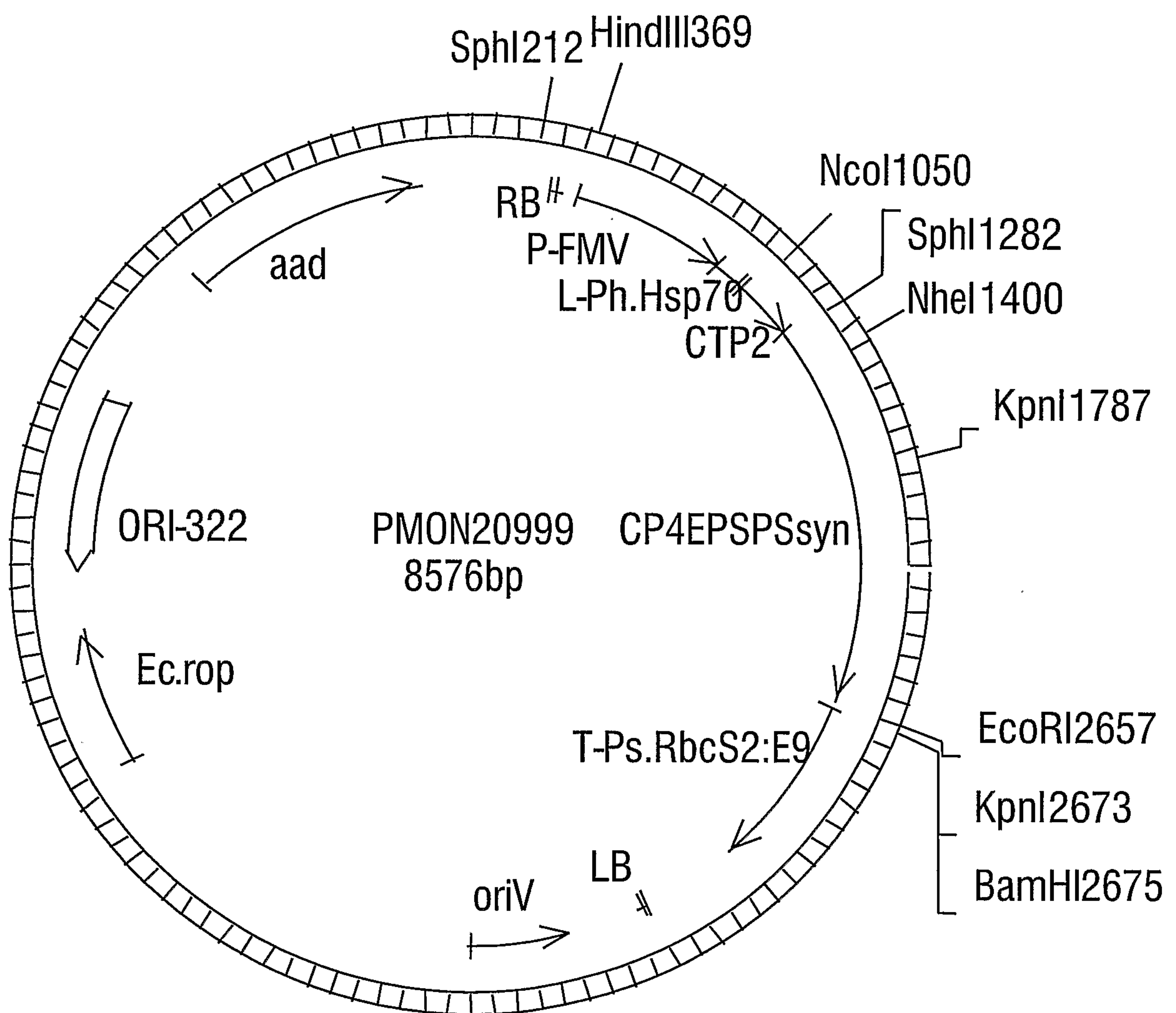


Figure 19

47/52

