PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C12N 15/29, 15/62, 15/64, 15/82	A1	(11) International Publication Number: WO 97/47771 (43) International Publication Date: 18 December 1997 (18.12.97)

(21) International Application Number: PCT/US97/10318

(22) International Filing Date: 13 June 1997 (13.06.97)

(30) Priority Data:
60/019,741 14 June 1996 (14.06.96) US

(71) Applicant *(for all designated States except US)*: RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY [US/US]; Old Queens, Somerset Street, New Brunswick, NJ 08903 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MALIGA, Pal [US/US]; 72 Yorktown Road, East Brunswick, NJ 08816 (US). CARRER, Helaine [BR/BR]; Apartment 24, Rua Sao Joao, 1344, CEP-13416-790 Piracicaba, SP (BR). CHAUDHURI, Sumita [US/US]; 68C Cedar Lane, Highland Park, NJ 08904 (US).

(74) Agents: HAGAN, Patrick, J. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19102 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, 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, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), 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: EDITING-BASED SELECTABLE PLASTID MARKER GENES

(57) Abstract

Disclosed are novel DNA constructs for selecting plastid transformants in higher plants. Also disclosed are editing based selectable marker genes which require editing at the transcriptional level for expression of the selectable marker gene. Vectors including such edited upstream sequences operably linked to slectable marker genes facilitate the isolation of plastid, rather than nuclear transformants in higher plants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

EDITING-BASED SELECTABLE PLASTID MARKER GENES

Pursuant to 35 U.S.C. §202(c), it is hereby acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from the National Science
5 Foundation.

FIELD OF THE INVENTION

This invention relates to the field of plant molecular biology. Specifically, DNA constructs are
10 provided that facilitate the selection of stably transformed plastids in multicellular plants for which the encoded RNA is modified post-transcriptionally.

15 **BACKGROUND OF THE INVENTION**

Several publications are referenced in this application by author names and year of publication in parenthesis in order to more fully describe the state of the art to which this invention pertains.
20 Full citations for these references are found at the end of the specification. The disclosure of each of these publications is incorporated by reference herein.

Genetic engineering of plants involves the
25 development and application of technology for genetic transformation through the direct manipulation of the plant genome and plant gene expression by the introduction of novel DNA. One method of transformation employs a derivative of the tumor
30 inducing (Ti) plasmid from the bacterium, *Agrobacterium tumefaciens*. Other methods utilize direct gene transfer into protoplasts using biolistics, electroporation, polyethylene glycol treatment.

While the incorporation of transforming DNA in the nucleus of plant cells is well known to those skilled in the art, transformation protocols that selectively identify transformed plastid DNA, to the exclusion of other genetic compartments have not yet been described. With the above-described methods, if plastid transformation only is desired, nuclear transformants may express the gene encoding the selectable marker and result in the generation of false positives.

The need for plastid-specific marker genes is based on this observation. In earlier work, selection for kanamycin resistance of pTNH32-bombarded tobacco leaves yielded a large number of nuclear transformants (Carrer et al., 1993). Indeed, recovery of nuclear gene transformants with other plastid *kan* genes (Cornelissen and Vandewiele 1989), and with promoterless *kan* constructs (Koncz et al. 1989) confirms that kanamycin resistant clones may be readily obtained by transformation with constructs that were not designed for expression in the nucleus. Additionally, nuclear gene transformants in tobacco may also be recovered by selection for spectinomycin resistance genes designed for expression in plastids.

Given the large number of plastid genomes in plant cells, the ability to select for the transformed genome in culture is a key element in achieving successful transformation. Selection markers have been identified by screening cultured plant cells for mutants resistant to various substances, such as antibiotics and herbicides. Such antibiotics and herbicides are listed in Table I, below. However, to date, a method has not been developed that will facilitate plastid transformation with the concomitant exclusion of the selection of

nuclear transformants. The development of such a system minimizes the false positives that result when a nuclear transformation event occurs.

RNA editing is a process that post-transcriptionally alters RNA sequences. Until recently, it was believed that chloroplasts, in contrast to mitochondria, did not utilize RNA editing and that the prediction of amino acid sequences from the corresponding gene sequences was generally correct. While most chloroplast genes begin with the canonical ATG start codon, genes have been identified that encode an ACG at a position that corresponds to the 5' terminal ATG in homologous genes in other species. Recently it has been shown that this ACG codon is not conserved at the mRNA level. It is converted to a functional AUG codon by C to U editing (Hoch et al., 1991). Most of the edited codons found to date, restore amino acids that are conserved in the corresponding peptides from chloroplasts of other species. This editing process is plastid specific. Genes edited in the plastid are not edited in the nucleus or other organelles of the plant.

The present invention provides DNA constructs and methods to facilitate the selection of stably transformed plastids, based upon a requirement for RNA editing in the transforming constructs which occurs exclusively in the plastid. Targeted manipulation of the plastid genome can now be performed with greater ease. Such manipulations include gene replacement, gene deletion, insertion of foreign genes and expression of recombinant proteins in plastids.

SUMMARY OF THE INVENTION

This invention provides DNA constructs and methods for the selection of stably transformed plastids of multicellular plants. The DNA constructs

of the invention can be used for the exclusive selection of plastid transformants. Nuclear transformants will not be selected with the constructs of the instant invention.

5 According to one aspect of the invention, chimeric DNA constructs are described containing an edited gene segment translationally fused to a selectable marker gene. Following editing at the RNA level, which occurs in the plastid, the selectable
10 marker gene is expressed. Cells or tissues are maintained on the selection medium until they have reached a homoplasmic condition, in which substantially all of the plastids of the cell or tissue have been transformed.

15 In a preferred embodiment of the invention, the above described chimeric construct is incorporated into a vector containing the necessary homologous sequences for targeted integration into the plastid genome. The targeting segment is of sufficient size
20 to promote homologous recombination with a pre-determined plastid genome sequence, thereby replacing that sequence in the genome of the transformed plastid. The vector may further comprise a foreign gene of interest to beneficially augment the
25 phenotype of the plant. In yet another embodiment of the invention, the chimeric DNA constructs may contain sequences that direct tissue specific regulation of the foreign gene of interest.

The method of the present invention is generally
30 applicable to the selection of stably transformed plastids in both monocotyledonous and dicotyledonous plants. Following selection, the cells or tissues expressing the selectable phenotype are regenerated into multicellular plants.

