Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2225652 C 2007/11/20

(11)(21) 2 225 652

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 1996/08/01
(87) Date publication PCT/PCT Publication Date: 1997/02/20
(45) Date de délivrance/Issue Date: 2007/11/20
(85) Entrée phase nationale/National Entry: 1998/02/09
(86) N° demande PCT/PCT Application No.: US 1996/012671
(87) N° publication PCT/PCT Publication No.: 1997/006250
(30) Priorité/Priority: 1995/08/10 (US60/002,136)

(51) Cl.Int./Int.Cl. *C12N 15/90* (2006.01), *A01H 5/00* (2006.01), *C12N 15/65* (2006.01), *C12N 15/82* (2006.01), *C12N 5/10* (2006.01)

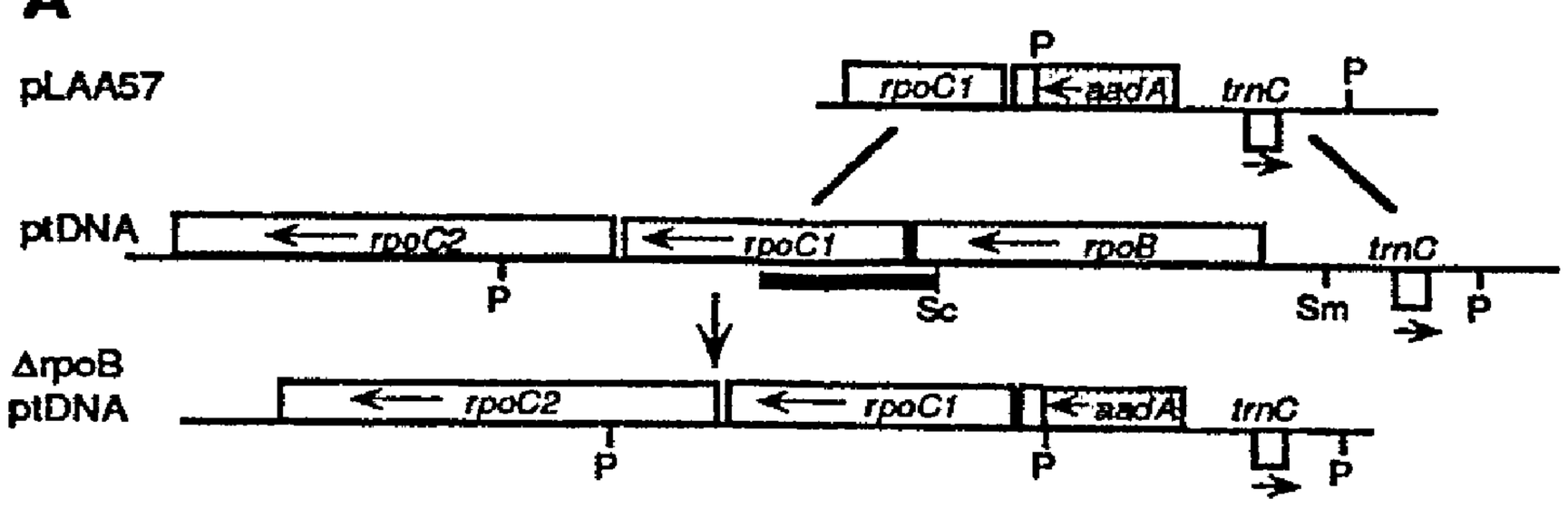
(72) Inventeurs/Inventors:
MALIGA, PAL, US;
ALLISON, LORI A., US;
HAJDUKIEWICZ, PETER T., US

(73) Propriétaire/Owner:
RUTGERS UNIVERSITY, US

(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : SYSTEME DE TRANSCRIPTION CODE PAR NUCLEOTIDES DANS LES PLASTES DE PLANTES SUPERIEURES

(54) Title: NUCLEAR-ENCODED TRANSCRIPTION SYSTEM IN PLASTIDS OF HIGHER PLANTS



(57) Abrégé/Abstract:
The present invention provides novel DNA constructs and methods for stably transforming the plastids of higher plants. The constructs described herein contain unique promoters that are transcribed by both nuclear encoded plastid polymerases and plastid encoded plastid polymerases. Use of the novel constructs of the invention facilitates transformation of a wider range of plant species and enables tissue specific expression of a transforming DNA in plastids of multicellular plants.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [6] :
C12N 15/00, 15/09, 15/11, 15/29, 15/63, 15/65, 15/82, 15/90, A01H 1/00, 1/04

A1

(11) International Publication Number: WO 97/06250

(43) International Publication Date: 20 February 1997 (20.02.97)

(21) International Application Number: PCT/US96/12671

(22) International Filing Date: 1 August 1996 (01.08.96)

(30) Priority Data:
60/002,136 10 August 1995 (10.08.95) US

(71) Applicant *(for all designated States except US)*: RUTGERS UNIVERSITY [US/US]; ASB Annex 2, Bevier Road, P.O. Box 1179, Piscataway, NJ 08855 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MALIGA, Pal [US/US]; 72 Yorktown Road, East Brunswick, NJ 08816 (US). ALLISON, Lori, A. [US/US]; 32A Cedar Lane, Highland Park, NJ 08904 (US). HAJDUKIEWICZ, Peter, T. [US/US]; 8 Larson Road, Somerset, NJ (US).

(74) Agents: HAGAN, Patrick, J. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: NUCLEAR-ENCODED TRANSCRIPTION SYSTEM IN PLASTIDS OF HIGHER PLANTS

(57) Abstract

The present invention provides novel DNA constructs and methods for stably transforming the plastids of higher plants. The constructs described herein contain unique promoters that are transcribed by both nuclear encoded plastid polymerases and plastid encoded plastid polymerases. Use of the novel constructs of the invention facilitates transformation of a wider range of plant species and enables tissue specific expression of a transforming DNA in plastids of multicellular plants.

NUCLEAR-ENCODED TRANSCRIPTION SYSTEM IN PLASTIDS OF HIGHER PLANTS

FIELD OF THE INVENTION

The present invention relates to plant genetic engineering and particularly to plastid transformation in higher plants. The invention provides a novel promoter sequences useful for the expression of foreign genes of interest in various plant species.

BACKGROUND OF THE INVENTION

Chloroplast genes are transcribed by an RNA polymerase containing plastid-encoded subunits homologous to the α, β and β' subunits of *E. coli* RNA polymerase. The promoters utilized by this enzyme are similar to *E. coli* σ^{70}-promoters, consisting of -35 and -10 consensus elements (G.L. Igloi and H. Kossel, Crit. Rev. Plant Sci. 10, 525, 1992; W. Gruissem and J.C. Tonkyn, Crit. Rev. Plant. Sci. 12:-19, 1993) Promoter selection by the plastid-encoded RNA polymerase is dependent on nuclear-encoded sigma-like factors ((Link et al. 1994, *Plant promoters and transcription factors*, Springer Verlag, Heidelberg, pp 63-83). In addition, transcription activity from some promoters is modulated by nuclear-encoded transcription factors interacting with elements upstream of the core promoter (L.A. Allison and P. Maliga, *EMBO J.*, **14**:3721-3730; R. Iratni, L. Baeza, A. Andreeva, R. Mache, S. Lerbs-Mache, *Genes Dev.* **8**, 2928, 1994). These factors mediate nuclear control of plastid gene expression in response to developmental and environmental cues.

There has been speculation of the existence of a second transcription system in plastids. However, direct evidence to support such a speculation has

heretofore been unavailable. Identification of a novel second transcription system in plastids represents a significant advance in the art of plant genetic engineering. Such a system enables greater flexibility and range in plant species available for plastid transformation, and facilitates tissue specific expression of foreign proteins and RNAs via constructs that can be manipulated by recombinant DNA techniques.

SUMMARY OF THE INVENTION

This invention provides DNA constructs and methods for stably transforming plastids of multicellular plants. The DNA constructs of the invention extend the range of plant species that may be transformed and facilitate tissue specific expression of foreign genes of interest.

According to one aspect of the invention, DNA constructs are provided that contain novel promoter sequences recognized by a nuclear encoded plastid (NEP) RNA polymerase. The DNA construct contains a transforming DNA, which comprises a targeting segment, at least one cloning site adapted for insertion of at least one foreign gene of interest, the expression of the foreign gene of interest being regulated by a promoter recognized by a NEP polymerase, and a plastid selectable marker gene.

The use of promoter elements recognized by plastid encoded plastid (PEP) RNA polymerase for enhancing expression of foreign genes of interest is another aspect of the instant invention. Like the constructs described above, these constructs also contain a targeting segment, and a cloning site for expression of a foreign gene of interest.

The promoters recognized by plastid encoded plastid RNA polymerase have been well characterized in photosynthetic tissues such as leaf. In contrast, the nuclear-encoded polymerase transcription system of the

present invention directs expression of plastid genes also in roots, seeds and meristematic tissue. In most plants, including maize, cotton and wheat, plant regeneration is accomplished through somatic embryogenesis (i.e., involving meristematic tissue). In a preferred embodiment of the invention, efficient plastid transformation in these crops will be greatly facilitated, through the use of the NEP plastid transcription system, promoters and polymerases of the present invention.

The NEP promoters of the invention are incorporated into currently available plastid transformation vectors and protocols for use thereof, such as those described in U.S Patent No. 5,451,513 and U.S. Patent No. 5,877,402, and also described by Svab & Maliga., *Proc. Natl. Acad. Sci. USA*, 90, 913 (1993). To obtain transgenic plants, plastids of non-photosynthetic tissues are transformed with selectable marker genes expressed from NEP promoters and transcribed by the nuclear-encoded polymerase. Likewise, to express proteins of interest, expression cassettes are constructed for high level expression in non-photosynthetic tissue, using the NEP promoter transcribed from the nuclear-encoded polymerase. In another aspect of the invention, PEP promoters of the invention are incorporated into currently available plastid transformation vectors and protocols for use thereof.

In yet another aspect of the invention, the NEP transcription system also may be combined with the σ^{70}-type system through the use of dual NEP/PEP promoters.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1. Deletion of *rpoB* from the tobacco plastid genome by targeted gene replacement. (A) Homologous recombination (diagonal lines) via plastid DNA

sequences flanking *aadA* in plasmid pLAA57 results in replacement of *rpoB* (Sac I to Sma I fragment) in the wild-type plastid genome (ptDNA) with *aadA* sequences, yielding the Δ*rpoB* plastid genome (Δ*rpoB* ptDNA). Abbreviations: *rpoB*, *rpoC1*, *rpoC2* are plastid genes encoding the β, β'and β'' subunits of the *E.coli*-like RNA polymerase; *aadA* is a chimeric spectinomycin-resistance gene. Restriction enzyme recognition sites: P, Pst I; Sm, Sma I; Sc, Sac I. **(B)** Pigment deficiency is associated with the deletion of *rpoB*. Total cellular DNA was isolated from green (lanes 1,3,5) and white (lanes 2,4,6) leaf tissue from three independently transformed lines (Line Nt-pLAA57-11B, lanes 1 and 2; line Nt-pLAA57-16B, lanes 3 and 4; line Nt-pLAA57-18C, lanes 5 and 6) and from wild-type green leaf tissue (Nt, lane 7). The DNA was digested with Pst I, and the gel blot was hybridized with a DNA fragment (nucleotide positions 22883-24486 of the ptDNA, numbering according to K. Shinozaki, *et al.*, (*EMBO J.* 5, 2043, 1986)) containing part of *rpoC1* (thick black line in Fig. 1A). The probe hybridizes to a 9.0 kb fragment from the wild-type genome and a 4.2 kb fragment from the Δ*rpoB* ptDNA. (C) DNA gel-blot analysis confirms the lack of wild-type ptDNA copies in white shoots of line Nt-pLAA57-10A (lane 2) and white seed progeny of a grafted chimeric plant from the same line (lane 3). DNA from wild-type green leaf tissue was loaded in lane 1. Note the absence of the wild-type pt DNA 9.0 kb fragment in Δ*rpoB* plants. The blot was prepared as for Fig. 1B.

Fig. 2. Deletion of *rpoB* results in a pigment-deficient phenotype. **(A)** Green wild-type (left), pigment-deficient Δ*rpoB* (right), and chimeric (center) plants are shown. **(B)** A flowering chimeric plant in the greenhouse. Note the white leaf margins indicating Δ*rpoB* plastids in the second leaf layer which forms the germline cells.

Fig. 3. (A) Plastids (P) in leaf mesophyll cells of Δ*rpoB* plants lack organized photosynthetic membranes. Abbreviations: N, nucleus; V, vesicles, M, mitochondrion (B) For comparison an electron micrograph of a wild-type leaf chloroplast (C) with thylakoid membranes (T) is shown. Magnification in both (A) and (B) is 7,800X.

Fig. 4. (Upper) Accumulation of plastid mRNAs for (A) photosynthetic genes and (B) genetic system genes in the Δ*rpoB* plants. Gel blots were prepared with total cellular RNA (A, 3 μg per lane; B, 5 μg per lane) from wild-type (lanes 1,3,5) and Δ*rpoB* (lanes 2,4,6) leaf tissue, and hybridized to the indicated plastid gene sequences. (Lower) Blots shown above were reprobed with *25S rDNA* sequences. Hybridization signals were quantified with a Molecular Dynamics PhosphorImager and normalized to the 25S rRNA signal. The fold excess of wild-type over Δ*rpoB* signal intensities for each probe is shown below the lanes.

Fig. 5. Transcription in the Δ*rpoB* plants initiates from a non-canonical promoter. (A) Primer extension analysis was used to map the 5' ends of *rbcL* and *16SrDNA* transcripts in wild-type (lanes 1,3) and Δ*rpoB* (lanes 2,4) plants. Primary transcripts are marked by circles (open for wild-type, closed for Δ*rpoB*), processed transcripts by a triangle. Transcripts of unknown origin are starred. The accompanying sequence ladders (loading order GATC) were generated using the same primers that served in the primer extension reactions. Numbers beside each extension product mark the distance from the first nucleotide of the coding sequence for *rbcL* and from the first nucleotide of the mature 16S rRNA. (B) Mapping primary transcripts for 16S rRNA in wild-type (lane 2) and Δ*rpoB* (lane 3) plants. Total leaf RNA (20 μg) was capped *in vitro*, and capped

16S rRNA species were identified by RNAse protection after hybridization with a complementary RNA probe. Capped protected products are marked as in (A). Lane 1 contains RNA standards of the sizes indicated. (C) DNA sequence of the *16SrDNA* upstream region with transcripts initiating from promoters for the plastid-encoded (σ^{70}-type, P1) and nuclear-encoded (P2) polymerases (designation of P1 and P2 is based on A. Vera and M. Sugiura, *Curr. Genet.* 27, 280, 1995). Consensus σ^{70} promoter elements (-35 and -10) are boxed. Initiation sites are marked by circles, as in (A) and (B). Numbering begins from the first nucleotide upstream of the *16SrDNA* coding region (-1 = nucleotide 102757 in the tobacco plastid genome).

Fig. 6. Accumulation of plastid mRNAs in wild-type and ΔrpoB tobacco leaves. Blots for the plastid genes (see Example I) are grouped as follows. (A) mRNA is significantly more abundant in the leaves of wild-type than in ΔrpoB plants. (B) Levels of mRNA are comparable in wild-type and ΔrpoB leaves, or (C) are higher in ΔrpoB leaves. Gel blots were prepared with total cellular RNA (3 μg per lane) from wild-type (lanes 1) and ΔrpoB (lanes 2) leaf tissue, and hybridized to the indicated plastid gene sequences. (Lower panel) To control for loading, blots shown above were reprobed with 25S rDNA sequences.