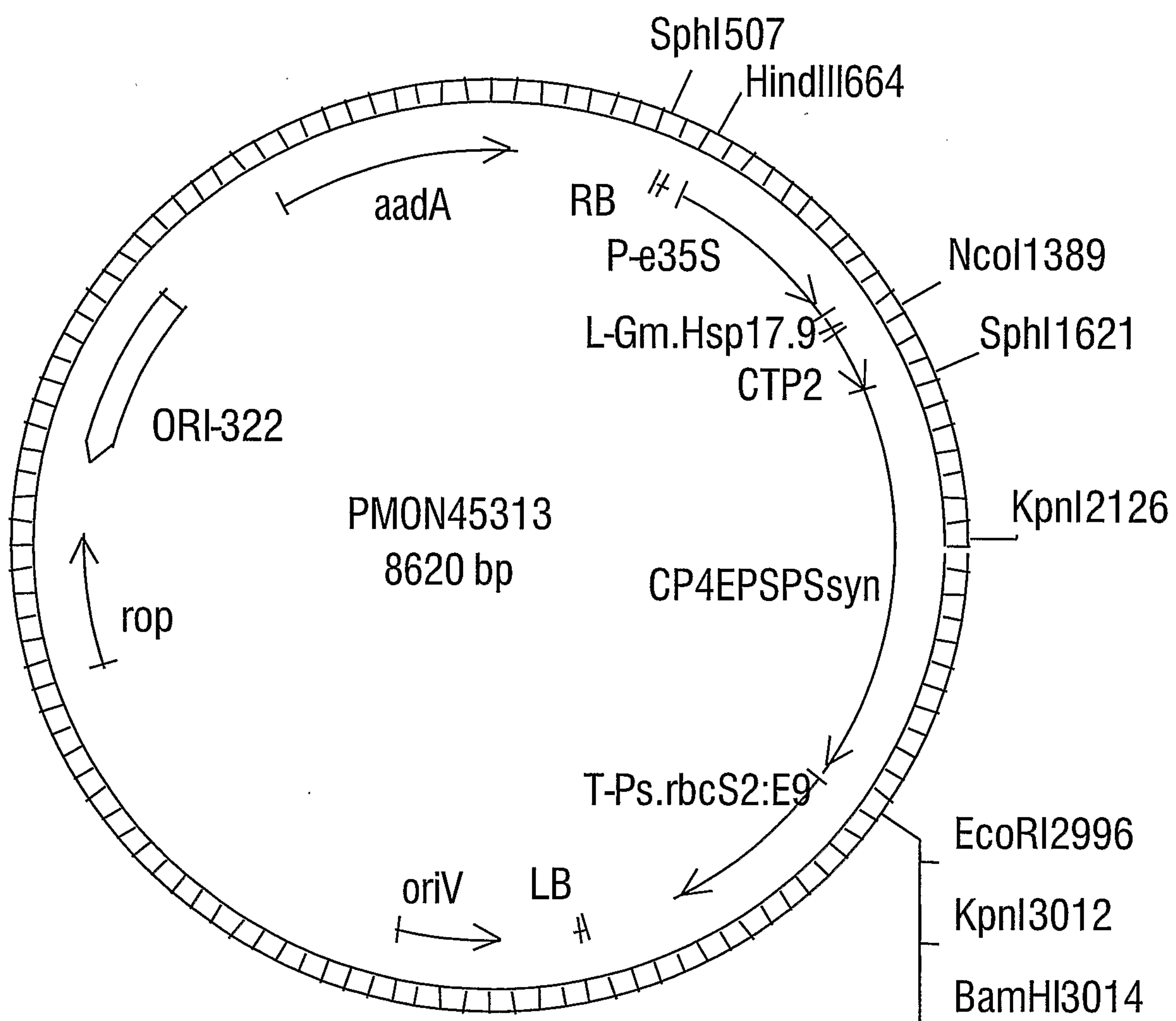


Figure 20

48/52