35

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic of the tobacco plastid *psb*E operon and the *psb*F and *psb*L DNA and amino acid sequences. The edited *psb*L initiation codon (from ACG to AUG) is underlined. The Δ*psb*F/Δ*psb*L region is bounded by dashed lines. The positions of oligonucleotides O1, O6, O7, O15 and O16 are marked. The DNA sequence is numbered according to Shinozaki et al. (1986).

Figure 2 is a physical map and partial DNA sequence showing, the E*aadA* gene in plasmid pJLM20, Figure 2A, and the E*kan* gene in plasmid pJLM18 in Figure 2B. The conserved -10/-35 promoter elements and ribosome binding site (RBS) are underlined in the *Prrn* sequence. DNA sequence derived from the Δ*psb*F/Δ*psb*L region is bounded by dashed lines, new sequence introduced during construction is in the solid box. The edited *psb*L initiation codon (from ACG to AUG) is underlined. T*rps*16 is the 3'-untranslated region of the plastid rps16 ribosomal protein gene. The positions for oligonucleotides O1, O2 and O3 in E*aadA*, and for O1, O4 and O5 in E*kan* are indicated. Abbreviation of restriction sites: B, *Bsp*HI; H, *Hind*III; N, *Nco*I; S, *Sac*I; X, *Xba*I.

Figure 3 shows a gel and an autoradiogram illustrating editing of the E*aadA* mRNA in the Nt-pHC94-1 plant, and of the E*kan* mRNA in the Nt-pJLM23-2 plant. Figure 3A shows PCR amplification products from DNA (lanes 1, 5) and cDNA (lanes 2, 6) with primers O1 and O2 for E*aadA*, and primers O1 and O4 for E*kan* templates. The location of primers is shown in Figures 2 A & B. Controls were amplification reactions carried out with DNase I-treated RNA (lanes 3, 7) and buffer only (lanes 4

and 8) using the same primers. Figure 3B shows the DNA and cDNA sequence of *EaadA* in the Nt-pHC94-1 plant, and Figure 3C shows *Ekan* in the Nt-pJLM23-2 plant. The amplified products were directly
5 sequenced with primers O3 (*EaadA*) and O5 (*Ekan*). Due to the polarity of primers, the sequence shown is complementary to the mRNA. The editing site in the sequence is marked by an arrowhead. Note a mixture of A and G nucleotides at the editing site in the
10 cDNA samples indicating partial editing.

Figure 4 depicts a gel and autoradiograms illustrating editing of the *psbL* mRNA in the transgenic plants. Figure 4A depicts the PCR
15 amplification products from DNA (lanes 1, 5, 9) and cDNA (lanes 2, 6, 10) from wild-type, Nt-pHC94-1 (*EaadA*) and Nt-pJLM23-2 (*Ekan*) plants with primers O1 and O6. See Figure 1. Controls were amplification reactions carried out with DNase I-treated RNA (lanes
20 3, 7, 11) and buffer only (lanes 4, 8, 12). The DNA and cDNA sequence of *psbL* in wild-type is shown in Figure 4B. Figure 4C shows the sequences in Nt-pHC94-1 (*EaadA*). Figure 4D shows the sequences in Nt-pJLM23-2 (*Ekan*) plants. The amplified products
25 were directly sequenced with primer O7. The sequence shown is complementary to the mRNA sequence due to the polarity of the O7 primer. The editing site is indicated by an arrowhead. Note nearly complete editing in the wild type (G* is very faint) and
30 partial editing in the transgenic plants.

Figure 5 shows a partial DNA map of the plastid genome and an autoradiogram depicting the steady-state levels of *psbL* and chimeric mRNAs.
35 Figure 5A depicts a partial map of the plastid genome with the *EaadA* and *Ekan* genes obtained by transformation with the pHC94 or pJLM23 plasmids.

The 16SrDNA and *trnV* genes, and the *rps*12/7 operon are marked. Horizontal arrows indicate mRNAs detected by the O14 oligonucleotide probe. The autoradiogram in the upper panel in Figure 5B shows that the O14 oligonucleotide detects the similar size (1.1-kb) *psb*E, E*aad*A and E*kan*, and the 2.2-kb E*kan-aad*A transcripts. Additional, minor uncharacterized RNA species are also visible which were not included in the quantitation. Total cellular RNA (2 μg per lane) was loaded from a wild-type plant (Wt), plasmid pHC94-transformed plants (Nt-pHC94-1, Nt-pHC94-11, Nt-pHC94-21) and plasmid pJLM23-transformed plants (Nt-pJLM23-2, Nt-pJLM23-14, Nt-pJLM23-18). The lower panel in Figure 5B illustrates the accumulation of *psb*E mRNA detected by the *psbJ* probe, and of the 16S rRNA as the loading control. The filter was stripped of the labeled O14 oligonucleotide, and probed with a mixture of the *psbJ* and 16SrDNA probes. The *psb*E probe was obtained by PCR amplification of the *psbJ* region with primers O15 and O16 shown in Figure 1. The 16SrDNA probe was a 2.4-kb *Eco*RI/*Eco*RV ptDNA fragment defined by the restriction sites at nucleotides 138448/141847 of the plastid genome (Shinozaki et al., 1986).

Figure 6 is a series of autoradiograms illustrating editing of the *rpo*B and the *ndh*B transcripts in the wild-type and transgenic Nt-pHC94-1 plants. The DNA and cDNA sequences corresponding to each gene were PCR amplified using the following primers: O8 and O9 for *rpo*B; O10 and O11 for *ndh*B. The sequencing primer for *rpo*B was O8, for *ndh*B was O1. The editing site in the sequence is marked by an arrowhead. Arrowhead points at C in DNA which is edited to T at sites I and II of the *rpo*B and *ndh*B transcripts.