Fig. 7. Mapping atpB transcription initiation sites in wild-type and ΔrpoB tobacco leaves. (A) Primer extension analysis. End-labeled primer extension products from wild-type (wt) and ΔrpoB (T57) samples were run alongside the homologous sequence obtained by using the same primer. Numbers alongside the sequence refer to the distance from the ATG translation initiation codon. Primary transcripts from NEP and PEP promoters are marked by filled and open circles, respectively. (B) In vitro

capping and RNase protection assay to identify primary transcript 5' ends. Lanes were loaded with ΔrpoB (T57; 1, 2) and wild-type (wt; 4,5) RNA samples with (2,4) and without (1,5) protecting complementary antisense RNA. Molecular weight (MW) markers (100, 200, 300, 400, and 500 nucleotides) were loaded in lane 3. The transcript 5' end in (A) corresponds to the protected fragment size in brackets: -255 (277nt), -289 (311). Note artifact slightly below the 200 nt marker which is present in the unprotected RNA samples. (C) Physical map of the atpB - rbcL intergenic region. Map position of the primary transcript 5' ends for the atpB NEP and PEP promoters are marked as in (A).

Fig. 8. Mapping atpI transcription initiation sites in wild-type and ΔrpoB tobacco leaves. (A) Primer extension analysis. End-labeled primer extension products from wild-type (wt) and ΔrpoB (T57) samples were run alongside the homologous sequence obtained by using the same primer. Numbers alongside the sequence refer to the distance from the ATG translation initiation codon. Primary transcripts from NEP and PEP promoters are marked by filled and open circles, respectively. (B) In vitro capping and RNAse protection assay to identify primary transcript 5' ends. Lanes were loaded with ΔrpoB (T57; 1, 2) and wild-type (wt; 4,5) RNA samples with (2,4) and without (1,5) protecting complementary antisense RNA. Molecular weight (MW) markers (100, 200, 300, 400, and 500 nucleotides) were loaded in lane 3. The transcript 5' end in (A) corresponds to the protected fragment size in brackets: -130 (235nt) , -207, -209, -212 (303, 305, 309; not resolved). Note artifact slightly below the 200 nt marker which is present in the unprotected RNA samples. (C) Physical map of the rps2 -atpI intergenic region. Map position of the primary transcript 5' ends for the atpI NEP and PEP promoters are marked as in (A).

Fig. 9. Mapping clpP transcription initiation sites in wild-type and ΔrpoB tobacco leaves. (A) Primer extension analysis. End-labeled primer extension products from wild-type (wt) and ΔrpoB (T57) samples were run alongside the homologous sequence obtained by using the same primer. Numbers alongside the sequence refer to the distance from the ATG translation initiation codon. Primary transcripts from NEP and PEP promoters are marked by filled and open circles, respectively. (B) In vitro capping and RNAse protection assay to identify primary transcript 5' ends. Lanes were loaded with ΔrpoB (T57; 1, 2) and wild-type (wt; 4,5) RNA samples with (2,4) and without (1,5) protecting complementary antisense RNA. Molecular weight (MW) markers (100, 200, 300, 400, and 500 nucleotides) were loaded in lane 3. The transcript 5' end in (A) correspond to the protected fragment size in brackets: -53 (96 nt), -95 (138 nt), -173 (216 nt) and -511 (69 nt). Note artifact slightly below the 200 nt marker which is present in the unprotected RNA samples. (C) Physical map of the clpP - psbB intergenic region. Map position of the primary transcript 5' ends for the clpP NEP and PEP promoters are marked as in (A).

Fig. 10. Mapping accD transcription initiation sites in wild-type and ΔrpoB tobacco leaves. (A) Primer extension analysis. End-labeled primer extension products from wild-type (wt) and ΔrpoB (T57) samples were run alongside the homologous sequence obtained by using the same primer. Numbers alongside the sequence refer to the distance from the ATG translation initiation codon. Primary transcript for the PaccD-129 NEP promoter is marked by a filled circle. (B) In vitro capping and RNAse protection assay to identify primary transcript 5' ends. Lanes were loaded with ΔrpoB (T57; 1, 2) and wild-type (wt; 4,5) RNA samples with (2,4) and without (1,5) protecting complementary antisense RNA. Molecular

weight (MW) markers (100, 200, 300, 400, and 500 nucleotides) were loaded in lane 3. The -57 transcript 5' end in (A) corresponds to the protected 103 nt fragment. Note artifact slightly below the 200 nt marker which is present in the unprotected RNA samples. (C) Physical map of the accD -- rbcL intergenic region. Map position of the primary transcript 5' end for the PaccD-129 NEP promoter is marked.

Fig. 11. Alignment of DNA sequences flanking the NEP promoter transcription initiation sites. Nucleotides with more than 6 matches are boxed. Consensus sequence adjacent to the transcription initiation site is shown below. Position of 5' ends are marked by filled circles. Note, that 5' ends for Prps12-152 and Prps16-107 were not capped and may not be primary transcripts. "Prrn-62, PORF2280-1577, PatpB-289, Prbs2-152, Prps16-107, PORF1901-41, PatpI-207, Pclp-511, Pclp-173, PaccD-129, NEP consensus sequence, and Pclp-53 respectively are shown."

Fig. 12. NEP and PEP polymerases, through recognition of distinct promoters, provide a mechanism for selective transcription of plastid genes. Note that some genes have only PEP promoters (photosystem I and photosystem II), others have both PEP and NEP promoters (most housekeeping genes), or only NEP promoters (accD).

Fig. 13. A schematic diagram of a chimeric plastid gene expressed from a NEP promoter.

DESCRIPTION OF THE INVENTION

Several reports have suggested the existence of an additional plastid-localized, nuclear-encoded RNA polymerase (reviewed in Gruissem and Tonkyn, 1993; Igloi and Kossel, 1992; Mullet, 1993; Link, 1994). By deleting the rpoB gene encoding the essential β subunit of the tobacco E. coli-like RNA polymerase. The existence of a second plastid transcription system which is encoded by the nucleus has been established (Allison et al., 1996, EMBO J. 15:2802-2809). Deletion of rpoB yielded photosynthetically defective, pigment-deficient

plants. An examination of plastid ultrastructure in leaf mesophyll cells of the ΔrpoB plants revealed proplastid-like organelles lacking the arrays of stacked thylakoid membranes which are characteristic of photosynthetically-active chloroplasts. Transcripts for the rbcL, psbA and psbD photosynthetic genes were low, whereas mRNAs for the rpl16, atpI and 16SrDNA genes accumulated to about wild-type, or higher than in wild-type levels. Lack of transcript accumulation for the photosynthetic genes was due to lack of σ^{70}-type promoter activity. While in wild-type tobacco leaves the ribosomal RNA operon is normally transcribed from a σ^{70}-type promoter, in the ΔrpoB plants the rRNA operon was transcribed from a non-σ^{70} promoter. The rRNA operon is the first transcription unit for which both a plastid-encoded and nuclear-encoded plastid RNA polymerase (PEP and NEP, respectively) was identified.

An analysis of the promoter regions of other genes has revealed that the rRNA operon is not unique. It is a member of a large class of plastid genes which have at least one promoter each for PEP and NEP, with a potential for expression by either of the two plastid RNA polymerases. In addition, plastid genes have been identified which are transcribed exclusively by NEP. Furthermore, the data suggest that additional gene-specific mechanisms regulate NEP transcript levels in different plastid types.

A NEP transcriptional start site has been identified about 62 bases upstream of the mature 16S rRNA 5' terminus. The sequence surrounding the initiation site is highly conserved among numerous plant species examined, and bears no resemblance to the PEP promoter consensus sequence. NEP promoter consensus sequences important for nuclear encoded polymerase recognition and binding (analogous to the -10 and -35 sequences of the *E. coli*-type transcription initiation site) are preferentially located within about 50 nucleotides in

either direction of the NEP transcription start site. As described in greater detail in Example 1, several different NEP promoters exist, and NEP promoters are sometimes found in conjunction with PEP promoters.

The polymerases of the invention may be purified by chromatography, using standard methods. The NEP polymerase activity in column fractions can be assayed utilizing DNA segments comprising the NEP promoter region as templates in *vitro* transcription reactions. Alternatively, NEP promoter segments may be attached to a matrix which is separable by some means (e.g., magnetic beads). The matrix-bound DNA is incubated with a plant extract under conditions in which the nuclear encoded polymerase is expected to bind the DNA. The matrix/DNA/polymerase complex is then separated from the plant extract, and the bound protein may then be isolated and characterized. The protein purified by either of the above mentioned protocols may be used to produce antibodies to probe expression libraries, for the purpose of isolating the nuclear genes or cDNAs encoding the nuclear encoded polymerase.

As an alternative approach for the isolation of the NEP polymerase, proteins with specific affinity for the promoter fragment can be isolated and the N-terminal amino acid sequence can be determined by microsequencing. The amino acid sequence can then be used to design appropriate PCR primers for gene isolation.

The activity of the previously-known plastid-encoded σ^{70}-type transcription system has been well characterized in photosynthetically active tissues, such as leaf. In contrast, the nuclear-encoded polymerase transcription system of the present invention directs expression of plastid genes also in roots, seeds and meristematic tissue. In most plants, including maize, cotton and wheat, plant regeneration is accomplished through somatic embryogenesis (i.e., involving meristematic tissue). Efficient plastid transformation

in these crops will be enabled, or greatly facilitated, through the use of the NEP plastid transcription system of the present invention.

The NEP promoters of the invention can be incorporated into currently available plastid transformation vectors and protocols for use thereof, such as those described in U.S. Patent No. 5,451,513 and U.S. Patent No. 5,877,402 and also described by Svab & Maliga., *Proc. Natl. Acad. Sci. USA*, 90, 913 (1993).

To obtain transgenic plants, plastids of non-photosynthetic tissues are transformed with selectable marker genes expressed from NEP promoters and transcribed by the nuclear-encoded polymerase. Likewise, to express proteins of interest, expression cassettes are constructed for high level expression in non-photosynthetic tissue, using the NEP promoter transcribed from the nuclear-encoded polymerase. The NEP transcription system also may be combined with the σ^{70}-type system through the use of dual NEP/PEP promoters. In some cases, expression of transgenes from NEP promoters in photosynthetic tissue also may be desirable.

The detailed description set forth in Examples I-III below describes preferred methods for making and using the DNA constructs of the present invention and for practicing the methods of the invention. Any molecular cloning and recombinant DNA techniques not specifically described are carried out by standard methods, as generally set forth, for example, in Ausubel (ed.), Current Protocols in Molecular Biology. John Wiley & Sons, Inc.(1994).

The following nonlimiting Examples describe the invention in greater detail.

EXAMPLE 1
Demonstration of a Second Distinct Plastid Transcription System by Deletion of *rpoB*

To establish the existence of a non-*E. coli*-

like RNA polymerase in plastids, the gene for one of the essential subunits of the *E. coli*-like enzyme was deleted from the tobacco plastid genome. mRNA levels were then assessed in mutant plastids. The data indicate that, in the absence of the plastid-encoded *E. coli*-like enzyme, expression of some photogenes is dramatically reduced. In contrast, transcript levels for the plastid genes encoding the gene expression apparatus are similar to levels in wild-type plants. Therefore the non-*E.coli*-like RNA polymerase selectively transcribes a subset of plastid genes. This second transcription apparatus does not initiate from typical *E. coli* σ^{70}-promoters but recognizes a novel promoter sequence.

Materials and Methods for Example I

Plasmid construction. Plasmid pLAA57 is a pBSKS+ (Stratagene) derivative which carries a Sac I to Bam HI fragment (nucleotides 22658 to 29820) of the ptDNA. An internal Sac I to Sma I DNA fragment within the ptDNA insert, between nucleotides 24456 and 28192, was replaced by a chimeric spectinomycin-resistance (*aadA*) gene. The *aadA* gene is identical to that described (Z. Svab and P. Maliga, *Proc. Natl. Acad. Sci. USA,* **90**, 913, 1993), except that the *psbA* 3' region is shorter and is contained in an Xba I to Dra I fragment as described (J.M. Staub and P. Maliga, *Plant J.* **6**, 547, 1994).

Plant Transformation. For plastid transformation tungsten particles were coated (Z. Svab and P. Maliga, *Proc. Natl. Acad. Sci. USA,* **90**, 913, 1993) with pLAA57 DNA, and introduced into the leaves of *Nicotiana tabacum* plants using the DuPont PDS1000He Biolistic gun at 1100 p.s.i. Transgenic shoots were selected aseptically on RMOP medium (Z. Svab, P. Hajdukiewicz, P. Maliga, *Proc. Natl. Acad. Sci. USA* **87**,

8526, 1990) containing 500mg/ml spectinomycin dihydrochloride. Transgenic cuttings were rooted and maintained on RM medium consisting of agar-solidified MS salts (T. Murashige and F. Skoog, *Physiol. Plant.*, 15, 493, 1962) containing 3% sucrose.

Electron Microscopy. Electron microscopy was done on fully expanded leaves from wild-type and Δ*rpoB* cuttings grown in sterile culture on RM medium with 3% sucrose. Tissue was fixed for 2 hours in 2% glutaraldehyde, 0.2M sucrose, 0.1M phosphate buffer (pH 6.8) at room temperature, and washed three times in 0.2M sucrose, 0.1M phosphate buffer. Fixed tissues were postfixed in buffered 1% osmium tetroxide with 0.2M sucrose, dehydrated in a graded ethanol series, embedded in Spurr's epoxy resin (hard), sectioned, and stained with uranyl acetate and lead citrate for transmission electron microscopy.

Gel blots. Total leaf DNA was prepared as described (I.J. Mettler, *Plant Mol. Biol. Rep.*, 5, 346, 1987), digested with restriction endonuclease Pst I, separated on 0.7% agarose gels, and transferred to Hybond N* (Amersham) using the Posiblot* Transfer apparatus (Stratagene). Hybridization to a random-prime labeled fragment was carried out in Rapid Hybridization Buffer (Amersham) overnight at 65°C. Total leaf RNA was prepared using TRIzol* (GIBCO BRL), following the manufacturer's protocol. The RNA was electrophoresed on 1% agarose/formaldehyde gels, then transferred to nylon membrane and probed as for the DNA blots.

Synthesis of Probes. Double-stranded DNA probes for *psbA*, *atpI*, and *rpl16* were prepared by random-primed ^{32}P-labeling of PCR-generated DNA fragments. The sequence of the primers used for PCR, along with their positions within the tobacco ptDNA (K. Shinozaki, *et al.*

* Trade-mark

, *EMBO J.* **5**, 2043, 1986) are as follows: *psbA*, 5' primer = 5'-CGCTTCTGTAACTGG-3' (complementary to nucleotides 1550 to 1536 of the ptDNA), 3' primer = 5'-TGACTGTCAACTACAG-3' (nucleotides 667 to 682); *atpI* 5' primer = 5'-GTTCCATCAATACTC-3' (complementary to nucleotides 15985 to 15971), 3' primer = 5'-GCCGCGGCTAAAGTT-3' (nucleotides 15292 to 15306); *rpl16* 5' primer = 5'-TCCCACGTTCAAGGT-3' (complementary to nucleotides 84244 to 84230), 3' primer = 5'-TGAGTTCGTATAGGC-3' (nucleotides 83685 to 83699). To generate probes for *rbcL*, *psbD/C* and 16S rRNA, the following restricted DNA fragments were ^{32}P-labeled: *rbcL*, a Bam HI fragment (nucleotides 58047 to 59285 in the ptDNA); *psbD/C*, a Sac II to Hind III fragment of the tobacco *psbD/C* operon (nucleotides 34691-36393); 16S rRNA, an Eco RI to Eco RV fragment (nucleotides 138447 to 140855 in the ptDNA).

The probe for tobacco 25S rRNA was from plasmid pKDR1 (D. Dempsey, K.W. Wobbe, D.F. Klessig, *Mol. Plant Path.* **83**, 1021, 1993) containing a 3.75 kb Eco RI fragment from a tobacco 25S/18S locus cloned in plasmid pBR325. When hybridizing gel-blots for 25S rRNA, ^{32}P-labeled double-stranded DNA probe was mixed with unlabeled plasmid pKDR1 corresponding to a 2-fold excess over the amount of RNA present on the filter.