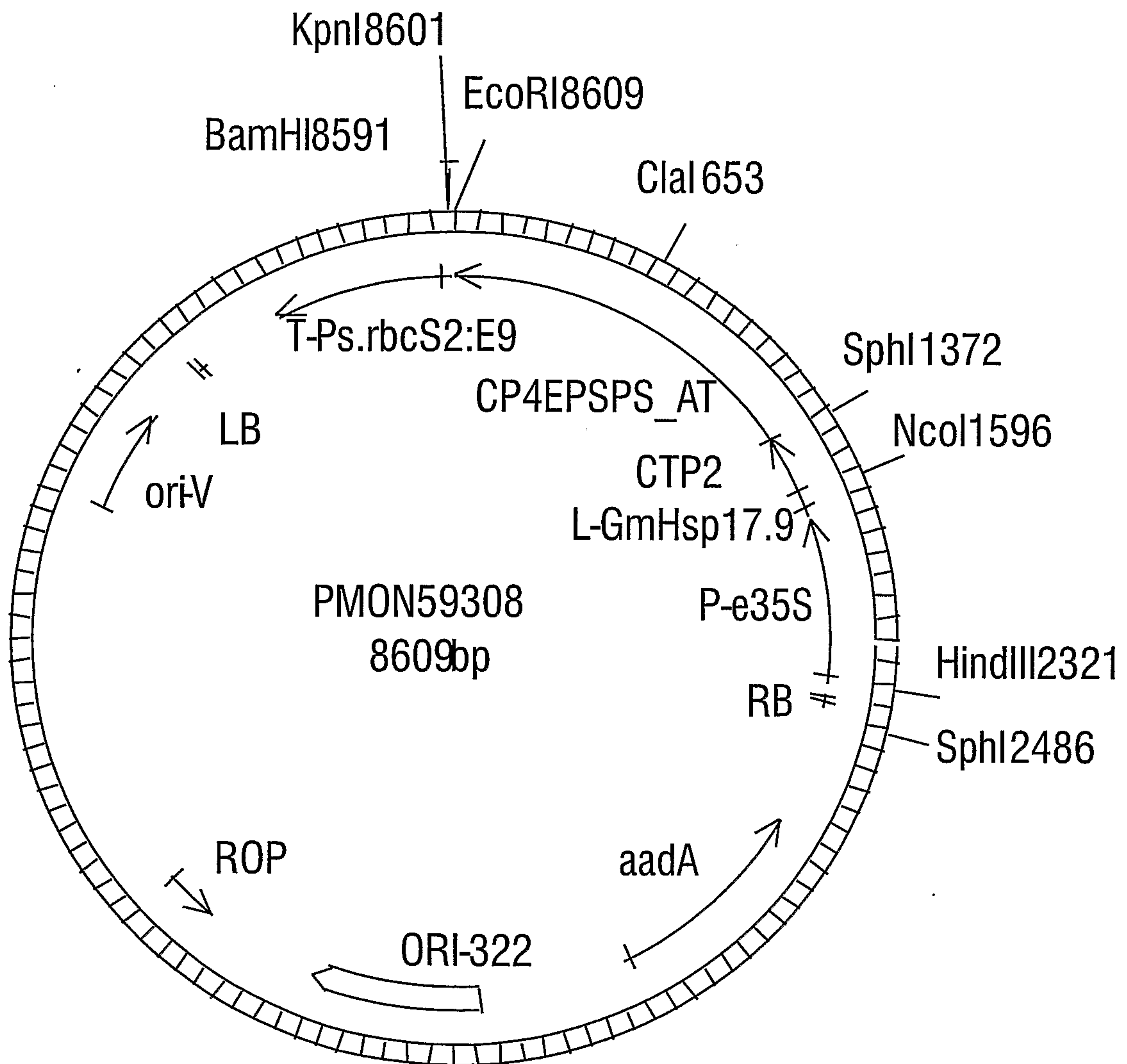


Figure 21