Figure 7 is a partial map and a series of autoradiograms illustrating the approach by which the region required for *psb*L editing was defined. Figure 7A shows the map of the chimeric Δ*psb*L/kan gene, with the 98 nt Δ*psb*F/Δ*psb* fragment enlarged at the top. The positions of primers 04, 05 and 017 are indicated. 66,780 and 66,683 are the nucleotides at the ends of the 98 nt Δ*psb*F/Δ*psb*L fragment in the tobacco plastid genome (Shinozaki et al., 1986). The lower portion of Figure 7A is a listing of pPRV111A plasmid derivatives which carry chimeric Δ*psb*L/kan genes. The nucleotide position at the end of the Δ*psb*F/Δ*psb*L deletion derivatives is given relative to the edited C (position 0; arrow). The efficiency of editing of the chimeric Δ*psb*L/*kan* mRNA (%), and the kanamycin resistance phenotype of the transgenic plants is listed. Figure 7B is an autoradiogram demonstrating editing of the *psb*L site in the chimeric mRNAs. The cDNAs were PCR amplified with primers 017 and 04 and directly sequenced with primer 05. Due to the polarity of 05, the sequence shown is complementary to the mRNA. Accordingly, A at the edited position indicates a C to U conversion event and a G an unedited C nucleotide.

Figure 8 is a partial DNA map and a pair of autoradiograms illustrating *psb*L editing in transgenic plants with mutations adjacent to the editing site. The position of the edited C (arrow) and the flanking nucleotides within the 98 nt Δ*psb*F/Δ*psb*L fragment are shown in the upper portion of Figure 8. Mutations in plasmids pSC14 and pSC15 are in lower case. Editing was tested by sequencing the chimeric cDNAs (bottom). Calculated editing efficiencies of the chimeric mRNA (%) are listed. For experimental details see legend to Figure 7.

Figure 9 shows a partial DNA map and a series of autoradiograms illustrating the existence of competition for the *psb*L-specific transfactor (*psb*L-SEF) in the transgenic plants. Figure 9A shows the map of the *psb*E operon containing the *psb*L gene, with the position of oligonucleotides 01, 06 and 07 used for PCR amplification and sequencing indicated. The 22 nt (-16/+5) sequence required for editing is shown. The edited C is marked by an arrow. The 16 nt segment competing for *psb*L-SEF is boxed. The plasmids used to obtain the transgenic plants are listed, as identified in Figures 7 and 8. Competition (+) was indicated by reduced editing efficiency of the *psb*L mRNA, as compared to nontransformed, wild-type plants. Figure 9B is an autoradiogram showing editing of *psb*L mRNAs. The cDNAs were PCR amplified with primers 01 and 06 and directly sequenced with primer 07. Due to the polarity of 07, the sequence shown is complementary to the mRNA. Accordingly, A at the edited position indicates a C to U conversion event and a G at the edited position, an unedited C nucleotide. A+G* denotes nearly complete editing (>99%) as in the wild-type plants. A+G denotes partial editing with ~10% unedited *psb*L transcripts.

Figure 10 is a partial DNA map and a series of autoradiograms illustrating that chimeric mRNAs containing the *ndh*D editing site do not compete for *psb*L-SEF. Figure 10A shows a partial map of the tobacco plastid genome containing the *ndh*D, *psa*C and *ndh*E genes, and the DNA sequence with the edited *ndh*D translation initiation codon (underlined). The genes are marked and the DNA sequence is numbered according to Shinozaki et al., 1986. The Δ*ndh*D segment in a dashed box was translationally fused with the *kan* gene, as shown in Figure 10B. The position of

primers 018, 019 and 020 are indicated. The A nucleotide 26 bp upstream of the editing site (underlined) was changed to a C during construction of the chimeric gene. Figure 10B shows the Δ*ndh*D/kan chimeric gene in plasmid pSC23 expressed in the *Prrn*/T*rps*16 cassette. The positions of primers 04, 05 and 017 are indicated. Figure 10C depicts autoradiograms demonstrating editing of *ndh*D and *psb*L sites in wildtype (Nt-wt), Nt-pSC23 and Nt-pSC2 plants. Editing of the *ndh*D site was studied in the endogenous *ndh*D, and the chimeric Δ*ndh*D/kan mRNAs. Editing of the *psb*L site was studied in the endogenous *psb*L, and the chimeric Δ*psb*L/*kan* mRNAs. The *ndh*D cDNA was amplified with primers 018 and 019 and sequenced with primer 020. The *psb*L cDNA was PCR amplified with primers 01 and 06 and directly sequenced with primer 07. The Δ*ndh*D/*kan* and Δ*psb*L/*kan* cDNAs were amplified with primers 017 and 04, and sequenced with 05. Due to the polarity of the sequencing primers, the sequence shown is complementary to the mRNA. Accordingly, A at the edited position indicates a C to U conversion event and a G, an unedited C nucleotide.

Figure 11 is a schematic illustration of the Δ*rpl*2/*kan* gene. A partial map of the maize plastid genome containing the *trn*I, *rpl*23, *rpl*2 and *trn*H genes, and the DNA sequence with the edited *rpl*2 translation initiation codon (underlined) is shown in Figure 11A. The genes are marked and the DNA sequence is numbered according to Maier et al., 1995. The Δ*rpl*2 segment in a dashed box was translationally fused with the *kan* gene, as shown in Figure 11B. Figure 11B demonstrates that the Δ*rpl*2D/*kan* chimeric gene in plasmid pSC22 is expressed in the *Prrn*/*Trps*16 cassette. The position of primers 04, 05 and 017 is indicated.

Figure 12 is a schematic drawing of minigenes used to test editing in segments of *ndh*B and *rpo*B RNAs. The *nhd*B plastid gene and its minigene are shown in the upper portion of Figure 12 and the *rpo*B plastid gene and its corresponding minigene (lower portion of Figure 12) are shown. Editing sites are indicated in the Figure. Minigene RNAs are expressed in a cassette which contains the rRNA operon promoter (*Prrn*) and the 3' untranslated region of the *rps*16 ribosomal protein gene (T*rps*16) required for mRNA stability (Zoubenko et al., 1994). The editing sites and the references are listed in Table IV.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, methods and DNA constructs are provided to facilitate the selection of transformed plastids following delivery of transforming DNA. The constructs of the invention will be expressed only if they are appropriately edited at the RNA level within the plastid. In so far as it is known, the methods and DNA constructs described herein have heretofore been unavailable for multicellular plants.

The following definitions will facilitate the understanding of the subject matter of the present invention:

Heteroplasmic: refers to the presence of a mixed population of different plastid genomes within a single plastid or in a population of plastids contained in plant cells or tissues.