Normalizing DNA levels by plastid genome copy number. To test whether changes in plastid genome copy number contributed to the estimated differences in gene expression, total cellular DNA and RNA were prepared from equal amounts of leaf tissue from wild-type and *ΔrpoB* plants. To compare the number of plastid genome copies per equivalent leaf mass, DNA gel-blots were carried out with an equal volume of each DNA preparation and probed with a radiolabeled Eco RI to Eco RV fragment (from nucleotides 138447 to 140845 of ptDNA (K. Shinozaki, *et al.*, *EMBO J.* **5**, 2043, 1986) of *16SrDNA* sequence.

Quantitation by PhosphorImage* analysis demonstrated an equal number of plastid genome copies in each sample. The amount of 16S rRNA from equal tissue samples, as measured by RNA gel-blots on equal volumes of each RNA preparation, was reduced by 2.5-fold in the Δ*rpoB* plants. This value is similar to the 3-fold reduction estimated when normalizing with the cytoplasmic 25S rRNA signal (Fig. 3B).

Primer extension reactions. Primer extension reactions were carried out on 3 μg (wild-type) or 10 μg (Δ*rpoB*) of total leaf RNA as described (L.A. Allison and P. Maliga, *EMBO J.*, in press) using the following primers: 16S rRNA: 5'-TTCATAGTTGCATTACTTATAGCTTC-3' (complementary to nucleotides 102757-102732); *rbcL*: 5'-ACTTGCTTTAGTCTCTGTTTGTGGTGACAT (complementary to nucleotides 57616-57587). Sequence ladders were generated with the same primers using the Sequenase II* kit (USB).

Identification of primary transcripts by in vitro capping. Total leaf RNA (20 μg) from wild-type and Δ*rpoB* plants was capped in the presence of [α-^{32}P]GTP (J.C. Kennell and D.R. Pring, *Mol. Gen. Genet.* 216, 16, 1989). Labeled 16S rRNAs were detected by ribonuclease protection (A. Vera and M. Sugiura, *Plant Mol. Biol.* 19, 309, 1992) using the RPAII kit (Ambion). To prepare the protecting complementary RNA, the *16SrDNA* upstream region (nucleotides 102526-102761 of the ptDNA) was PCR-amplified using the following primers: 5' primer was 5'-CCTCTAGACCCTAAGCCCAATGTG-3' corresponding to nucleotides 102526 and 102541 of the ptDNA (K. Shinozaki, *et al.*, *EMBO J.* 5, 2043, 1986), underlined) plus an XbaI site; 3' primer was 5'-CCGGTACCGAGATTCATAGTTGCATTAC-3' complementary to nucleotides 102761 to 102742 of the ptDNA (underlined) plus a Kpn I site. The amplified product was cloned as an Xba I to Kpn I fragment into Xba

* Trade-mark

I and Kpn I-restricted pBSKS+ vector (Stratagene). To generate unlabeled RNA complimentary to the 5' end of 16S rRNAs, the resulting plasmid was linearized with Xba I, and transcribed in a Megascript* (Ambion) reaction with T3 RNA polymerase. Markers (100, 200, 300, 400, and 500 nucleotides) were prepared with the RNA Century Markers Template Set (Ambion), following the manufacturer's protocol. The 72 nucleotide marker was the mature processed transcript from the plastid *trnV* gene, and was generated by RNAse protection.

Results and Discussion

Disruption of the E. coli-like RNA polymerase activity in tobacco plastids results in a pigment-deficient phenotype. To avoid disrupting plastid genes for other functions the *rpoB* gene was targeted for deletion, since it is the first reading frame of an operon encoding exclusively subunits of the *E. coli*-like plastid polymerase (K. Shinozaki, *et al.* , *EMBO J.* 5, 2043, 1986). The deletion was accomplished by replacing most of the *rpoB* coding region (3015 out of 3212 base pairs) and 691 bp of upstream non-coding sequence, with a chimeric spectinomycin resistance (*aadA*) gene (Z. Svab and P. Maliga, *Proc. Natl. Acad. Sci. USA,* 90, 913, 1993) in a cloned plastid DNA (ptDNA) fragment. The resulting plasmid was introduced by particle bombardment into tobacco chloroplasts, where the *aadA* gene integrated into the plastid genome via flanking plastid DNA sequences as diagrammed in Fig. 1A. Since the plastid genetic system is highly polyploid, with every leaf cell containing up to 10,000 identical copies of the ptDNA, selective amplification of transformed genome copies was carried

* Trade-mark

out by growing the bombarded tissue on spectinomycin-containing medium (P. Maliga, *Trends Biotechnol*. 11, 101, 1993).

From the initial round of selection several spectinomycin-resistant plants exhibiting sectors of white leaf tissue were obtained (Fig. 2A). DNA gel-blot analysis of white and green sectors indicated that the pigment-deficiency was correlated with deletion of *rpoB* in three independently transformed lines (Fig. 1B). Most DNA samples from the pigment-deficient tissue, for example lane 4 in Fig. 1B, contained a mix of wild-type and transformed genome copies. The complete absence of wild-type ptDNA copies was critical for the interpretation of the data. Therefore, to obtain plants containing only transformed plastid genomes, shoots were regenerated from the white tissue sectors. This procedure yielded uniformly white plants (Fig. 2A) which contained no wild-type ptDNA as judged by DNA gel-blot analysis (Fig. 1C). Regeneration from these white leaves on spectinomycin-free medium yielded exclusively pigment-deficient shoots, confirming the complete absence of wild-type plastid genomes in all leaf layers and cell types.

It is difficult to obtain seed from tobacco plants grown in sterile culture. Fortuitously, during plant regeneration from primary transformants, we obtained a periclinal chimera (S. Poethig, *Trends Genetics* 5, 273, 1989) homoplasmic for the plastid mutation in the L2 leaf layer (Fig. 2A). This line was grafted on wild-type tobacco and was raised to maturity in the greenhouse (Fig. 2B). Seed from self-pollinated flowers gave rise to uniformly white seedlings, in which no wild-type plastid genomes could be detected by DNA gel-blot analysis (Fig. 1C).

Plastids in leaves of the ΔrpoB plants lack thylakoid membranes. The pigment-deficient *ΔrpoB* plants

were unable to grow photoautotrophically. However, if maintained on sucrose-containing medium to compensate for their lack of photosynthesis, they grew normally but at a reduced rate compared to wild-type plants, and exhibited no noticeable changes in organ morphology. Moreover, Δ*rpoB* seedlings germinated at a high efficiency, and developed into plants. These observations indicate that the *E. coli*-like plastid RNA polymerase is not required for maintenance of the nonphotosynthetic plastid functions necessary for plant growth and differentiation.

An examination of plastid ultrastructure in leaf mesophyll cells of the Δ*rpoB* plants revealed that the mutant plastids were smaller and rounder than wild-type chloroplasts, averaging 2-5 μm in length as compared to 5-9 μm for wild-type chloroplasts. The Δ*rpoB* plastids are thus larger than undifferentiated proplastids whose average size is 1 μm (M.R. Thomas and R.J. Rose, *Planta* **158**, 329, 1983). In addition, Δ*rpoB* plastids typically contained multiple vesicles of irregular size and shape, and lacked the arrays of stacked thylakoid membranes which are characteristic of photosynthetically-active chloroplasts (Fig. 3).

Transcription of plastid genes is maintained in ΔrpoB plastids. In the absence of the β subunit, no transcription was expected from plastid σ^{70}-type promoters. To determine whether any transcription activity was maintained in the Δ*rpoB* plastids, accumulation of RNAs was examined by RNA gel-blot analysis. Transcripts were surveyed for two different classes of plastid genes (K. Shinozaki, *et al.*, *EMBO J.* **5**, 2043, 1986). The first group included genes encoding subunits of the photosynthetic apparatus: the *psbD/C* operon, encoding subunits D2 and CP43 of photosystem II; *rbcL*, encoding the large subunit of ribulose-1,5-

bisphosphate carboxylase; and *psbA*, encoding the D1 subunit of the photosystem II reaction center. The second group contained genes for components of the gene expression apparatus: *rpl16* encoding a ribosomal protein subunit, and the *16SrDNA* gene. All plastid RNA quantitations were normalized to cytoplasmic 25S ribosomal RNA levels.

Surprisingly, accumulation of mRNAs was detected for all the genes examined. However, the effect of the *rpoB* deletion on transcript accumulation was dramatically different for the two classes of genes. The steady-state mRNA levels of the photosynthetic genes *psbD/C*, *rbcL*, and *psbA*, were reduced 40- to 100-fold compared to wild-type levels (Fig. 4A; signals were visible in all Δ*rpoB* lanes upon longer exposure). In contrast, transcript levels for nuclear encoded polymerase genes were much less affected. A 3-fold reduction for 16S rRNA was measured, and an actual increase for the multiple transcripts arising from the polycistronic operon containing the *rpl16* gene was also observed (Fig. 4B). These data indicate that, while expression of genes encoding the photosynthetic apparatus is defective in the Δ*rpoB* plants, the RNAs for genes involved in housekeeping functions accumulate to approximately wild-type, or higher, levels.

The 16SrDNA gene is transcribed from a novel promoter in ΔrpoB plants. The accumulation of plastid RNAs confirmed that there is RNA polymerase activity in plastids lacking the β subunit of the *E. coli*-like enzyme. However, migration of plastid genes to the nucleus has been documented (S.L. Baldauf and J.D. Palmer, *Nature* **334**, 262, 1990; J.S. Gantt, S.L. Baldauf, P.J. Calie, N.F. Weeden, J.D. Palmer, *EMBO J.* **10**, 3073, 1991; M.W. Gray, *Curr. Op. Genet. Dev.* 3, 884, 1993). Therefore, transcription in Δ*rpoB* plastids could still

conceivably initiate from σ^{70}-type promoters if there existed a nuclear copy of *rpoB*, whose product could be imported into plastids and assembled into functional *E. coli*-like enzyme. To establish whether the plastid transcripts detected in Δ*rpoB* plants were products of transcription from a σ^{70}-type promoter, the 5' transcript ends for four genes were mapped, *rbcL* (K. Shinozaki and M. Sugiura, *Gene* **20**, 91, 1982), *16SrDNA* (A. Vera and M. Sugiura, *Curr. Genet.* **27**, 280, 1995), *psbA* (M. Sugita and M. Sugiura, *Mol. Gen. Genet.* **195**, 308, 1984) and *psbD* (W.B. Yao, B.Y. Meng, M. Tanaka, M. Sugiura, *Nucl. Acids Res.* **17**, 9583, 1989), for which the transcription initiation sites have been established previously. None of the 5' ends mapped to σ^{70}-type promoter initiation sites (data are shown for *rbcL* and *16SrDNA* in Fig. 5A). Therefore it was concluded that the residual RNA polymerase activity in the Δ*rpoB* plastids was not due to an *E. coli*-like enzyme, but represents a second unique plastid transcription system. This distinct RNA polymerase enzyme is referred to as the Nuclear Encoded Plastid RNA polymerase (NEP), to distinguish it from the *E. coli*-like enzyme which we designate Plastid Encoded Plastid RNA polymerase (PEP). Since the tobacco plastid genome has been fully sequenced, and since the few unidentified reading frames bear no sequence similarity to known RNA polymerase subunits (K. Shinozaki, *et al.*, *EMBO J.* **5**, 2043, 1986), transcription by the nuclear encoded RNA polymerase relies on nuclear gene products.

In the absence of transcription from σ^{70}-type promoters in the Δ*rpoB* plants, the question remained: what promoters were the source of the plastid RNAs. The 16S rRNA 5' end detected in the Δ*rpoB* plants mapped 62 nucleotides upstream of the mature 16S rRNA 5' terminus (Fig. 5A). This 5' end was determined to be a primary transcript by *in vitro* capping (Fig. 5B). A prominent primary transcript with a similar 5' end was recently reported in proplastids of heterotrophically-

cultured tobacco cells, and was designated P2 (A. Vera and M. Sugiura, *Curr. Genet.* **27**, 280, 1995; This transcript is also present at very low levels in wild-type leaf cells (A. Vera and M. Sugiura, *Curr. Genet.* **27**, 280, 1995; Fig. 5A longer exposure, not shown). The sequence surrounding the initiation site is highly conserved among all plant species examined, and bears no resemblance to the σ^{70} consensus sequence (A. Vera and M. Sugiura, *Curr. Genet.* **27**, 280, 1995). Based on its prominent usage in the Δ*rpoB* plants, it was concluded that this unique promoter is utilized by the NEP transcription apparatus.

In contrast to the 16S rRNA, the major transcripts for the photosynthetic genes *rbcL*, and *psbD/C* mapped to previously-characterized processed ends (data shown for *rbcL* Fig 5; L. Hanley-Bowdoin, E.M. Orozco, N.-H. Chua, *Mol. Cell. Biol.* **5**, 2733, 1985; J.E. Mullet, E.M. Orozco, N.-H. Chua, *Plant Mol. Biol.* **4**, 39, 1985; S. Reinbothe, C. Reinbothe, C. Heintzen, C. Seidenbecher, B. Parthier, *EMBO J.* **12**, 1505, 1993). Additional minor transcript ends mapped upstream of the processed termini. Therefore, the low levels of transcript accumulation for these photosynthetic genes are the result of upstream promoter activity and subsequent processing of the readthrough RNAs to yield correctly-sized transcripts.

Proposed roles for the two plastid transcription systems. In the Δ*rpoB* plants there is accumulation of RNAs transcribed by the NEP system. This indicates a role for the nuclear encoded RNA polymerase in maintaining the expression of plastid housekeeping genes. Apparently these expression levels are sufficient to support the growth and differentiation of non-photoautotrophic plants. In contrast, the *E. coli*-like

PEP RNA polymerase is required to provide the high levels of plastid gene transcripts necessary for development of photosynthetically active chloroplasts. The proposed role for the nuclear encoded RNA polymerase implies a high demand for its function during the early phases of chloroplast development, before the PEP RNA polymerase is active (J.E. Mullet, *Plant Physiol.*, **103**, 309, 1993). Developmental regulation of a nuclear-encoded RNA polymerase is supported by the observation that the nuclear encoded polymerase P2 promoter of the *16SrDNA* gene is more active in proplastids of cultured tobacco cells than in leaf chloroplasts (A. Vera and M. Sugiura, *Curr. Genet.* **27**, 280, 1995).

EXAMPLE II

Transcription by two distinct RNA polymerases is a general regulatory mechanism of gene expression in higher plants

As described in Example I, accumulation of transcripts in plants lacking the PEP polymerase led to the identification of a NEP promoter for the plastid ribosomal RNA operon (Allison et al. 1996, EMBO J. **14**:3721-3730). To facilitate mapping of additional NEP promoters, mRNA accumulation was examined in ΔrpoB plants for most classes of plastid genes. The novel promoter sequences described herein may be used to extend the range of species within such plastid transformation is feasible and to drive expression of foreign genes of interest in a tissue specific manner.