49/52

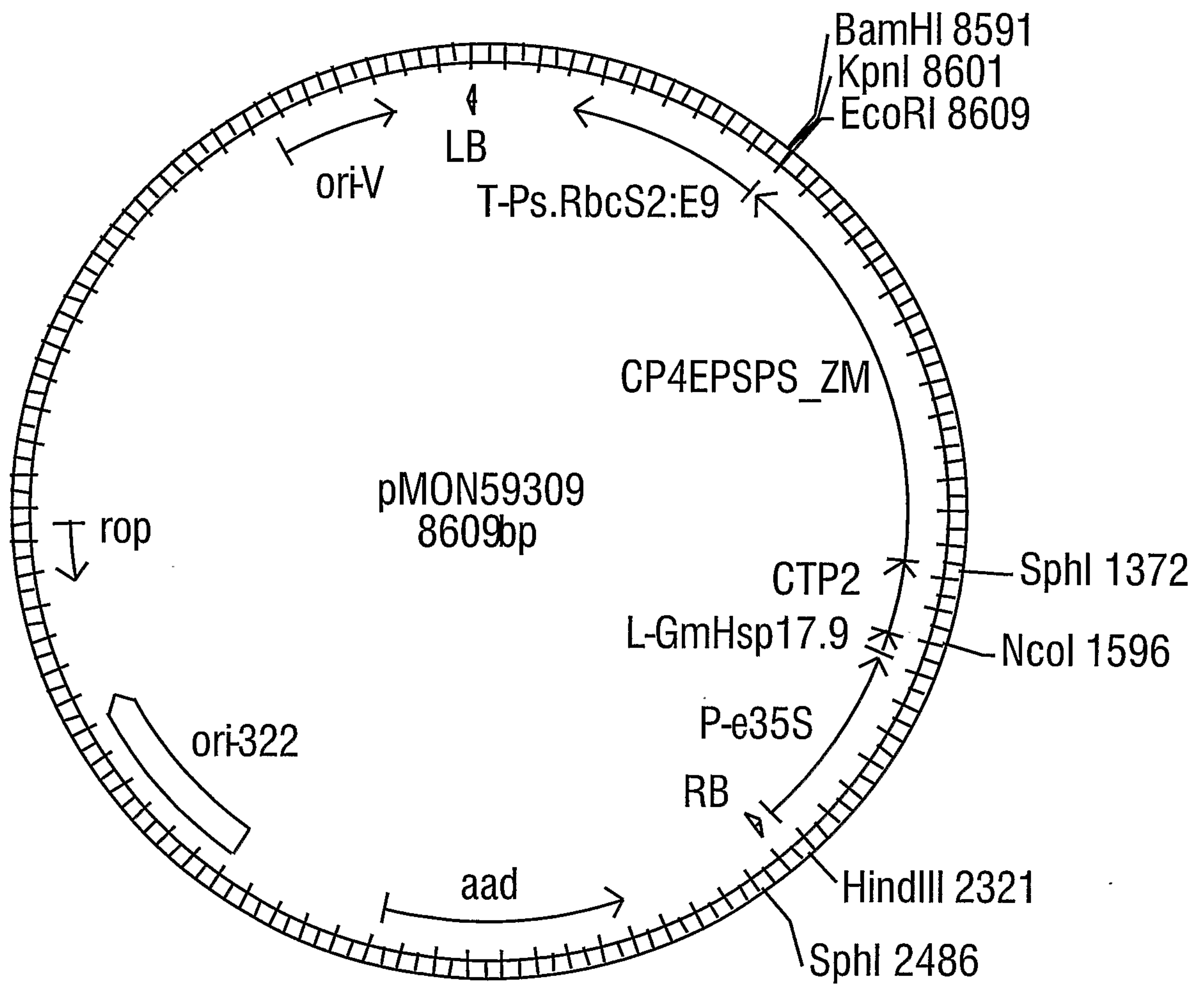


Figure 22