Homoplasmic: refers to a pure population of plastid genomes, either within a plastid or within a population contained in plant cells and tissues. Homoplasmic plastids, cells or tissues are genetically stable because they contain only one type of plastid genome. Hence, they remain homoplasmic even after the selection pressure has been removed,

and selfed progeny are also homoplasmic. For purposes of the present invention, heteroplasmic populations of genomes that are functionally homoplasmic (i.e., contain only minor populations of wild-type DNA or transformed genomes with sequence variations) may be referred to herein as "functionally homoplasmic" or "substantially homoplasmic." These types of cells or tissues can be readily purified to homoplasmy by continued selection on the non-lethal selection medium. Most seed progeny of such plants are homoplasmic in the absence of selection pressure, due to random sorting of plastid genomes.

Plastome: the genome of a plastid.

Transplastome: a transformed plastid genome.

Transformation of plastids: stable integration of transforming DNA into the plastid genome that is transmitted to the seed progeny of plants containing the transformed plastids.

Selectable marker: the term "selectable marker" refers to a phenotype that identifies a successfully transformed organelle, cell or tissue, when a gene or allele encoding the selectable marker is included in the foreign DNA used for transformation.

Transforming DNA: refers to homologous DNA, or heterologous DNA flanked by homologous DNA , which when introduced into plastids becomes part of the plastid genome by homologous recombination.

Edited gene segment: refers to a region of DNA which encodes an RNA which is post-transcriptionally altered.

Translationally fused: refers to two coding regions of two separate genes spliced together in a construct such that both regions will be expressed at the protein level. In accordance with the present invention translation of the chimeric protein is

dependent on appropriate editing of the upstream coding region at the mRNA level.

The detailed description as follows provides examples of preferred methods for making and using
5 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
10 Sambrook et al., "DNA Cloning, A Laboratory Manual," Cold Spring Harbor Laboratory, 1989.

In the detailed description and examples set forth hereinbelow, a preferred embodiment comprises a DNA segment that encodes an edited RNA segment
15 operably linked to a second DNA segment which encodes a selectable marker. In the following examples in tobacco chloroplasts are exemplified. Transformation vectors containing such combinations will be useful in enabling plastid-specific transformation.
20 References made to positions and sequences on the tobacco chloroplast genome are taken from Shinozaki et al., EMBO J., 5: 2043-49 (1986), which discloses the complete nucleotide sequence of the Nicotiana tabacum chloroplast genome. Although tobacco is
25 exemplified, it will be appreciated by those skilled in the art that the DNA constructs and methods of the present invention can be adapted to plastids of other plant species.

Plastid transformation requires: (1) a method
30 for delivering DNA through the double membrane of the plastid; (2) integration of the heterologous DNA without interfering with the normal function of the plastid genome; and (3) efficient selection for the transplastome. Methodology for performing efficient
35 transformation of plastids of multicellular plants is set forth in U.S. Patent No. 5,451,513 issued

September 19, 1995, the entire disclosure of which is incorporated by reference herein.

In accordance with the present invention, it has been discovered that the selection criterion for identifying transplastomes is critical to the success of stable plastid transformation in higher plants. Accordingly, the selection technique of the present invention employs DNA encoding a selectable phenotype ("selectable marker") in the transforming DNA. Selection greatly facilitates obtaining transplastomic lines, due in part to the large number of identical plastid genome copies present in each plant cell (3,000-12,000 copies localized in up to 100 plastids in tobacco, as compared with 80 copies carried by a single plastid in *Chlamydomonas*). Selectable phenotypes can include antibiotic resistance, herbicide resistance, drug resistance or resistance to toxic analogs of metabolites.

The present invention provides selectable marker genes that require RNA-editing processes which occur in plastids only, not in the nucleus or in mitochondria. Such plastid specific marker genes will greatly enhance the ability to obtain stably transformed plastids in multicellular plants. The novel combination of plastid specific editing site controlling expression of a selectable marker is hereinafter described.

Certain mRNA sequences can be altered post-transcriptionally by a process known as RNA editing, so that their final nucleotide sequence differs from that encoded by the DNA sequence. The process has been detected in divergent organisms including trypanosomes, *Physarum polycephalum*, mammals, viruses and higher plants involving widely different molecular mechanisms (reviewed in Benne, 1994; Chan, 1993; Gray and Covello, 1993; Innerarity et al., 1996; Simpson and Thiemann, 1995).

In higher plants, editing of plastid and mitochondrial RNAs involves C to U conversions and rare cases of U to C changes in mitochondria. The number of editing sites in plastids is estimated to be about 25 (Maier et al., 1995) while in plant mitochondria it is 1000 or more (Schuster and Brennicke, 1994). Comparison of sequences surrounding editing sites have failed to identify any conserved primary sequence and/or structural motifs that could direct the site-selection process. The recent development of an *in vitro* editing system should lead to accelerated progress in the analysis of RNA editing in plant mitochondria (Araya et al., 1992; Yu and Schuster, 1995). Although an *in vitro* system for editing in plastids is still lacking, the availability of plastid transformation allows an *in vivo* approach to study plastid editing (Bock et al., 1994; Bock and Maliga, 1995; Sutton et al., 1995,.

While RNA editing has been reported to occur in plastids and mitochondria, it has not been observed in the nucleus (Kossel et al., 1993; Hanson et al., 1995). Furthermore, mitochondrial transcripts are not edited in plastids (Sutton et al., 1995). The discovery of RNA editing and the problem of recovering large numbers of nuclear transformants after bombardment of plant cells with plastid directed constructs, led to the design of the plastid transgenes of the invention which are expressed in plastids but not in other genetic compartments of the cell. The following examples describe the transgenes of the invention.

Briefly, one example of an editing based selectable marker gene utilizes the N-terminal segment of an edited plastid gene translationally fused to the coding region of an antibiotic resistance gene. The expression of the antibiotic resistance gene is dependent upon RNA editing of the

construct. In a specific example, the N-terminal segment of *psb*L is fused to the coding region of *aad*A gene. Translation of the *aad*A gene is dependent on editing, and is used to recover plastid transformants by direct selection (Chaudhuri et al., 1995). Additionally, a *psb*L based editing selectable marker can comprise the similar Δ*psb*L/*kan* chimeric gene (Chaudhuri and Maliga, 1996), in which kanamycin resistance is a reliable measure of the editing of the translation initiation codon (Chaudhuri et al., 1995; Chaudhuri and Maliga, 1996). Although these genes confer kanamycin resistance to the plant cell when present in each of the plastid genome copies in a cell, they could not be used for direct selection. Presumably this is because plastid transformants can only be directly selected by the kanamycin resistance marker if the chimeric genes are expressed at high levels. Direct selection for E*kan* genes should be feasible by improving their expression level through appropriate engineering.