Materials and Methods for Example II

RNA Gel blots Total leaf RNA was prepared using TRIzol (GIBCO BRL), following the manufacturer's

protocol. The RNA was electrophoresed on 1% agarose/formaldehyde gels, then transferred to Hybond N (Amersham) using the Posiblot Transfer apparatus (Stratagene). Hybridization to random-primer labeled fragment was carried out in Rapid Hybridization Buffer (Amersham) overnight at 65°C. Double-stranded DNA probes were prepared by random-primed ^{32}P-labeling of PCR-generated DNA fragments. The sequence of the primers used for PCR, along with their positions within the tobacco ptDNA (Shinozaki, et al. , 1986, supra) are as follows:

Gene	5' nucleotide position in plastid DNA	Sequence
accD	60221	GGATTTAGGGGCGAA
	60875	GTGATTTTCTCTCCG
atpB	56370(C)	AGATCTGCGCCCGCC
	55623	CCTCACCAACGATCC
atpI	15985(C)	GTTCCATCAATACTC
	15292	GCCGCGGCTAAAGTT
clpP	73621(C)	GACTTTATCGAGAAAG
	73340	GAGGGAATGCTAGACG
ndhA	122115(C)	GATATAGTGGAAGCG
	121602	GTGAAAGAAGTTGGG
ndhB	97792(C)	CAGTCGTTGCTTTTC
	97057	CTATCCTGAGCAATT
ndhF	113366(C)	CTCGGCTTCTTCCTC
	112749	CTCCGTTTTTACCCC
ORF1901	129496(C)	GTGACTATCAAGAGG
	128895	GACTAACATACGCCCG
ORF2280	92881	GCTCGGGAGTTCCTC
	93552	TGCTCCCGGTTGTTC
petB	78221	GGTTCGAAGAACGTC
	78842	GGCCCAGAAATACCT
psaA	43467(C)	TTCGTTCGCCGGAACC
	42743	GATCTCGATTCAAGAT
psbB	75241	GGAGCACATATTGTG
	75905	GGATTATTGCCGATG
psbE	66772(C)	CAATATCAGCAATGCAGTTCATCC
	66452	GGAATCCTTCCAGTAGTATCGGCC
rps14	38621	CACGAAGTATGTGTCCGGATAGTCC
rpl33/rpl18	70133	GGAAAGATGTCCGAG
	70636	GTTCACTAATAAATCGAC

The rps16 mRNA was probed with an EcoRI fragment isolated from plasmid pJS40, containing

sequences between nucleotides 4938/5363 and 6149/6656 of the tobacco ptDNA (Shinozaki et al., 1986, supra). The probe for tobacco 25S rRNA was from plasmid pKDR1 (Dempsey et al., Mol. Plant Path. **83**:1021, 1993) containing a 3.75 kb EcoRI fragment from a tobacco 25S/18S locus cloned in plasmid pBR325. When hybridizing gel-blots for 25S rRNA, ^{32}P-labeled double-stranded DNA probe was mixed with unlabeled plasmid pKDR1 corresponding to a 2-fold excess over the amount of RNA present on the filter.

Primer extension reactions Primer extension reactions were carried out on 10 μg (wild-type) or 10 μg (ΔrpoB) of total leaf RNA as described (Allison and Maliga, 1995 EMBO J. **15**:2802-2809). The primers are listed below. Underlined oligonucleotides were also used to generate the capping constructs.

Gene	5' nucleotide position in plastid DNA	Sequence
accD	59758	<u>CCGAGC</u>TCTTATTTCCTATCAGACTAAGC
atpB	56736	CCCCAGAACCAGAAGTAGTAGGATTGA
atpI	15973	GTATTGATGGAACATGATAGAACAT
clpP#1	74479	GGGACTTTTGGAACACCAATAGGCAT
clpP#2	74947	<u>GGGAGC</u>TCCATGGGTTTGCCTTGG
ORF1901	31451	CTTCATGCATAAGGATACTAGATTACC
ORF2280	87419	<u>GGGAGCT</u>CTACATGAAGAACATAAGCC
rps2	16921	CCAATATCTTCTTGTCATTTCTCTC
rps16	6185	CATCGTTTCAAACGAAGTTTTACCAT

Sequence ladders were generated with the same primers using the Sequenase II kit (USB).

Identification of primary transcripts by in vitro capping Total leaf RNA (20 mg) from wild-type and ΔrpoB plants was capped in the presence of [a-32P]GTP (Kennell and

Pring, 1989 Mol Gen. Genet., **216**:16-24). Labeled RNAs were detected by ribonuclease protection (Vera and Sugiura, 1992, supra) using the RPAII kit (Ambion). To prepare the protecting complementary RNA, the 16SrDNA upstream region (nucleotides 102526-102761 of the ptDNA) was PCR-amplified using the primers listed below. The 5' primers were designed to add an XbaI restriction site(underlined) upstream of the amplified fragment. The 3' primers were designed to add a KpnI site (underlined) downstream of the amplified sequence. The amplified product was cloned as an XbaI to KpnI fragment into XbaI- and KpnI-restricted pBSKS+ vector (Stratagene). To generate unlabeled RNA complementary to the 5' end of RNAs, the resulting plasmid was linearized with XbaI, and transcribed in a Megascript (Ambion) reaction with T3 RNA polymerase. Markers (100, 200, 300, 400, and 500 nucleotides) were prepared with the RNA Century Markers Template Set (Ambion), following the manufacturer's protocol.

Gene	5' nucleotide position in plastid DNA	Sequence
accD	59758	CCGAGCTCTTATTTCCTATCAGACTAAGC
	59576	CCGGTACCATAGGAGAAGCCGCCC
atpB	56750	CCGAGCTCGTAGTAGGATTGATTCTCA
	57131(C)	CCGGTACCGGAGCCAATTAGATACAAA
atpI	15895	CCGAGCTCTGACTTGGAAACCCCC
	16277(C)	CCGAATTCTAGTATTCGCAATTTGT
clpP	74462	GGGAGCTCCAGGACTTCGGAAAGG
	74752(C)	GGGGTACCAATACGCAATGGGG
	74947	GGGAGCTCCATGGGTTTGCCTTGG
	75080(C)	GGGGTACCGCTAATTCATACAGAG
ORF1901	31424	GGGAGCTCCGACCACAACGACCG
	31724(C)	GGGGTACCCTTACATGCCTCATTTC
ORF2280	87419	GGGAGCTCTACATGAAGAACATAAGCC
	87154	GGGGTACCGTGCCTAAGGGCATATCGG

DNA sequence analysis DNA sequence analysis was carried out utilizing the Wisconsin Sequence Analysis Package (Genetics Computer Group, Inc.).

Results and Discussion

Based on the accumulation of mRNAs in wild-type and ΔrpoB leaves, the plastid genes may be divided into three classes. The first class includes genes for which the mRNAs accumulate to high levels in wild-type leaves, and to very low levels in the leaves of ΔrpoB plants (Figure 6A). Genes which belong to this class are psaA (photosystem I gene), psbB and psbE (photosystem II genes), petB (cytochrome b6/f complex gene), ndhA (respiratory chain NADH dehydrogenase homologue; Matsubayashi et al., 1987 Mol. Gen. Genet. **210**:385:393) and rps14 (ribosomal protein gene). The second class includes plastid genes for which the mRNAs accumulate to about equal levels in the wild-type and (rpoB leaves (Figure 6B). This class includes atpB (ATP synthase gene), ndhF (respiratory chain NADH dehydrogenase homologue gene; Matsubayashi et al., 1987, supra), rps16 (ribosomal protein gene) and ORF1901 (a gene with unknown function; Wolfe et al., 1992, J. Mol. Biol. **223**:95-104). The third class includes genes for which there is significantly more mRNA in the ΔrpoB leaves than in the leaves of wild-type plants (Figure 6C). Typical for this class are rpl33 and rpl18 (ribosomal protein genes), accD (encoding a subunit of the acetyl-CoA carboxylase; Sasaki et al., 1993 Plant Physiol. **108**:445-449) and ORF2280 (putative ATPase with unknown function; Wolfe 1994, Curr. Genet. **25**:379-383). Two additional genes of this class, ndhB (respiratory chain NADH dehydrogenase homologue; Matsubayashi et al., 1987, supra) and clpP (encoding the proteolytic subunit of the Clp ATP-dependent protease; Maurizi et al., 1990 J. Biol. Chem. **265**:12546-12552; Gray et al., 1990 Plant Mol. Biol.

15:947-950) form a subgroup of this class which demonstrate significant levels of mRNA in wild-type leaves.

The atpB and atpI ATP synthase genes have both NEP and PEP promoters.

The RNA gel blot analysis identified a number of genes and operons for which high transcript levels are maintained in ΔrpoB leaves. To identify additional NEP promoters, the 5'-end of several transcripts has been mapped by primer extension analysis. 5' ends may be those of primary transcripts identifying a promoter, or generated by RNA processing. Since primary plastid transcripts retain triphosphate groups at their 5' ends, specific [^{32}P]GMP transfer to these RNA molecules by the enzyme guanylyltransferase allowed accurate discrimination between primary transcripts and processed ends. For the tobacco atpB operon, transcript 5'- ends have been identified by Orozco et al. (1990 Curr. Genet. 17:65-71.) at nucleotide positions -611, -502, -488, -289 and -255 upstream of the translation initiation codon (Figure 7C). The 5' ends are numbered relative to the translation initiation codon (ATG) when the nucleotide directly upstream of A is at position -1. Primer extension analysis identified each of these 5'-ends in our wild-type plants (Figure 7A). In the ΔrpoB sample only the -289 RNA species was present, the 5' end of which was a substrate for guanylyltransferase (Figure 7B). Therefore, the -289 RNA is transcribed from a NEP promoter, PatpB-289. Interestingly, the -289 transcript is present in the wild-type leaves, although it is less abundant than in the ΔrpoB plants. The -255, -488 and -611 transcripts are absent in the ΔrpoB plants (Figure 7A). DNA fragments containing these promoters (but not PatpB-289) are recognized by the *E. coli* RNA polymerase (Orozco et al., 1990, supra), and are transcribed by PEP

in plastids. The atpA operon includes the atpI, -atpH-atpF-atpA genes (Figure 8C). In wild-type tobacco leaves, mRNA 5' ends have been mapped to three regions upstream of atpI,: the -209 region, with 5'ends mapping to nucleotides -212, -209 and -207, and 5'-ends at nucleotides -130 and -85. In ΔrpoB leaves only the -207 transcript is detectable (Figure 8A). This transcript could be capped in the ΔrpoB RNA sample (Figure 8B), therefore it is transcribed from a NEP promoter. A signal at this position was also obtained in the in vitro capping reaction of wild-type RNA samples. The -209 and -212 transcripts may be due to the activity of an overlapping PEP promoter, or formation of multiple transcripts from the NEP promoter in wild-type plants. The -130 transcript which is present only in wild-type leaf RNA could also be capped (Figure 8A, 8B). Since there are sequences similar to -10/-35 elements at the correct spacing upstream of this 5'-end, it is transcribed by the PEP polymerase.

A clpP NEP promoter is highly expressed in chloroplasts.

The clpP protease subunit gene also belongs to the class which has both NEP and PEP promoters. Primer extension analysis in the wild-type plants identified RNA 5'-ends at nucleotide positions -53, -95, and -173, while in ΔrpoB plants 5' ends map to the -53, -173, and -511 nucleotide positions (Figure 9A). In vitro capping reaction verified that each of these are primary transcripts (Figure 9B). Three of the transcripts derive from NEP promoters. The PclpP-53 promoter is highly expressed in both wild-type-and ΔrpoB plants, thus represents a distinct class of NEP promoters with a potential for high-level expression in different tissue types. The PclpP-53 promoter is well conserved in spinach (Westhoff, 1985 Mol. Gen. Genet. **201**:115-123). Additional clpP promoters for NEP are PclpP-173 and

PclpP-511. Since the PclpP-511 transcript accumulates only in ΔrpoB plants (Figure 9A) it is a candidate regulated NEP promoter. Note also, that the PclpP-511 is located within the psbB coding region, and it's expression may be affected by the convergent psbB PEP promoter (Figure 9C).

The only PEP promoter directly upstream of clpP is PclpP-95. RNAs from this promoter accumulate only in wild-type leaves and PclpP-95 has upstream sequences reminiscent of the -10-/-35 conserved elements (not shown).

The accD gene is transcribed exclusively from a NEP promoter.

For the lipid biosynthetic gene accD, mRNA accumulates to high levels only in ΔrpoB plants. A major transcript initiates at nucleotide position -129 (Figure 10A), which can be capped in vitro (Figure 10B). Therefore, this RNA is transcribed from a NEP promoter. Since PaccD-129 does not have a significant activity in the photosynthetically active leaf mesophyll cells, it serves as a candidate for a regulated NEP promoter with a distinct tissue-specific expression pattern.

NEP promoters share a loose consensus adjacent to the transcription initiation site.

Sequences flanking the transcription initiation sites were aligned to identify conserved NEP promoter elements (Figure 11). Included in the sequence alignment are nine promoters identified in this study and Prrn-62, the NEP promoter described in Allison et al. (1996, supra). Sequences for PORF2280-1577, and PORF1901-41 for which the 5' ends were shown to be primary transcripts by capping in vitro are also included (data not shown). Both of these promoters are active in ΔrpoB leaves but not in the leaves of wild-type plants. Also included in the sequence alignment are tentative NEP promoters for

rps2 and rps16 , for which there is more mRNA in ΔrpoB leaves. The 5' ends of these transcripts were mapped by primer extension analysis. The in vitro capping assay failed due to low abundance of the mRNAs (data not shown). Multiple sequence alignment of the regions immediately flanking the NEP 5' ends identified a loose 10 nucleotide consensus around the transcription initiation site (Figure 11). Conservation of additional nucleotides upstream and downstream is also apparent. Striking is the lack of sequence conservation between the PclpP-53 and other NEP promoters which is the only NEP promoter highly active in chloroplasts. Given the lack of sequence similarity, this sequence was not included in the alignment. Sequences around the PclpP-53 transcription initiation site are shown separately at the bottom of Figure 11.

NEP and PEP polymerases, through recognition of distinct promoters, provide a mechanism for selective transcription of plastid genes (Figure 12). The data provided herein demonstrate that some genes have only PEP promoters or NEP promoters while others have both PEP and NEP regulatory sequences.

EXAMPLE III

NEP Promoters for the Expression of Selectable Marker Genes

For versatility and universal applications, expression of selectable marker genes for plastid transformation is desirable in all tissue types at a high level. Selectable marker genes in the currently utilized plastid transformation vectors are expressed from PEP promoters recognized by the plastid encoded RNA polymerase. The PEP polymerase transcribes photosynthetic genes and some of the housekeeping genes, therefore appears to be the dominant RNA polymerase in photosynthetically active leaf tissues. Efficient plastid transformation has been achieved in tobacco based on chloroplasts transformation in leaf cells. However,

plant regeneration is not feasible, or is not practical from the leaves of most agronomically important cereal crops, including maize, rice, wheat and in cotton. In these crops, transgenic plants are typically obtained by transforming embryogenic tissue culture cells or seedling tissue. Given that these tissues are non-photosynthetic, expression of marker genes by NEP promoters which are active in non-green tissues appears to be particularly advantageous, and will facilitate transformation of plastids in all non-photosynthetic tissue types.

A particularly suitable promoter to drive the expression of marker genes is the PclpP-53 promoter. This promoter is highly expressed in the proplastids of ΔrpoB plants, therefore it may also be highly expressed in the proplastids of embryogenic cell cultures which yield transgenic cereal plants. Marker genes expressed from these promoters will also be suitable to select plastid transformants in bombarded leaf cultures, since this promoter was found to be active in chloroplasts. Marker genes expressed from promoters such as the PclpP-53 promoter will have wide application to obtain transformed plastids.

The selectable marker genes will be constructed using the principles outlined in U.S. Patent No. 5,451,513 and U.S. Patent No. 5,877,402. A transforming DNA construct is illustrated in Figure 13. More specifically, the PclpP-53 promoter will be cloned upstream of a DNA segment encoding a plastid-selectable marker. Signals for translation will be provided by incorporating suitable DNA sequences between the promoter fragment and the selectable marker coding region. 3' untranslated segments of a plastid gene to provide signals for transcription termination and to stabilize the chimeric mRNA will be cloned downstream of the selectable marker. Utilization of the 3'untranslated region of genes

expressed from NEP promoters is preferred since the requirements for transcription termination for the NEP and PEP polymerases may be different.

PclpP-53 is a particularly strong NEP promoter. However, plants with transformed plastids may be obtained with weak promoters as well. There are several examples for such weak NEP promoters in the preceding examples, for example PclpP-173.