50/52

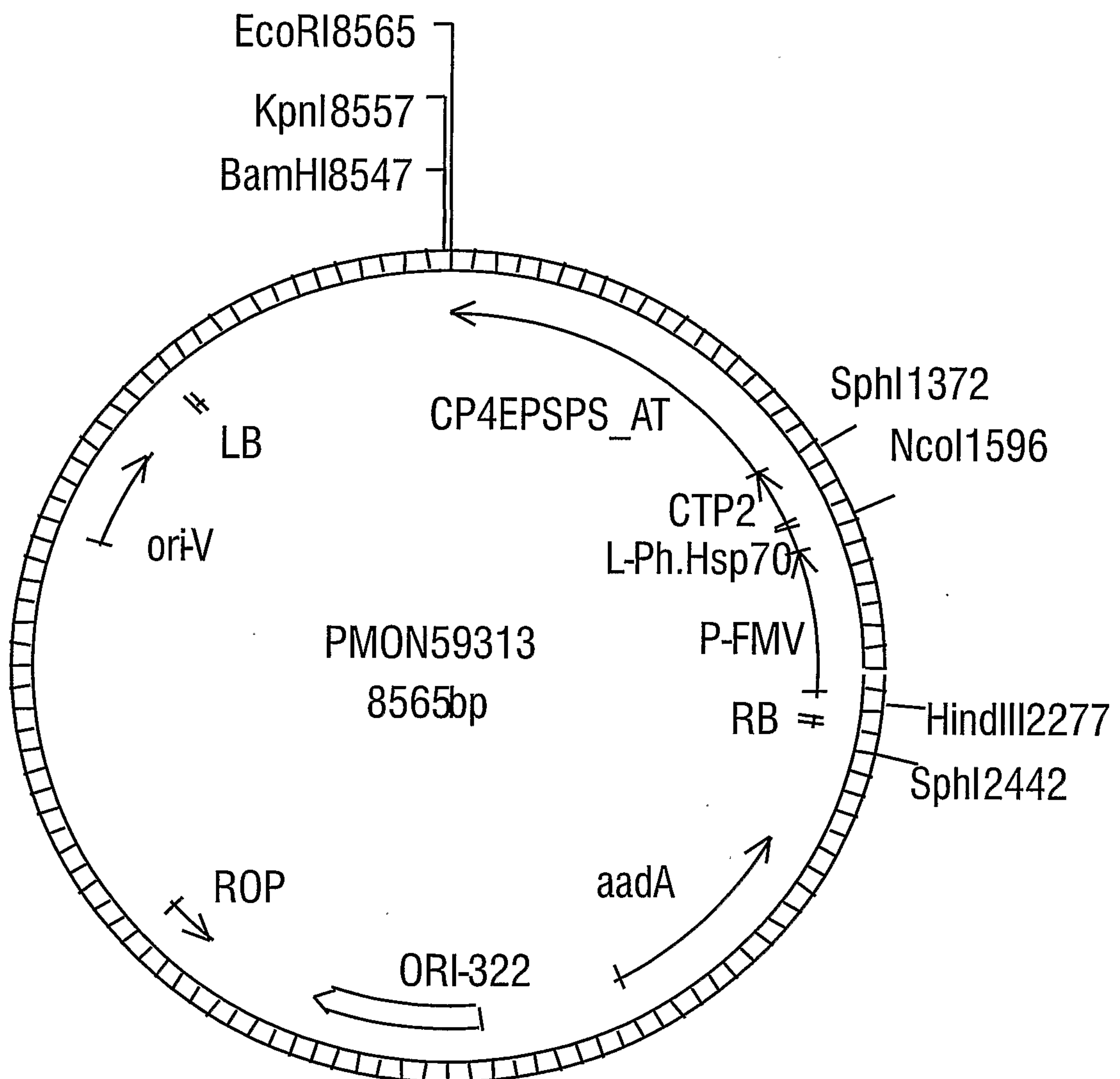


Figure 23