A second, novel example of an editing based selectable marker gene utilizes the *ndh*D edited segment. This construct, Δ*ndh*D/*kan* was obtained by fusing the N-terminal segment of *ndh*D and the coding region of *kan* (Chaudhuri and Maliga, 1996). The above described selectable marker genes will facilitate the selection of transformants in certain dicot species.

A third type of editing based marker gene, Δ*rpl*2/*kan*, is created by fusing the *rpl*2 edited segment to the kanamycin coding region. Such chimeric selectable marker genes will be used to advantage in selecting plastid transformants in monocots.

In tobacco plastids, functional *psb*L and *ndh*D mRNA is created by editing an ACG codon to an AUG translation initiation codon. To determine if

editing may occur in a chimeric mRNA, the N-terminal part of *psb*L containing the editing site was translationally fused with the *aad*A and *kan* bacterial genes as described in the following examples. The chimeric constructs were introduced into the tobacco plastid genome by targeted gene insertion. Deletion derivatives of a 98 nt fragment were expressed as parts of chimeric transcripts to define the *cis* sequences required for *psb*L editing. In accordance with the instant invention, it has been found that a 22 nt fragment is sufficient to direct *psb*L editing. Although the 22 nucleotides were required for editing, only 16 nucleotides competed for the *psb*L-specific editing factor.

Expression of the chimeric gene transcripts led to a significant decrease in the editing efficiency of the endogenous *psb*L mRNA. However, the efficiency of editing in the transplastomic lines was unchanged for four sites in the *rpo*B and *ndh*B mRNAs. Reduced efficiency of *psb*L editing, but not of the other four sites, in the transplastomic lines indicates depletion of *psb*L-specific editing factor(s). This finding implicates the involvement of site-specific factors in editing of plastid mRNAs in higher plants.

In addition to *psb*L, editing was shown to create the AUG translation initiation codon for *ndh*D in tobacco (Neckermann et al., 1994). To test whether editing of initiation codons involves a common depletable *trans*-factor, a chimeric gene containing the *ndh*D editing site was expressed in tobacco plastids. The data show that, as for *psb*L, editing of the *ndh*D site requires a depletable *trans*-factor. However, this *trans*-factor is distinct from that required for *psb*L editing.

In maize plastids, the translation initiation codon of *rpl*2 is created by editing (Hoch et al., 1991). To test, whether the ACG codon in the maize

*rpl*2 context is edited in tobacco plastids, a Δ*rpl2/kan* gene was constructed by translationally fusing the N-terminal segment of *rpl*2 with the *kan* coding region. The chimeric mRNA is not edited in tobacco, but provides a useful selectable marker in cereals such as maize and rice.

The following examples are provided to merely illustrate typical protocols for carrying out the instant invention. They are not intended to limit the scope of the invention in any way.

EXAMPLE I

EDITING-BASED Ekan and EaaDA SELECTABLE MARKER GENES

A. Construction of the chimeric genes

The *psb*L gene encodes a peptide of photosystem II and is part of the *psb*E operon (Carillo et al., 1986; Figure 1 above). A 98-nucleotide fragment spanning the *psb*L editing site, Δ*psb*F/Δ*psb*L, was cloned upstream of the spectinomycin resistance gene (*aad*A) coding sequence such that the N-terminus of *psb*L was translationally fused with *aad*A. The Δ*psb*F/Δ*psb*L fragment contains 40 nucleotides of the *psb*F C-terminus, 22 nucleotides of the intergenic region between *psb*F and *psb*L and 36 nucleotides of the *psb*L N-terminus. The construct was cloned in the *Prrn*/T*rps*16 plastid expression cassette (Figure 2A). *Prrn* contains the plastid rRNA operon promoter, a ribosome binding site and a translational initiation codon (ATG). In the chimeric construct, the truncated *psb*F coding region forms an open reading frame with the *Prrn* initiation codon (ATG), whereas the translation of the E*aad*A reading frame (*psb*L-*aad*A fusion peptide) is dependent on the creation of a translation initiation codon (AUG from ACG) by editing the *psb*L site. Note that the two coding

regions are in different reading frames in the *Eaad*A mRNA. See Figure 2A.

5 The *Ekan* gene was obtained by translationally fusing *psb*L with kan, a kanamycin resistance gene encoding neomycin phosphotransferase using the same Δ*psb*F/Δ*psb*L fragment, shown in Figure 2B. *Ekan* is similar to the *Eaad*A gene, except that it has 39 nucleotides instead of 36 nucleotides of the *psb*L N-terminus. A detailed technical description of the
10 gene construction is set forth below.

The *Ekan* gene, shown in Figure 2B, in plasmid pJLM18 was constructed in a pBluescript KS+ plasmid (Stratagene). The *Ekan* coding region in pJLM18 is expressed in the *Prrn*/*Trps*16 cassette. The *Prrn*
15 5'-regulatory region consists of the plastid rRNA operon promoter and a ribosome binding site and is on an *Eco*RI/*Nco*I fragment. *Prrn* derives from plasmid pZS195, the progenitor of plasmid pZS197 (Svab and Maliga, 1993) in which the translational initiation
20 codon (ATG) is included in the *Nco*I site. The *Nco*I site of *Prrn* was ligated to the *Bsp*HI site of a *Bsp*HI/ *Xba*I fragment; the *Nco*I/*Bsp*HI fusion eliminated both restriction sites. The *Bsp*HI/*Xba*I oligonucleotide was obtained by annealing the
25 overlapping 5'CATTCATGACTTTGGGATCAATATCAGCATATGCA GTTCATCCAACGATAAACTTAATCCGAATTATAGAGC-3' and 5'CGGTCTGAATTCAATTCAACATTTTGTTCGTTCGGGTTTGATTGTGTCGTA GCTCTATAATTCGGATTAAG-3' single-stranded oligonucleotides and extension with the Klenow
30 fragment of DNA polymerase I. The *Bsp*HI/*Xba*I fragment contains the sequence framed in Figure 2B, including the ΔpsbF/Δ*psb*L sequence encoding the C-terminal end of *psb*F, the intergenic region and the N-terminal portion of *psb*L. As the result of the
35 *Nco*I/*Bsp*HI fusion, the C-terminal end of *psb*F is translated from the *Prrn* translational initiation codon (ATG). To translationally fuse the 14