Expression of Tissue Specific Plastid Transgenes Driven by NEP Promoters

Tissue specific expression of plastid transgenes is desirable in many applications. Tissue specific expression of a protein that makes the plant tissue repellent or toxic for root nematodes may be desirable in roots. However, expression of the same protein in the leaves would drain the plants resources and may effect utilization of the aerial plant parts. Since most often expressed in non-green tissues, the NEP promoters described in this application, and the promoters expressed from the NEP polymerase in general, are a rich source of tissue-specific promoters for transgene expression.

Several of the NEP promoters, for example PclpP-511, are highly expressed in proplastids of ΔrpoB plants. Proplastids are present in the edible part of cauliflower. Therefore, high level expression of foreign genes in cauliflower is anticipated from this promoter in the edible parts of the plant.

The plastid gene accD encodes a subunit of the prokaryotic acetyl-CoA carboxylase, an enzyme involved in lipid biosynthesis. Interestingly, in wild-type leaves the level of accD mRNA is low while it is high in the proplastids of ΔrpoB plants. This observation suggests that PaccD-129 is active in non-green plastids of tissues

actively involved in lipid biosynthesis, such as the plastids of developing seed which is rich in oil.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Rutgers, The State University of New Jersey

(ii) TITLE OF THE INVENTION: NUCLEAR ENCODED TRANSCRIPTION SYSTEM IN THE PLASTIDS OF HIGHER PLANTS

(iii) NUMBER OF SEQUENCES: 74

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Dann, Dorfman, Herrell and Skillman
(B) STREET: 1601 Market Street Suite 720
(C) CITY: Philadelphia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19103-2307

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US96/12671
(B) FILING DATE: 01-AUG-1996
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/002,136
(B) FILING DATE: 10-AUG-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Rigaut, Kathleen D.
(B) REGISTRATION NUMBER: P43,047
(C) REFERENCE/DOCKET NUMBER: 95-08031EPO

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 215-563-4100
(B) TELEFAX: 215-563-4044
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGCTTCTGTA ACTGG 15

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TGACTGTCAA CTACAG 16

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTTCCATCAA TACTC 15

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GCCGCGGCTA AAGTT 15

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCCCACGTTC AAGGT 15

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TGAGTTCGTA TAGGC 15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TTCATAGTTG CATTACTTAT AGCTTC 26

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACTTGCTTTA GTCTCTGTTT GTGGTGACAT 30

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCTCTAGACC CTAAGCCCAA TGTG 24

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCGGTACCGA GATTCATAGT TGCATTAC 28

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGATTTAGGG GCGAA 15

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GTGATTTTCT CTCCG 15

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGATCTGCGC CCGCC 15

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCTCACCAAC GATCC 15

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTTCCATCAA TACTC 15

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GCCGCGGCTA AAGTT 15

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACTTTATCG AGAAAG 16

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAGGGAATGC TAGACG 16

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GATATAGTGG AAGCG 15

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTGAAAGAAG TTGGG 15

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CAGTCGTTGC TTTTC 15

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTATCCTGAG CAATT 15

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CTCGGCTTCT TCCTC 15

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CTCCGTTTTT ACCCC 15

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GTGACTATCA AGAGG 15

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GACTAACATA CGCCCG 16

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCTCGGGAGT TCCTC 15

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TGCTCCCGGT TGTTC 15

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGTTCGAAGA ACGTC 15

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GGCCCAGAAA TACCT 15

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TTCGTTCGCC GGAACC 16

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GATCTCGATT CAAGAT 16

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGAGCACATA TTGTG 15

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GGATTATTGC CGATG 15

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CAATATCAGC AATGCAGTTC ATCC 24

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGAATCCTTC CAGTAGTATC GGCC 24

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CACGAAGTAT GTGTCCGGAT AGTCC 25

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGAAAGATGT CCGAG 15

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GTTCACTAAT AAATCGAC 18

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CCGAGCTCTT ATTTCCTATC AGACTAAGC 29

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CCCCAGAACC AGAAGTAGTA GGATTGA 27

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GTATTGATGG AACATGATAG AACAT 25

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGGACTTTTG GAACACCAAT AGGCAT 26

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GGGAGCTCCA TGGGTTTGCC TTGG 24

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CTTCATGCAT AAGGATACTA GATTACC 27

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GGGAGCTCTA CATGAAGAAC ATAAGCC 27

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CCAATATCTT CTTGTCATTT CTCTC 25

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CATCGTTTCA AACGAAGTTT TACCAT 26

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CCGAGCTCTT ATTTCCTATC AGACTAAGC 29

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CCGGTACCAT AGGAGAAGCC GCCC 24

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CCGAGCTCGT AGTAGGATTG ATTCTCA 27

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CCGGTACCGG AGCCAATTAG ATACAAA 27

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CCGAGCTCTG ACTTGGAAAC CCCC 24

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

CCGAATTCTA GTATTCGCAA TTTGT 25

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GGGAGCTCCA GGACTTCGGA AAGG 24

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGGGTACCAA TACGCAATGG GG 22

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GGGAGCTCCA TGGGTTTGCC TTGG 24

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GGGGTACCGC TAATTCATAC AGAG 24

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GGGAGCTCCG ACCACAACGA CCG 23

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GGGGTACCCT TACATGCCTC ATTTC 25

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GGGAGCTCTA CATGAAGAAC ATAAGCC 27

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GGGGTACCGT GCCTAAGGGC ATATCGG 27

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

ATTTGCTCCC CCGCCGTCGT TCAATGAGAA TGGATAAGAG GCTCGTGGGA TTGACGTGAG 60
GGGGCAGGGA TGGCTATATT TCTGGGAGCG AACTCCGGGC GAATATGAAG CGCATGGATA 120
CAAGTTATGC CTTGGAATGA AAGACAATTC CGAATCCGCT TTGTCTACGA ACAAGGAAGC 180
TATAAGTAAT GCAACTATGA A 201

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GAGCGAACTC CGGGCGAATA TGAAGCGCAT GGATACAAGT TATGCCTTGG AATGAAAGAC 60
AATTCCGAA 69

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CTGGATGTAG ATGATGATAT CTATACAGAT GGATCTTATA TATATCGTAG AATGAAGTAC 60
CACATGGGT 69

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

AATGAATCGT AATAGAAATA GAAAATAAAG TTCAGGTTCG AATTCCATAG AATAGATAAT 60

ATGGATGGG 69

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GCTATCTGCA TTTGGTATGG TTATTTGCTT TGGTAATAAA AAGAATAATG AATAGAAAAG 60
AATAATGGT 69

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TAGCGATGGG GTCTTACTAA AGAAAAATAT TTATCCACCT ATCTCTATAG TATATAGATA 60
TAGAACAA 68

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

ATCTCTTTCT ATTTCTTTTC TATATATGGA AAGTTCAAAA ATCATCATAT AATAATCCAG 60
AAATTGCAA 69

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

TTGTGAGGGT CGTTCTAGCT ATATAAGAAA TCCTTGATTA ATAATAACAT AATAAGATAA 60
ATAACTTAC 69

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

ATACAGAGCC ATCGAACCGG CCCAACCAGC AACCAGAGCT GTATGCATTA TATGAACAGA 60
AAGCAACCG 69

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TGAGGGGTA ACGCATAACC ATGAATAGCT CCATACTTAT TTATCATTAG AAAGACCTAT 60
TCGTAATAA 69

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

AATTATATTA TTTTAAATAA TATAAAGGGG GTTCCAACAT ATTAATATAT AGTGAAGTGT 60
TCCCCCAGA 69

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 0 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

AGACAATAAAAAAAATTGTTACGTTTCCACCTCAAA

CLAIMS:

1. A DNA construct for stably transforming plastids of multicellular plants, which comprises a transforming DNA having a targeting segment which effects insertion of said transforming DNA into said plastid genome by homologous recombination, a selectable marker gene conferring a selectable phenotype to plant cells containing said transformed plastids, a cloning site for insertion of an additional expressible DNA encoding a foreign gene of interest, and a 5' promoter element which is recognized and transcribed by a nuclear encoded plastid (NEP) RNA polymerase.

2. The DNA construct of claim 1, wherein said construct is incorporated into a vector suitable for transformation of plastids.

3. A DNA construct according to claim 1, wherein said DNA construct further comprises a second 5' promoter element which is recognized and transcribed by a plastid encoded plastid RNA polymerase.

4. The DNA construct according to claim 1 wherein said promoter element is selected from promoter elements of plastid genes selected from the group consisting of Prrn-62, PORF2280-1577, PatpB-289, PORF1901-41, Pr[b]ps2-152, Prps16-107, PatpI-207, PclpP-511, PclpP-173, PclpP-53 and PaccD-129.

5. A DNA construct for stably transforming plastids of a plant cell and for expression of at least one additional gene product therein, comprising:

a) a targeting segment comprising a DNA sequence substantially homologous to a predetermined plastid genomic sequence with a plastid to be transformed, said targeting segment enabling homologous recombination with said pre-determined plastid genomic sequence; and

b) a first selectable marker gene disposed within said targeting segment, said first selectable marker gene conferring a non-lethal selectable phenotype to cells containing plastids with said DNA construct; and

c) an additional DNA segment comprising a transcription unit of a gene encoding a protein or a precursor thereof; and

d) a promoter sequence which is operably linked to said transcription unit, said promoter sequence being recognized by a nuclear encoded polymerase, said gene encoding said protein being regulated by said promoter.

6. A DNA construct according to claim 5, wherein said transcription unit encodes a second selectable marker gene.

7. A DNA construct according to claim 6, wherein said selectable marker gene is regulated by a promoter transcribed by a nuclear encoded plastid polymerase.

8. A DNA construct according to claim 5, wherein said transcription unit encodes a reporter gene.

9. A DNA construct according to claim 5, wherein said construct is incorporated into a vector suitable for transformation of plastids.

10. A plant cell stably transformed with the DNA construct of claim 1.

11. A plant cell stably transformed with the DNA construct of claim 2.

12. A method for obtaining a plant cell or a multicellular plant, the plastids of which cell have been stably transformed by at least one foreign gene of interest, which comprises administering to a plant cell a DNA construct comprising:

a) a targeting segment comprising a DNA sequence substantially homologous to a predetermined plastid genomic sequence with a plastid to be transformed, said targeting segment enabling homologous recombination with said pre-determined plastid genomic sequence; and

b) a selectable marker gene disposed within said targeting segment, said selectable marker gene conferring a selectable plastid phenotype to cells containing plastids with said DNA construct; and

c) and a foreign gene of interest, said gene being regulated by a promoter recognized by a nuclear encoded plastid RNA polymerase; and

d) selecting for cells which express said phenotype; and

e) regenerating a plant from said cell containing stably transformed plastids.

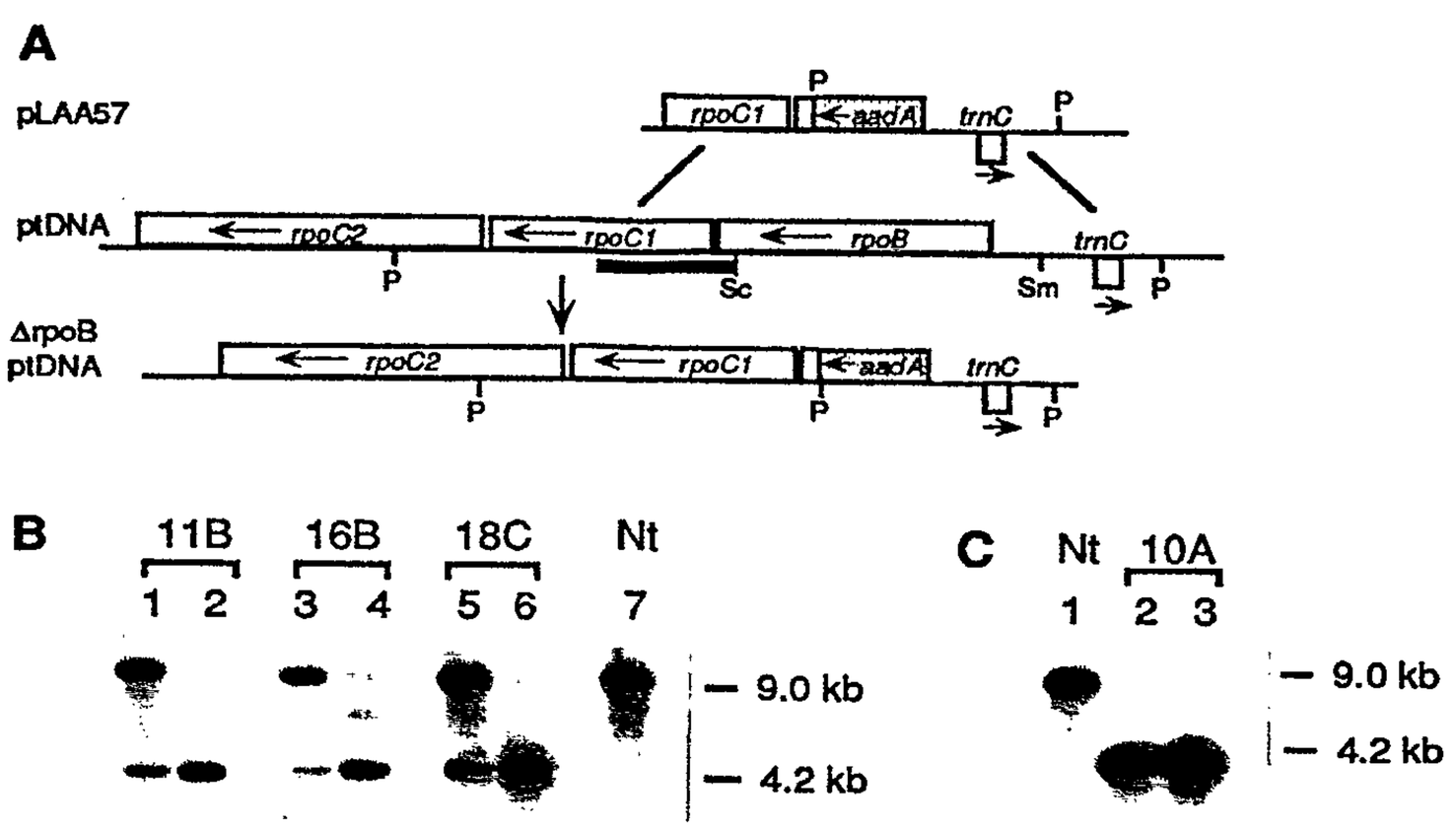
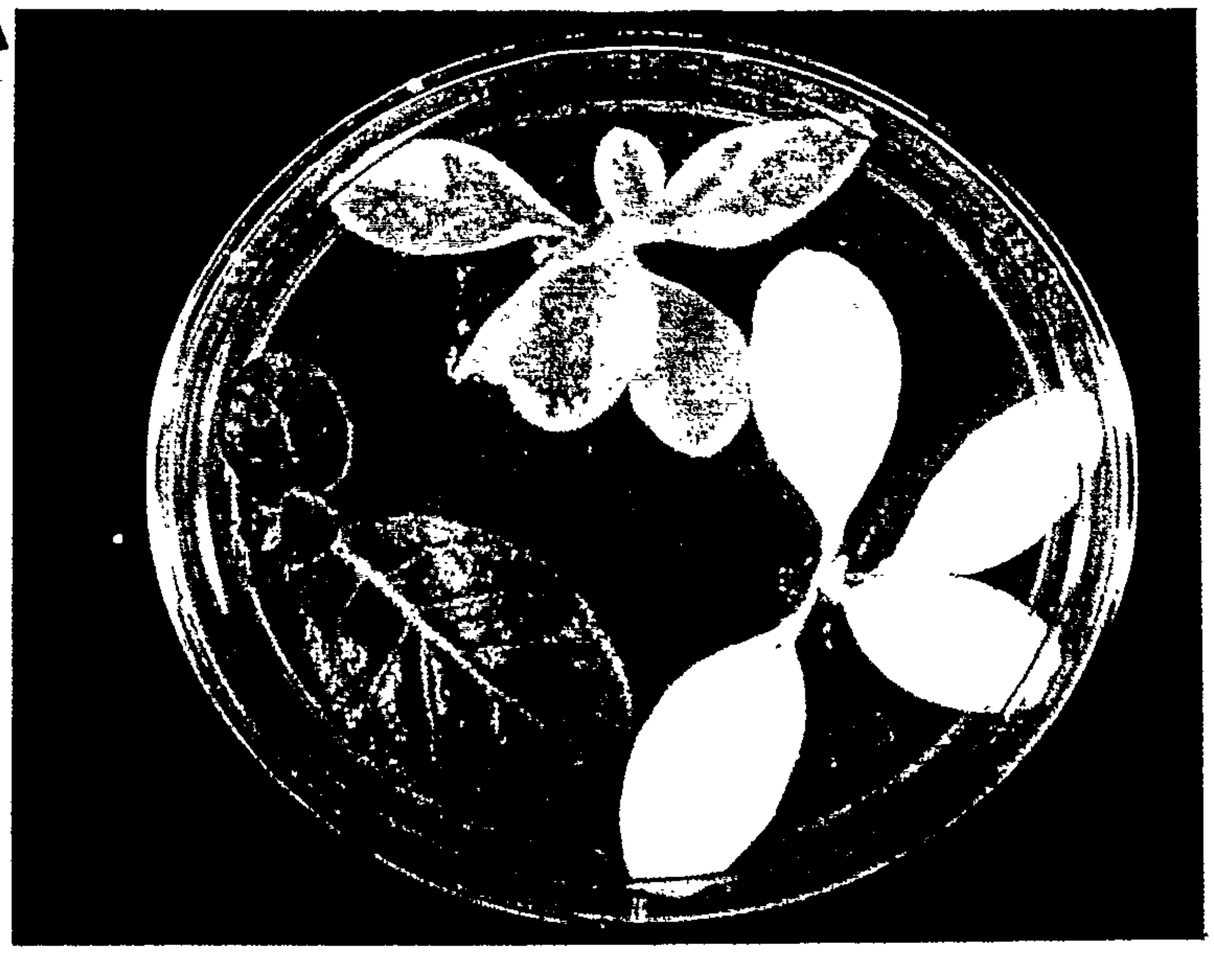