51/52

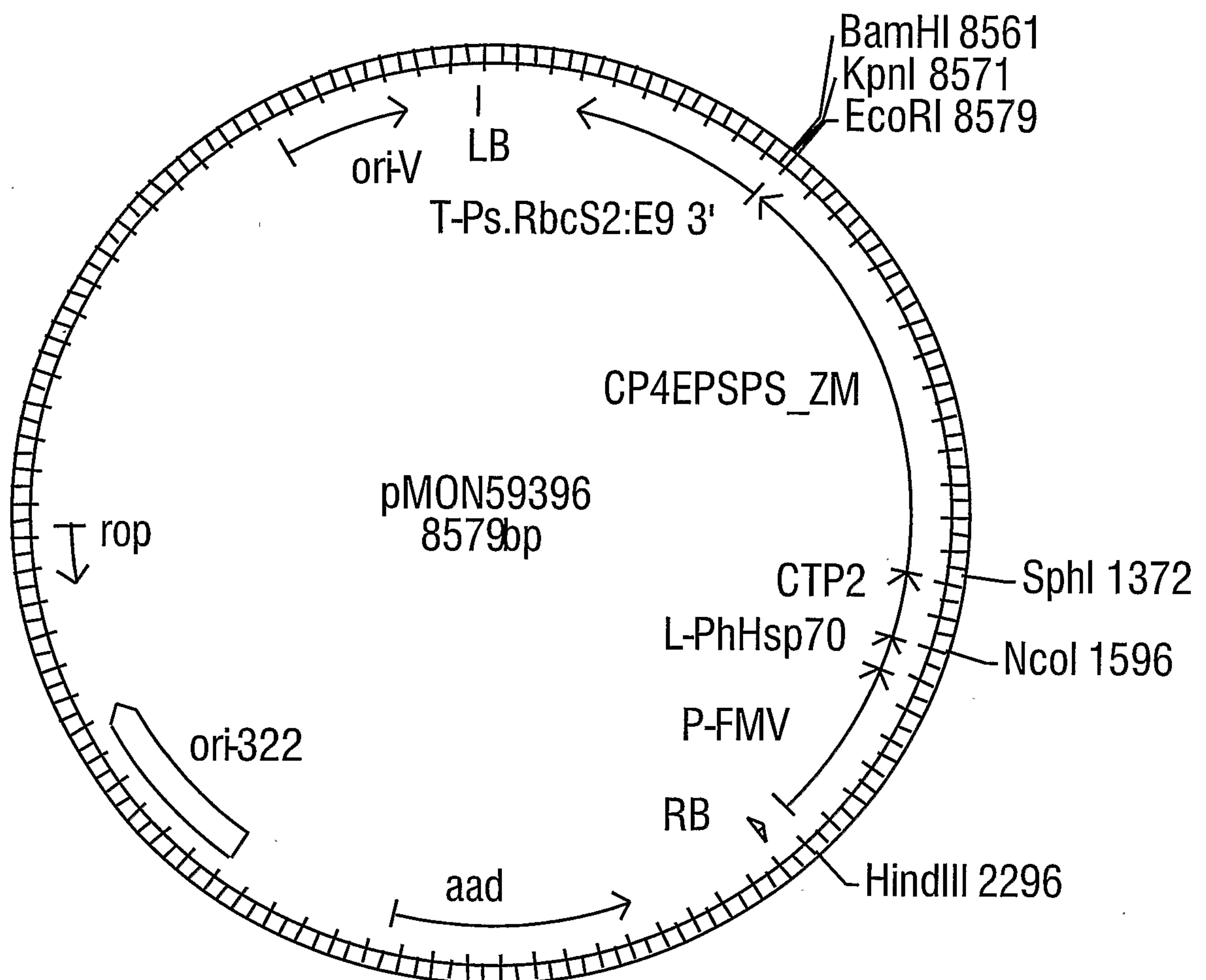


Figure 24