N-terminal codons of *psbL* with the kan coding region, the *Xba*I single-stranded overhang of the *Bsp*HI/*Xba*I fragment and the single-stranded overhang of the *Nco*I site of *kan* (including the translational initiation codon) was removed by mung bean nuclease treatment, and subsequently ligated. The *kan* coding region derives from plasmid pTNH4 as an *Nco*I/*Xba*I-fragment (Carrer et al., 1993). The T*rps*16 fragment is contained within an *Xba*I/ *Hind*III fragment, and was linked to the E*kan* coding region via the *Xba*I site. The Trps16 fragment contains the *rps*16 gene 3'-regulatory region between nucleotides 5,087 to 4,939 in the ptDNA (Shinozaki et al., 1986). The *Xba*I-site at the 5'-end of the fragment was created by oligonucleotide -directed mutagenesis; the 3'-end of the fragment was excised from the plastid genome at an EcoRI-site at nucleotide position 4,938. (Staub and Maliga, 1994). The *Eco*RI-site was subsequently converted to a *Hind*III-site by linker-ligation. For introduction into the plastid genome, the E*kan* construct was cloned as an *Eco*RI/*Hind*III fragment in the multiple cloning site of plastid vector pPRV111B (Zoubenko et al., 1994; Gene Bank Accession No. U12813), which is adjacent to a selectable *aad*A gene.

The E*aad*A gene shown in Figure 2A, in plasmid pJLM20 was constructed in a pBluescript KS+ plasmid as described for the E*kan* gene. The *Nco*I/*Xba*I fragment containing the aadA coding region is derived from plasmid pHC1 (Carrer et al., 1991) and the *aad*A coding region is translationally fused with the 12 N-terminal codons of the tobacco *psb*L gene. For introduction into the plastid genome, the E*aad*A gene was cloned in plastid insertion vector pPRV100B (Zoubenko et al., 1994, Gene Bank Accession No. U12811). The pPRV100B vector carries a multiple cloning site flanked by ptDNA sequences, but no selectable plastid marker gene.

While kanamycin, spectinomycin and/or streptomycin resistance is exemplified herein, the use of other selectable marker genes is contemplated. A list of such genes is set forth in Table I below
5 (Potrykus et al., (1995) in Gene Transfer to Plants, Springer Verlag.

TABLE I

Selectable marker genes for plant transformation

Selective agent	Marker gene	Gene Product
Kanamycin, G418	*nptII*	Neomycin; Phosphotransferase II
Gentamycin	*aacC3* *aacC4*	Gentamycin-3-N-acetyltransferase
Hygromycin	*hph, hpt*	Hygromycin phosphotransferase
Methotrexate	*dhfr*	Dihydrofolate reductase
Spectinomycin	*16S rDNA* *aadA*	16S rRNA Aminoglycoside-3'-adenyltransferase
Streptomycin	*SPT* *16S rDNA* *aadA*	Streptomycin phosphotransferase 16S rRNA Aminoglycoside-3'-adenyltransferase
Bleomycin Phleomycin	*ble*	
Blasticidin	*bsr*	Blasticidin S deaminase
Sulfonamide	*sul*	Dihydropteroate synthase
Phosphinothricin	*bar*	Phosphinothricin acetyltransferase
Chlorsulfuron	*als* *csr-1*	Acetolactate synthase
Bromoxynil	*bxn*	Bromoxynil nitrilase
Glyphosate	*EPSPS*	5-enolpyruvyl-shikimate-3-phosphate synthase
2,4-D	*tfdA*	2,4-dichlorophenoxyacetate monooxygenase
Atrazine	*psbA*	Q_B protein
2,2-DCPA		Dehalogenase
4-methyl-tryptophane	*tdc*	Tryptophane decarboxylase
Nitrate	*NR*	Nitrate reductase
S-aminoethyl-L-cysteine	*DHPS*	Dihydropicolinate synthase
lysine/threonine	*AK*	Aspartate kinase
aminoethyl-cysteine	*osc*	Octopine synthase

B. Transformation and selection of antibiotic resistant transplastomic lines

The *EaadA* gene was cloned into the plastid transformation vector pPRV100B (Zoubenko et al., 1994) to yield plasmid pHC94 which was introduced into tobacco chloroplasts by the biolistic process. The chimeric gene integrated into the plastid genome via two homologous recombination events in the *trnV-rps*7/12 intergenic region. In a sample of 50 bombarded leaves, selection for spectinomycin resistance resulted in the isolation of 43 spectinomycin resistant clones. Out of these, 34 were confirmed to carry the *EaadA* gene by DNA gel blot analysis (data not shown). Expression of antibiotic resistance indicated editing of the chimeric *EaadA*. The efficiency of selection for the *EaadA* gene, approximately one plastid transformant per bombarded leaf sample, was comparable to the efficiency of selection for an *aadA* gene whose expression was independent of editing (Svab and Maliga, 1993). Three independently transformed lines, Nt-pHC94-1, Nt-pHC94-10 and Nt-pHC94-11, were further studied. As direct selection for kanamycin resistance is inefficient (Carrer et al., 1993), the *Ekan* gene was linked to a spectinomycin resistance gene in transformation vector pPRV111B to yield plasmid pJLM23. Direct selection of plastid transformants was attempted after bombardment with pJLM23-plasmid coated tungsten particles. No kanamycin resistant clones were obtained in a sample of 200 bombarded leaves (100 each selected on 50 μg/ml and 100 μg/ml kanamycin sulfate). However, transgenic plants containing the *Ekan* gene were obtained by selection for the linked spectinomycin-resistance gene. Three independently transformed lines, Nt-pJLM23-2, Nt-pJLM23-14 and Nt-pJLM23-18, were further studied. Leaf segments

from each of the clones proliferated on kanamycin medium (50 μg/ml) indicating phenotypic expression of the E*kan* gene. The methods used for plastid transformation are described in greater detail below.