A
pLAA57
rpoC1
aadA
trnC
P
ptDNA
rpoC2
rpoC1
rpoB
trnC
Sc
Sm
ΔrpoB
ptDNA
B
11B
16B
18C
Nt
1 2 3 4 5 6 7
9.0 kb
4.2 kb
C
Nt
10A
1 2 3
9.0 kb
4.2 kb

Figure 1

A



B

Figure 2

Fig. 3A

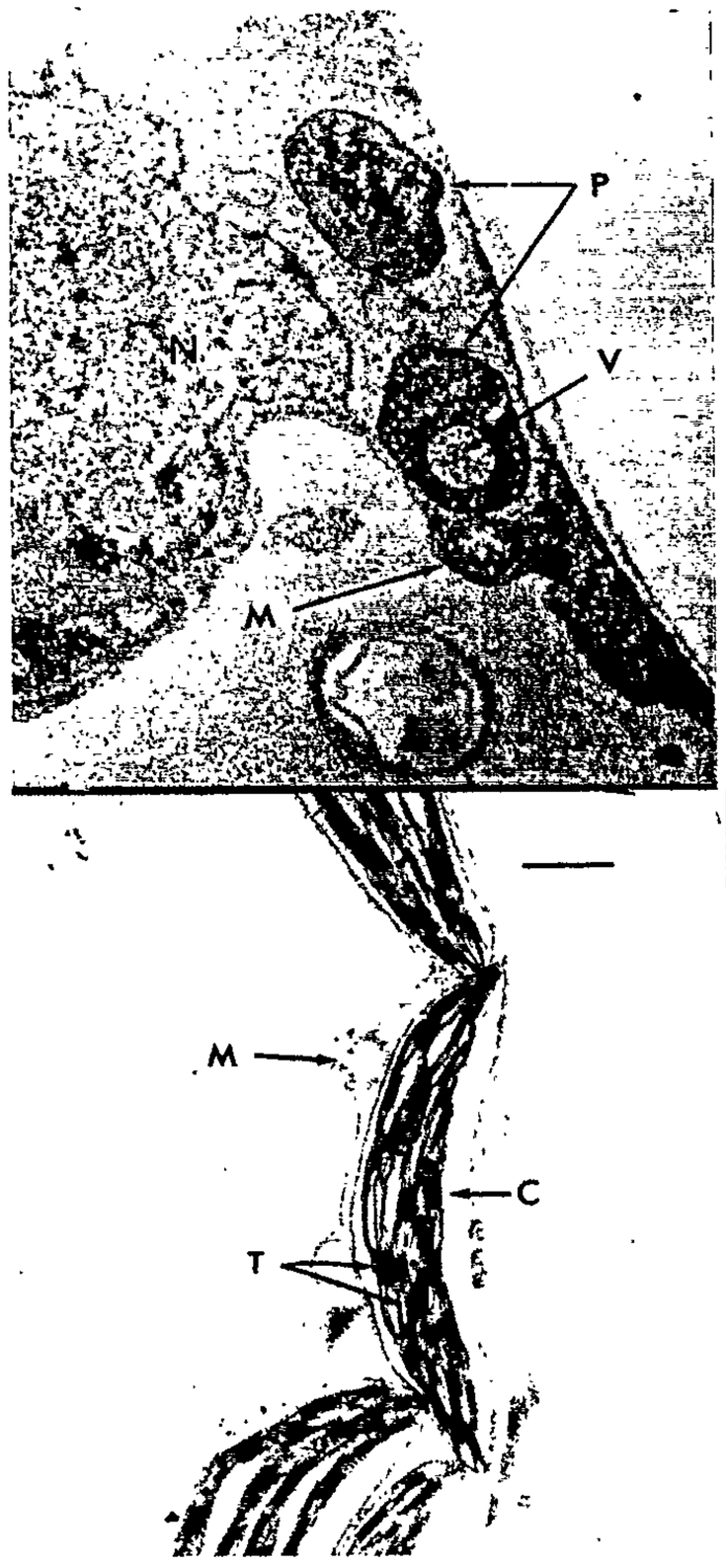


Fig. 3B

Figure 3

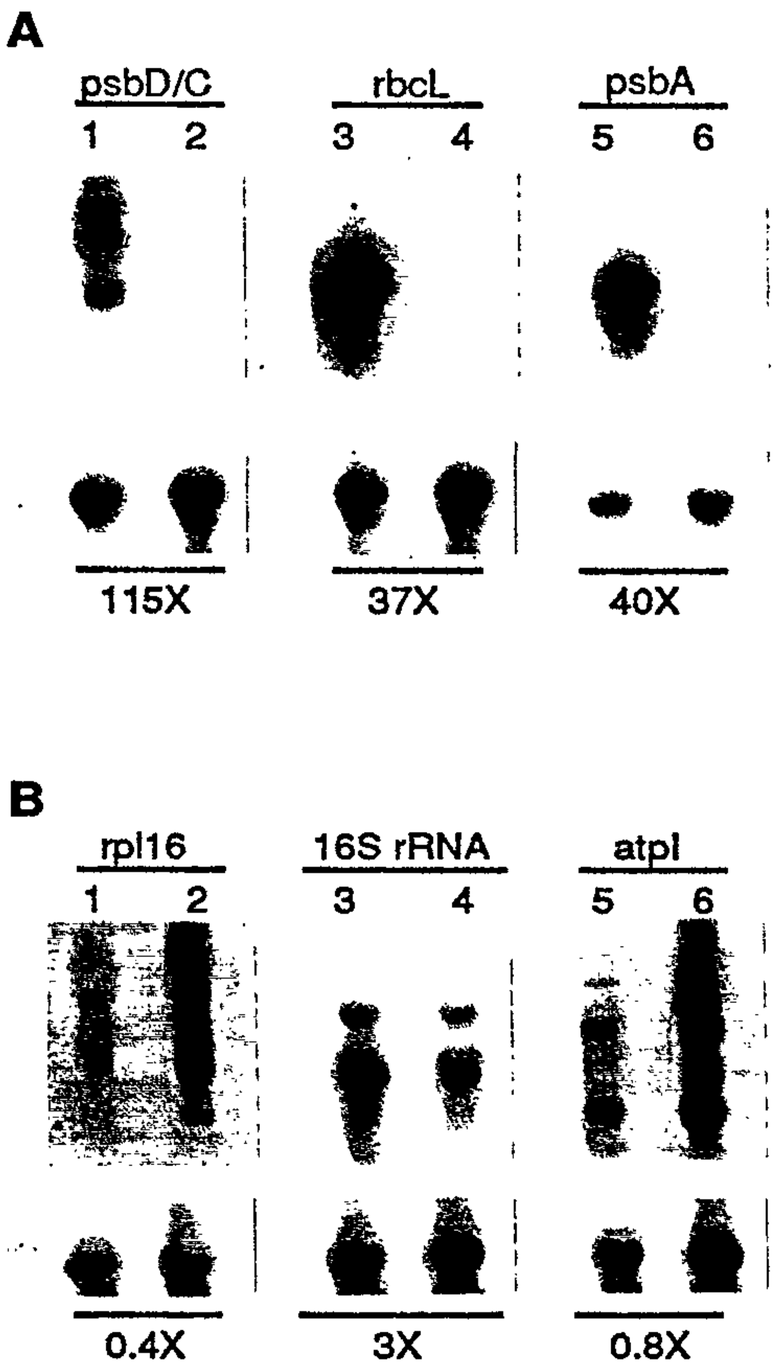
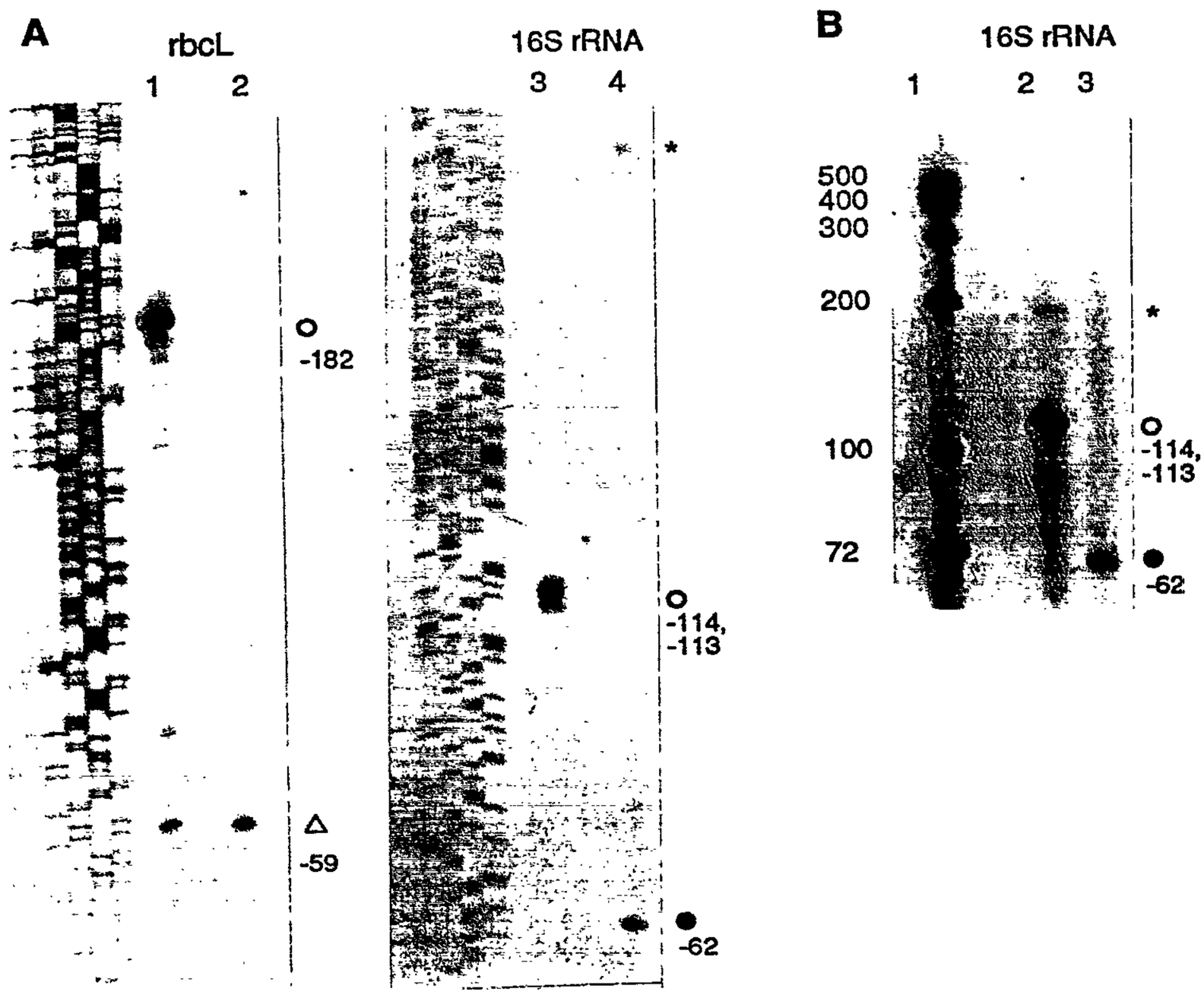

A
psbD/C
rbcL
psbA
1
2
3
4
5
6
115X
37X
40X
B
rpl16
16S rRNA
atpI
1
2
3
4
5
6
0.4X
3X
0.8X

Figure 4



C

-201 ATTTGCTCCCCCGCCGTCGTTCAATGAGAATGGATAAGAGGCTCGTGGGA

-114,-113 (P1)

-151 TTGACGTGAGGGGGCAGGGATGGCTATATTTCTGGGAGCGAACTCCGGGC

-62 (P2)