52/52

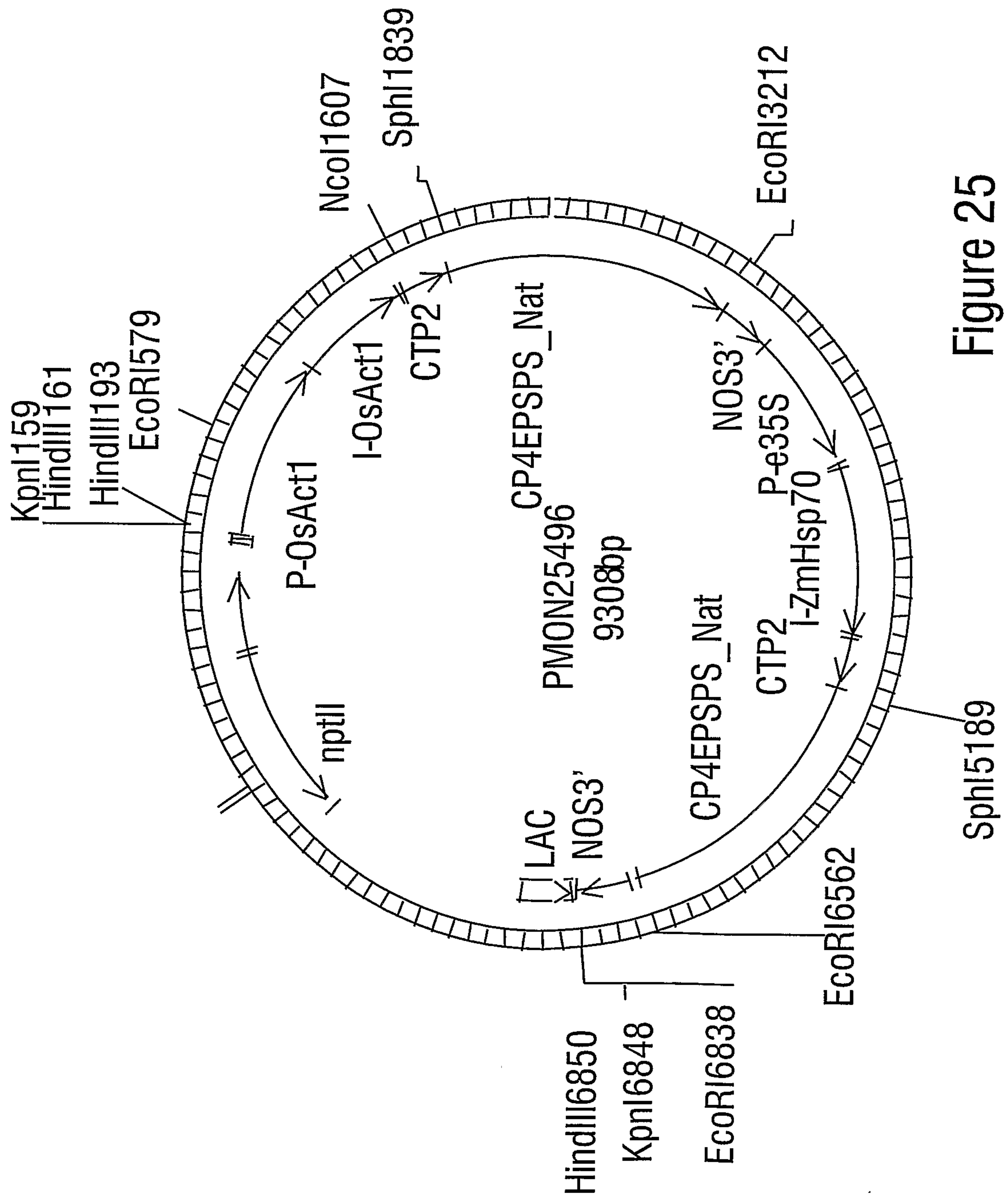


Figure 25