5 Tobacco (*Nicotiana tabacum* cv. Petit Havana) plants were grown aseptically on agar-solidified medium containing MS salts (Murashige and Skoog, 1962) and sucrose (30 g/l). Leaves were placed abaxial side up on RMOP media for bombardment. The

10 RMOP medium consists of MS salts, N6-benzyladenine (1 mg/l), 1-naphthaleneacetic acid (0.1 mg/l), thiamine (1 mg/l), inositol (100 mg/l), agar (6 g/l) at pH 5.8, and sucrose (30 g/l). The DNA was introduced into chloroplasts on the surface of 1 μm

15 tungsten particles using the DuPont PDS1000He Biolistic gun (Maliga, 1995). Spectinomycin resistant clones were selected on RMOP medium containing 500 μg/ml of spectinomycin dihydrochloride. Resistant shoots were regenerated

20 on the same selective medium, and rooted on MS agar medium (Svab and Maliga, 1993). Kanamycin resistant clones were selected on RMOP medium containing 50 or 100 μg/ml kanamycin sulfate (Carrer et al., 1993).

25 C. <u>Editing of E*aad*A, E*kan* and *psb*L transcripts.</u>

The phenotypic expression of antibiotic-resistance by E*aad*A and E*kan* plants indicated that the chimeric genes were edited since their translation was made dependent on the editing

30 of an ACG to an AUG initiation codon. To directly test for editing of E*aad*A and E*kan* mRNAs, cDNAs were PCR-amplified with primer O1 within the *psb*F coding region and primers O2 and O4 within the E*aad*A and E*kan* coding sequences, respectively. The position of

35 primers is shown in Figure 2. The PCR amplification products are shown in Figure 3A. Direct sequencing of the PCR products from three independently

transformed *Eaad*A lines and phosphorimager analysis indicated that approximately 70% of the *Eaad*A transcripts are edited. See Figure 3B and Table II.

TABLE II

Unedited mRNAs (%) in the wild-type and transgenic leaves

Plant Line	Sample	*psb*L	*Eaad*A	*Ekan*
Nt-wt	1	<0.1		
	2	0.7		
	3	0.3		
Nt-pJLM23-2	1	8.8		30.3
Nt-pJLM23-14	1	9.2		28.2
Nt-pJLM23-18	1	10.2		28.4
Nt-pHC94-1	1	9.5	28.7	
Nt-pHC94-10	1	9.0	30.4	
Nt-pHC94-11	1	10.4	29.9	

Radioactivity in bands in Figure 3 corresponding to nucleotides was determined by phosphorimager analysis. The values were normalized for DNA loading and labeling efficiency against six other bands in the same lanes. Percent unedited mRNA = [corrected unedited signal/(corrected edited + corrected unedited signal)] x 100.

A similar extent of editing was found for the E*kan* mRNAs as shown in Figure 3C and Table II. The partial editing was not due to the presence of contaminating DNA in the RNA samples since no PCR-amplified products were obtained from non-reverse transcribed DNase I-treated RNA samples. See Figure 3A; lanes 3 and 7. The *psb*L site in the chimeric transcripts was only partially (approximately 70%) edited while in leaves of wild-type plants the *psb*L mRNA is >99% edited (Kudla et al., 1992; Bock et al., 1993). Therefore it was of interest to determine whether or not the editing of the *psb*L mRNA is affected in the transgenic plants. The *psb*L cDNAs

were PCR-amplified with primers O1 and O6 within the *psb*F and *psb*J coding regions as shown in Figure 1A, from wild type and transgenic plants. Direct sequencing of the PCR products revealed that the transgenic plants contained approximately 10% unedited *psb*L mRNA. This indicates a >10-fold increase in the level of unedited *psb*L mRNA in the transgenic plants. See Figure 4 and Table II. Artifacts due to DNA contamination of RNA samples were excluded by the lack of PCR products from non-reverse transcribed DNase I-treated RNA samples. See Figure 4A, lanes 3, 7 and 11. Methods utilized to study the editing in plastid mRNAs are set forth below.

Total cellular DNA was isolated according to Mettler (1987). Total cellular RNA was extracted using TRIzol (Gibco BRL). Reverse transcription of proteinase K- and DNAse I- treated RNA samples were carried out as described by Kudla et al.(1992). DNA and cDNA were amplified by PCR according to standard protocols: 1 min at 92°C, 2 min at 55°C, 1.5 min at 72°C, 30 cycles.

The PCR amplification products were separated in 1.5% agarose gels and purified using the Geneclean II kit (BIO 101 Inc.). Direct sequencing of DNA was performed as described (Bachmann et al., 1990) using the Sequenase kit (USB) and the detergent Nonidet P-40.

The following is a list of primers used for PCR.

O1	5' -CAATATCAGCAATGCAGTTCATCC- 3'
O2	5' -CCAAGCGATCTTCTTCTTGTCCAA- 3'
O3	5' -GCGCTCGATGACGCCAAC- 3'
O4	5' -CACGACGAGATCCTCGCCG- 3'
O5	5' -GAATAGCCTCTCCACCCA- 3'
O6	5' -GGAATCCTTCCAGTAGTATCGGCC- 3'
O7	5' -GGAAAATAAAACAGCAAGTAC- 3'
O8	5' -CAAATATTGCAAAGTCCCGG- 3'

O9 5' -CCGGATCGCCACCTACAC- 3'
O10 5' -TGGCTATAACAGAGTTTCTC- 3'
O11 5' -GGATTTCCAGAAGAAGATGCC- 3'
O14 5' -GTTCGTTCGGGTTTGATTGTG- 3'
O15 5' -GAACTCAACGGGCCCTTCCCC- 3'
O16 5' -GGAGGGAAGTGGAGTAAATGGCCG- 3'

D. Relative abundance of *psb*L, E*aad*A and E*kan* mRNAs

Accumulation of partially edited *psb*L mRNA in the transgenic lines could be due to its competition with the chimeric E*aad*A or E*kan* mRNAs for a limiting common factor(s) that is required for editing. Therefore, the relative abundance of the *psb*L and chimeric transcripts was determined. It should be noted that both the polycistronic *psb*E (Carillo et al., 1986) and the E*aad*A and E*kan* mRNAs are approximately 1.1-kb in size. To quantify the accumulation of these transcripts, differential DNA probes on Northern blots were utilized. See Figure 5. Probing with the *psb*J coding sequence fragment indicated that the 1.1-kb *psb*E operon mRNA, which contains the *psb*J and *psb*L reading frames, accumulates to a similar extent in the wild-type and transformed plants. See Figure 5B, lower panel. The O14 oligonucleotide probe hybridizes to the mRNA containing the Δ*psb*F/Δ*psb*L region present in both the *psb*E operon and the chimeric E*aad*A and E*kan* transcripts. The O14 probe detected about 4x more RNA in the transgenic plants indicating a 1:3 ratio of the polycistronic *psb*E to chimeric mRNAs. See Figure 5B, upper panel. Procedures used for RNA gel analysis are discussed below.