-101 GAATATGAAGCGCATGGATACAAGTTATGCCTTGGAATGAAAGACAATTC

-51 CGAATCCGCTTTGTCTACGAACAAGGAAGCTATAAGTAATGCAACTATGAA

Figure 5

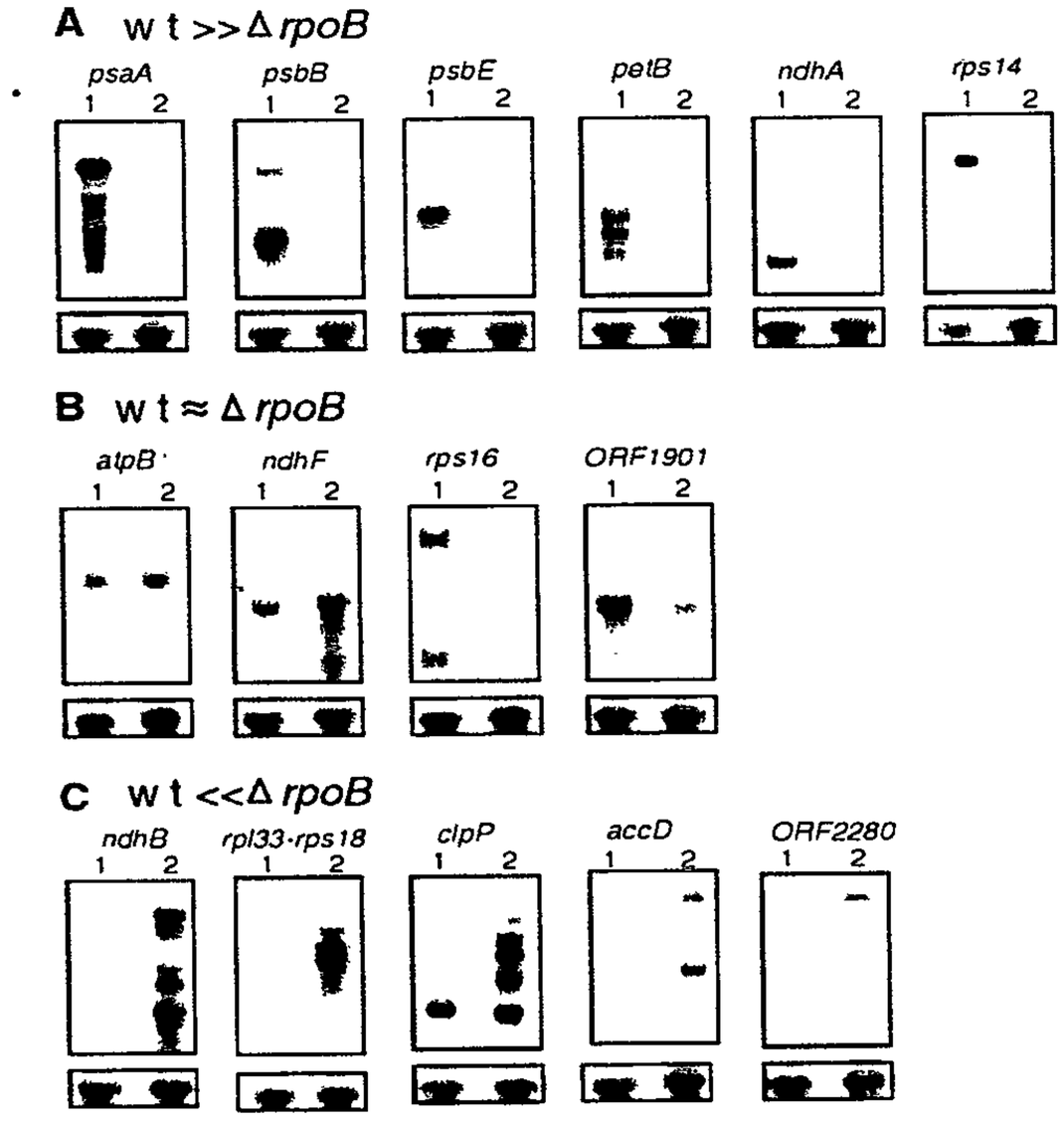

A wt >> Δ rpoB
psaA
psbB
psbE
petB
ndhA
rps14
1 2
B wt ≈ Δ rpoB
atpB
ndhF
rps16
ORF1901
C wt << Δ rpoB
ndhB
rpl33-rps18
clpP
accD
ORF2280

Figure 6

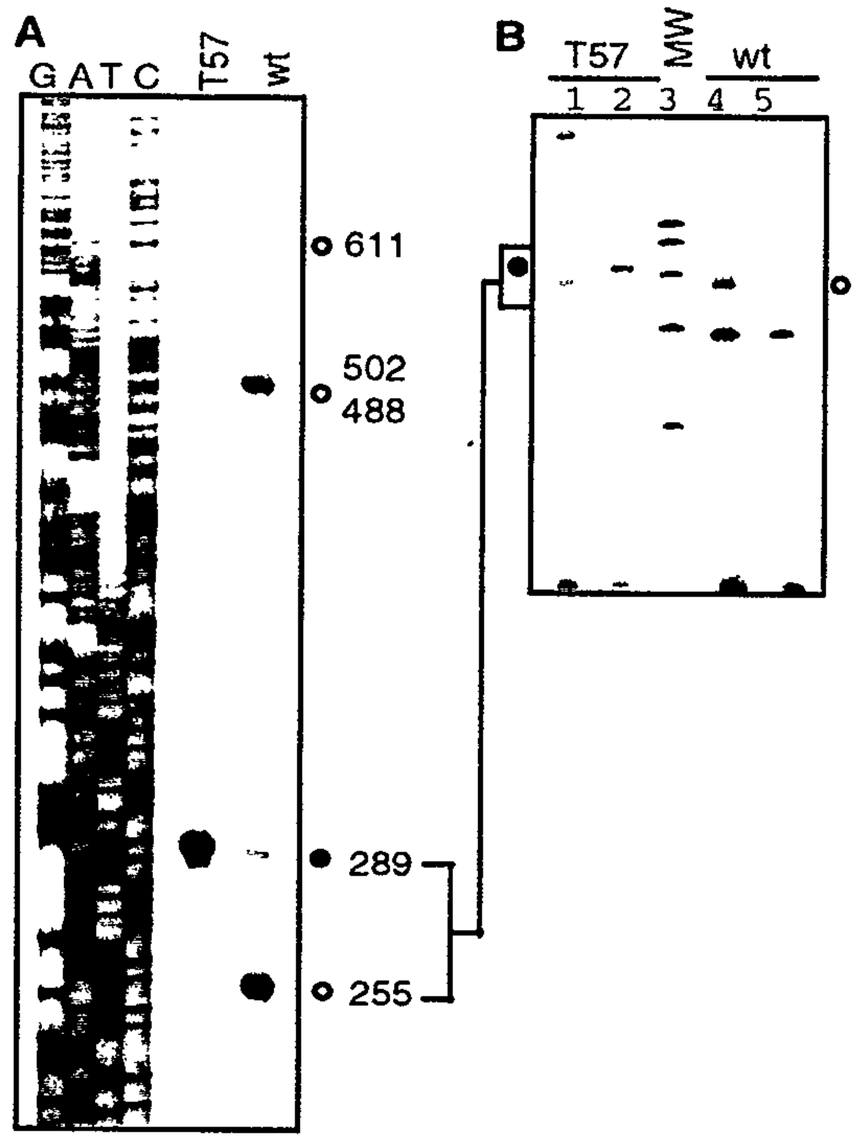

A
G A T C
T57
wt
611
502
488
289
255
B
T57
MW
wt
1 2 3 4 5
C
rbcL
atpB
atpE
rbcL
-611
-502
-488
-289
-255
+1
atpB

Figure 7

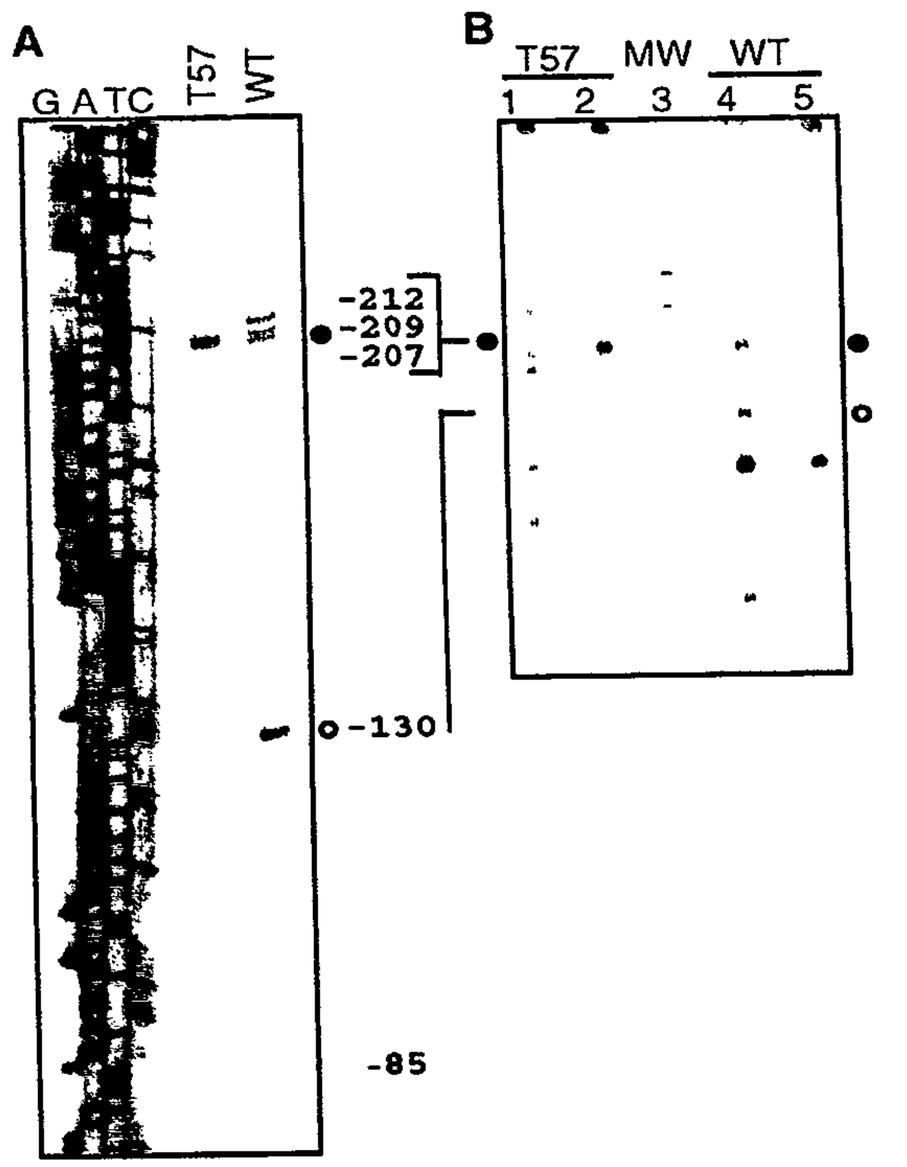

A
G A T C
T57
WT
-212
-209
-207
-130
-85
B
T57
MW
WT
1 2 3 4 5
C
rpoC2
rps2
atpI
atpH
atpF
atpA
trnR
rps2
-212
-209
-207
-130
-85
+1
atpI

Figure 8

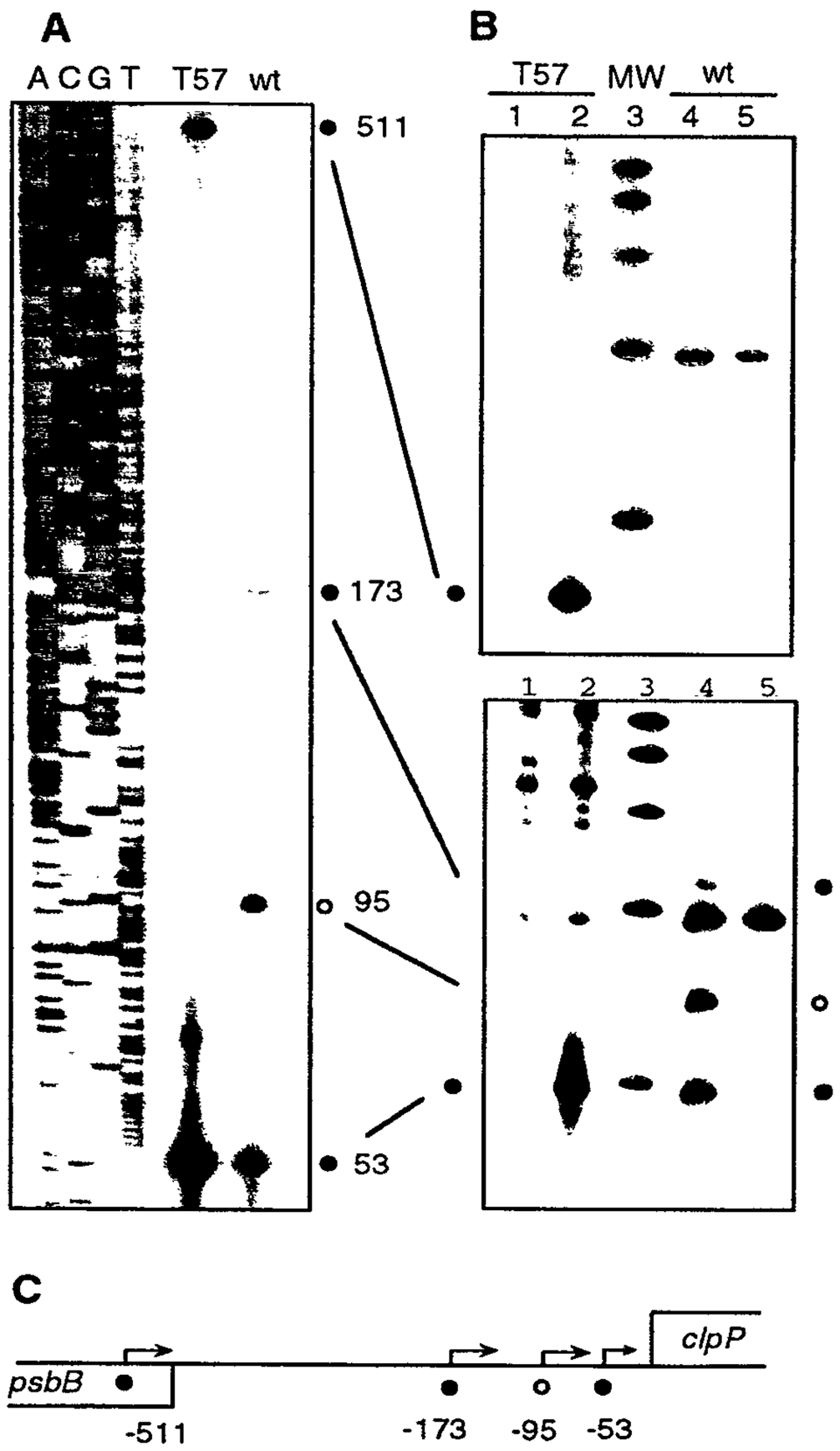

A
ACGT T57 wt
511
173
95
53
B
T57 MW wt
1 2 3 4 5
1 2 3 4 5
C
psbB
clpP
-511
-173
-95
-53

Figure 9

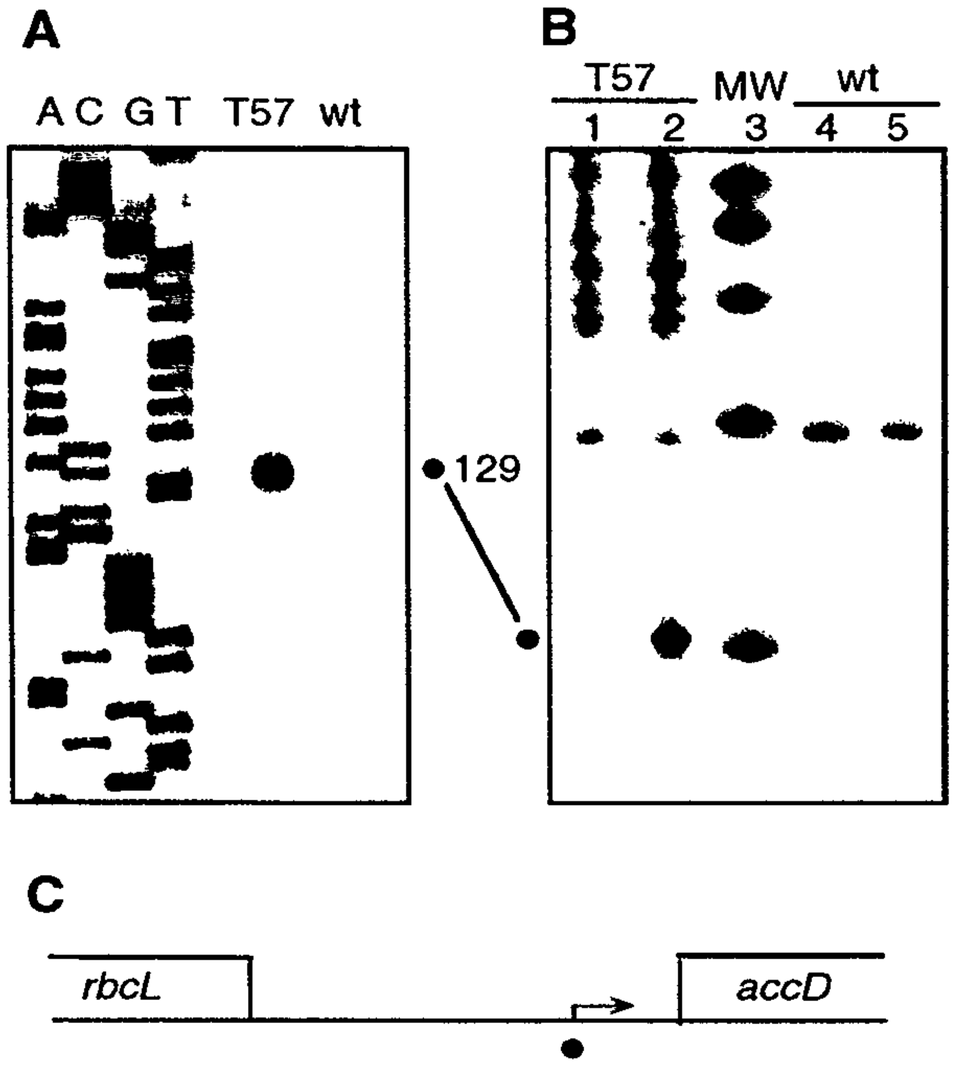


Figure 10

```
Prrn-62          GAGCGAACTCCGGGCGAATATGAAGCGCATGGA
PORF2280-1577    CTGGATGTAGATGATGAATCTATCAGATGGA
PatpB-289        AATGAATCGTAATAGAAATAGAAATAAAGTTC
Prps2-152        GCTATCTGCATTTGGTAATGGTTATTTGGG
Prps16-107       TAGCGATGGGGTCTTACTAAAGAAAATTTA
PORF1901-41      ATCTCTTTCTATTCTTTCTTTCTATATATGGAAAG
Patpl-207        TTGTGAGGGTCGTTCTAGCTATATAGAAATCC
PclpP-511        ATACAGAGCCATCGAACCGGCCCAACCAGCAAC
PclpP-173        TGAGGGGGTAACGCATAACCCATGAATAGCTCCA
PaccD-129        AATTATATATTATTAAATAATATAAGGGGGTT

Prrn-62          TACAAGTTATGCCTTGGGAATGAAAGACATTCCGAA
PORF2280-1577    TCTTATATATATCGTAGAATGAAGTACCACATGGGT
PatpB-289        AGGTTCGAATTCCATAGAATAGAGTAATGGATGGG
Prps2-152        TAATAAAAAGAATAATGAAATAGAAAGAATAATGGT
Prps16-107       TCCACCTATCTTACTATGTATTA-GAATAGAACAA
PORF1901-41      TTCAAAATCATCATAATTAAATCCAAGAAATTGCAA
Patpl-207        TTGATTAATAATAACATAATTAAGATAAAACTTAC
PclpP-511        CAGAGCTGTATGCATTATGAACAGAAAGCAACCG
PclpP-173        TACTTATTTATCATTAGAAAGACCTATTCGTAATAA
PaccD-129        CCAACATATTAATATAGTGAAGTGTTCCCCCAGA

Consensus                    ATAGAATAAA
                                  -
                                  G

PclpP-53         AGACAATAAAAAAAATTGTTACGTTTCCACCTCAAA
```

FIG. 11

CA 02225652 2004-03-29

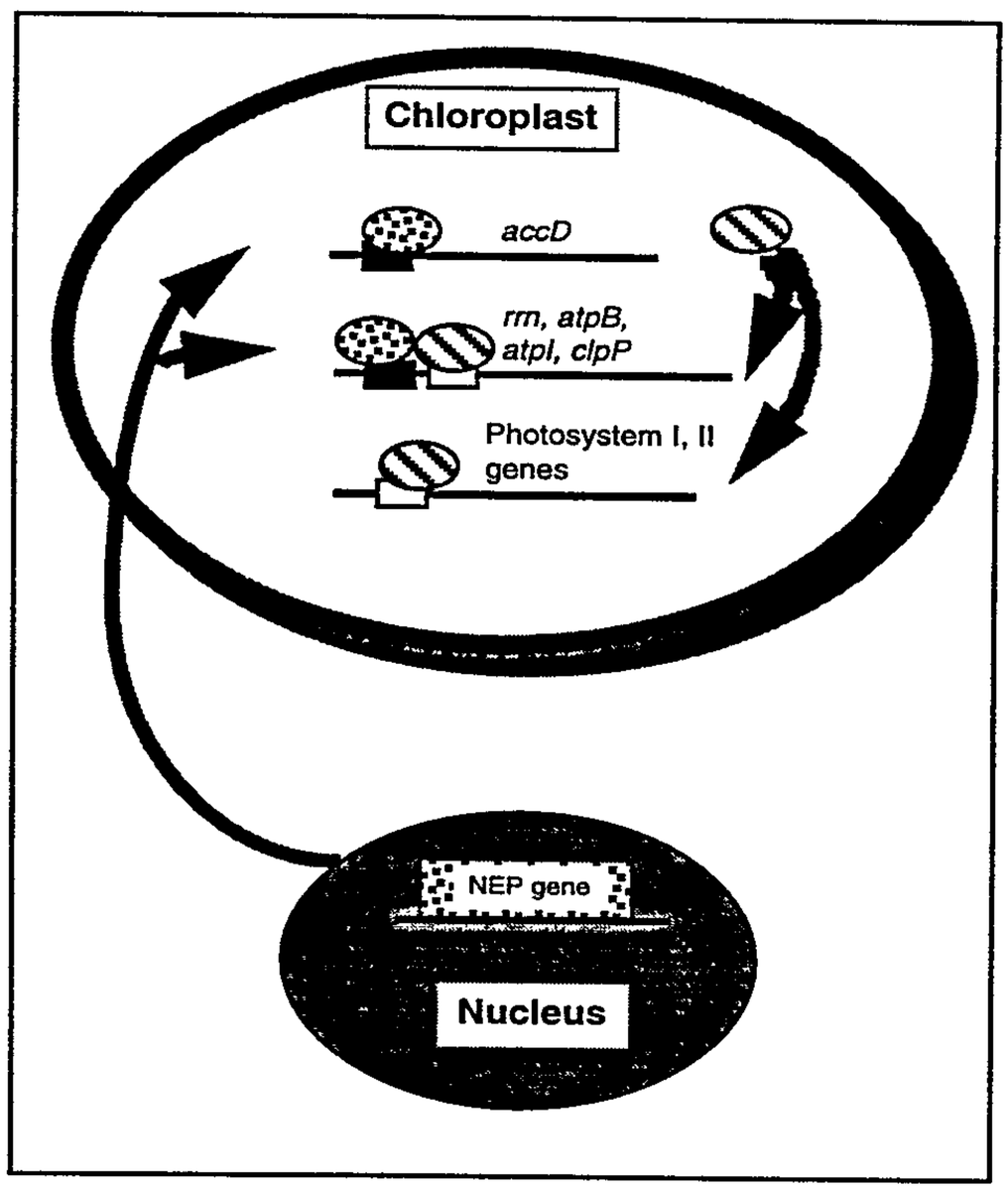

Chloroplast
accD
rrn, atpB,
atpI, clpP
Photosystem I, II
genes
NEP gene
Nucleus
NEP polymerase
PEP polymerase

FIGURE 12

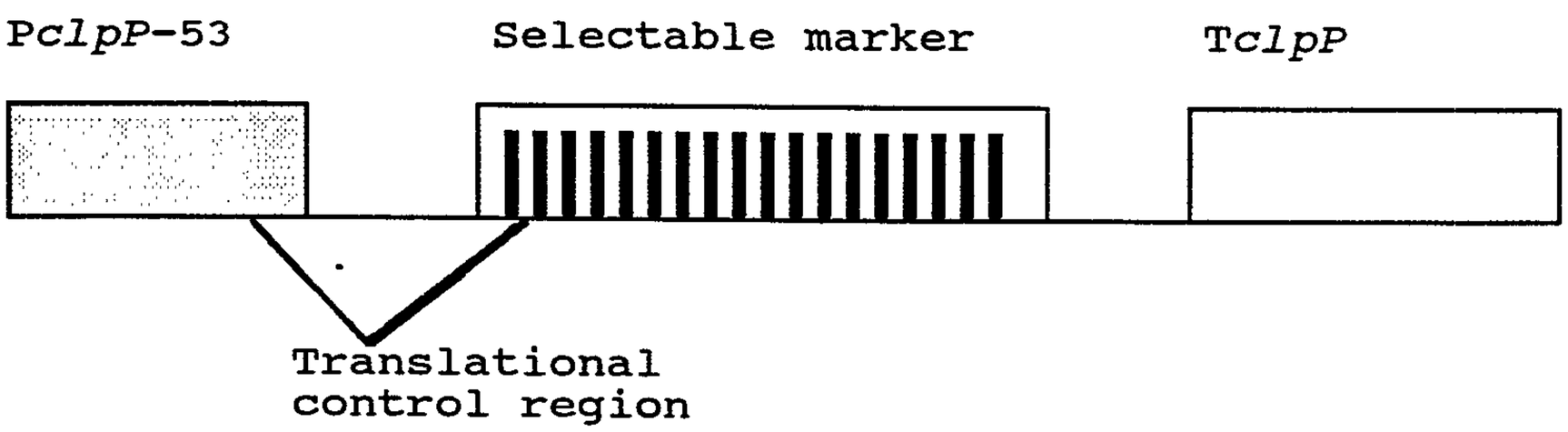
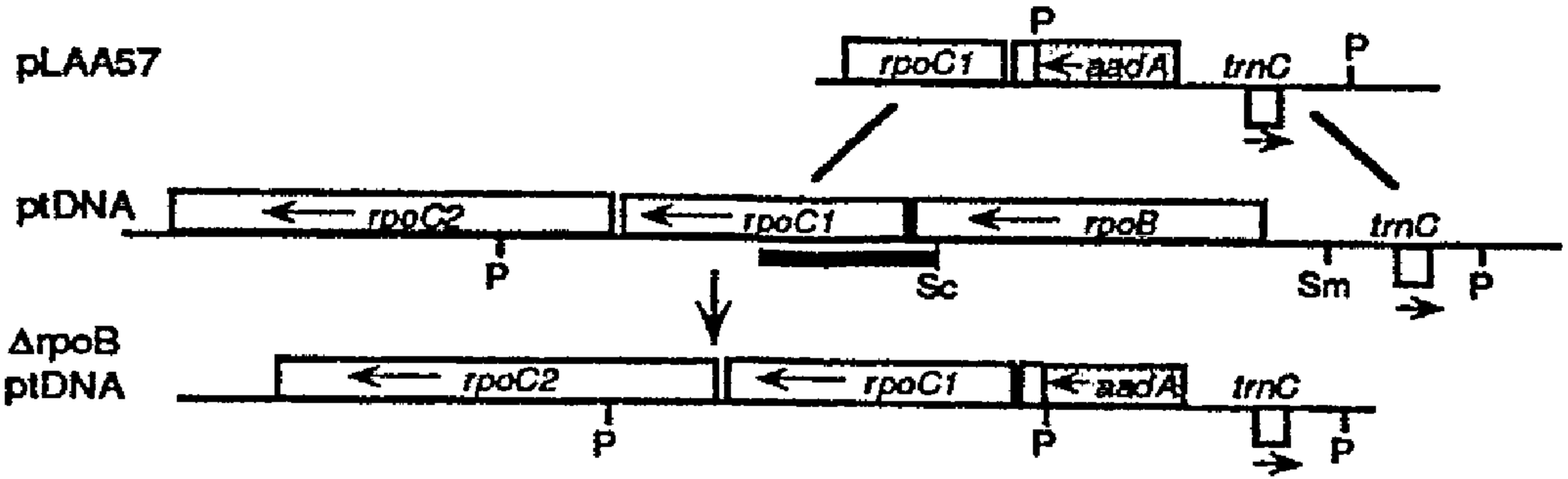
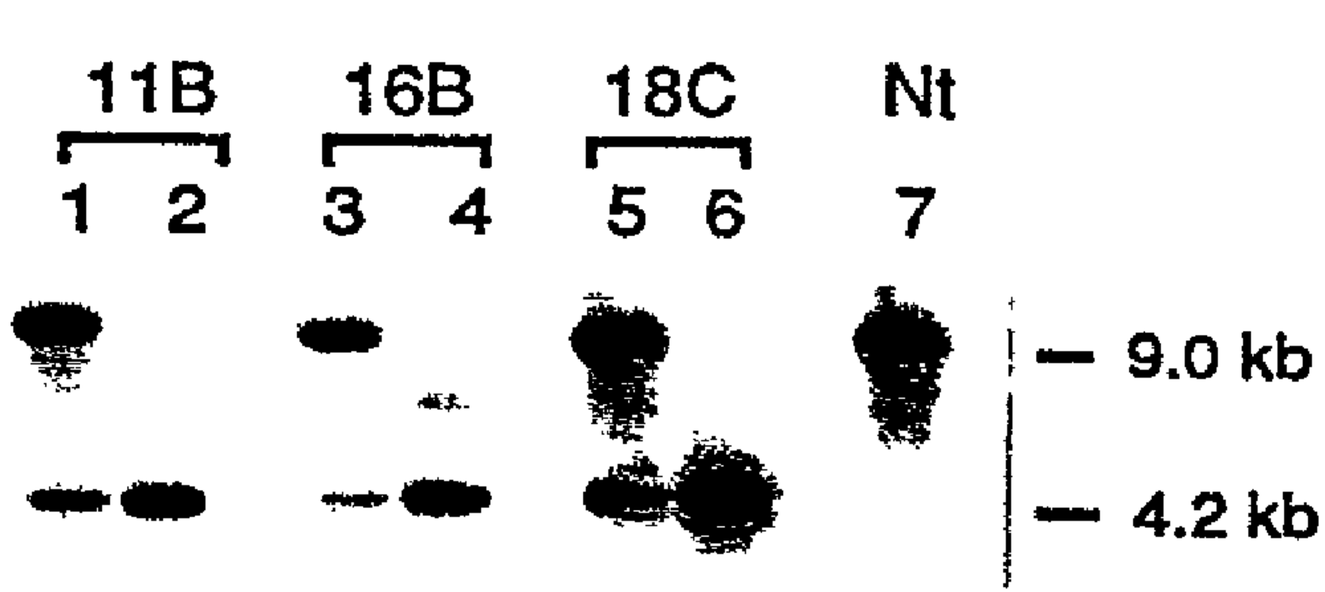

PclpP-53
Selectable marker
TclpP
Translational
control region

FIGURE 13

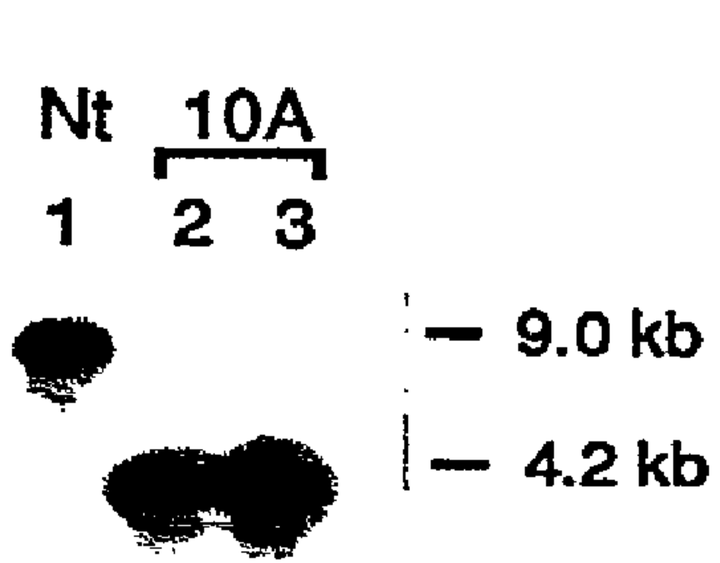
A


pLAA57
P
rpoC1
aadA
trnC
P
ptDNA
rpoC2
rpoC1
rpoB
trnC
P
Sc
Sm
P
ΔrpoB
ptDNA
rpoC2
rpoC1
aadA
trnC
P
P
P

B


11B
16B
18C
Nt
1 2
3 4
5 6
7
9.0 kb
4.2 kb

C


Nt
10A
1 2 3
9.0 kb
4.2 kb