Total RNA was extracted using TRIzol (Gibco BRL). RNA was electrophoresed in formamide-containing 1% agarose gel and transferred to nylon membrane (Amersham). Hybridization to ^{32}P-end-labeled

oligonucleotide probe 014 was carried out in 6 X SSPE, 0.5% SDS, 10x Dendardt's solution, 100 mg/ml tRNA, 0.1% Sodium Pyrophosphate at 45°C. Hybridization to random primed (Boehringer Mannheim) ^{32}P-labeled DNA fragment probes was carried out at 65°C in rapid hybridization buffer (Amersham). RNA levels in samples that hybridized to the probes were quantitated by PhosphorImager analysis (Molecular Dynamics).

E. Editing of other mRNAs is not affected in the transgenic plants.

Increased demand for *psb*L editing in the transgenic plants led to a reduction in its editing efficiency. Experiments were performed to determine if editing of other mRNAs is also affected in the transgenic plants. Two sites were tested in the *rpo*B, and two in the *ndh*B transcripts. See Table III.

TABLE III

List of tested editing sites in wild-type and transgenic plants

Editing site	Codon no.		Codon (amino acid)	
	Maize	Tobacco	Unedited	Edited
rpoB site I[a]	156	158	TCG (Ser) →	TTG (Leu)[c]
rpoB site II[a]	182	184	TCA (Ser) →	TTA (Leu)
*ndh*B site I[b]	156	156	CCA (Pro) →	CTA (Leu)
*ndh*B site II[b]	196	196	CAT (His) →	TAT (Tyr)

[a]Reference: Zeltz et al., 1993
[b]Reference: Maier et al., 1992
[c]In tobacco, a TCA codon is edited to a TTA codon.

The *rpo*B and *ndh*B editing sites were originally reported for maize, and confirmed for tobacco in this study. Editing sites I and II of *rpo*B are almost fully edited in wild-type tobacco, shown in Figure 6,

as has been observed for maize and barley (Zeltz et al., 1993). Similarly, sites I and II of the *ndh*B transcript are fully edited in wild type tobacco also shown in Figure 6, as reported for maize (Maier et al., 1992). The editing efficiency for the same sites was tested in three lines each of the E*aad*A- and E*kan*-expressing plants. No significant difference in the editing efficiency between wild-type and transformed plants was found for any of the four sites. Data in Figure 6 are shown for a Nt-pHC94-1 plant, one of the E*aad*A-expressing lines. Lack of change in the editing efficiency at any of the sites other than *psb*L indicates that expression of the chimeric genes specifically compromises the editing efficiency of the *psb*L site.

DISCUSSION

The above described examples are the first demonstration of the editing of chimeric mRNAs in plastids. Editing of both E*aad*A and E*kan* transcripts indicates that 98 and 101 nucleotides, respectively, of the Δ*psb*F/Δ*psb*L fragment are sufficient to direct editing at the *psb*L site. Accumulation of E*aad*A or E*kan* mRNA at levels approximately 3-fold above that of the *psb*E polycistronic message containing the *psb*L reading frame led to a significant (>10-fold) increase in the level of unedited *psb*L transcript. Increase in the level of unedited *psb*L mRNA from <1% to approximately 10% did not have any deleterious consequence that could have been detected at the phenotypic level. The chimeric mRNAs were also partially edited in the transgenic plants.

Partial editing of both *psb*L and chimeric mRNAs suggests depletion of a limiting trans-acting factor(s) that is required for editing of the shared site. However, the editing efficiency of four other sites was unaffected suggesting that the depleted

factor is specifically required for editing of the *psb*L transcript and is not a component of the general editing machinery. It is therefore conceivable that each of the editing sites in the chloroplast genome requires some factors for editing that are unique to them. This conclusion is reinforced by the lack of any obvious sequence motif common to the 98 nucleotide Δ*psb*F/Δ*psb*L fragment and sequences surrounding the other four tested editing sites. Therefore, it appears likely that the editing of these sites is directed by sequences and factors that are unique to each.

As an alternative to depletion of a site-specific factor, existence of "strong" and "weak" editing sites was also considered. Accordingly, the *psb*L site would be weak and its editing frequency would be lowered by the presence of excess chimeric RNA competing for a limiting but common editing factor, whereas the others would be strong sites that remain unaffected. This explanation is considered unlikely based on other data in the literature which are consistent with the existence of site-specific editing factors in plastids. The *psb*F mRNA is edited in spinach plastids by a C to U conversion, changing a serine to a conserved phenylalanine codon. In tobacco at this position a phenylalanine codon is already present at the DNA level. When the tobacco *psb*F gene was modified to match the spinach sequence, the heterologous editing site was unedited, although the adjacent *psb*L site is edited in both species (Bock et al., 1994). It appears therefore that tobacco lacks the capacity to edit the spinach *psb*F mRNA while maintaining the capacity to edit the *psb*L site which is common to both species. Another case consistent with site-specific editing is site IV of the *rpo*B mRNA which is edited in maize but not in barley,

although the sequences surrounding the site are highly conserved. Interestingly, the editing of three other sites in the same transcript is conserved between the two species (Zeltz et al., 1993). These observations suggest that the editing capacity of an individual site may be lacking without affecting the editing capacity of other sites, supporting site-specific editing in plastids.

EXAMPLE II

EDITING BASED ΔpsbL/kan AND Δ*ndh*D/kan SELECTABLE MARKER GENES

In plastids, editing of an ACG codon to an AUG codon creates the translation initiation codon for the *psb*l and *ndh*D transcripts in tobacco. To identify the RNA segment required for *psb*L editing, chimeric kanamycin resistance genes were constructed containing *psb*L deletion derivatives, and tested *in vivo* for editing in transgenic plants. The data demonstrate that a 22 nucleotide segment is sufficient to direct efficient *psb*L editing, including 16 nucleotides upstream and 5 nucleotides downstream of the editing site. Mutation of the A nucleotide to a C upstream of the editing site completely abolished editing, while mutation of the downstream G to a C only reduced the editing efficiency. Out of the 22 nt editing target sequence, the 16 upstream nucleotides were found to compete with the endogenous *psb*L transcript for a depletable *trans*-factor. To test whether editing of initiation codons involves a common *trans*-factor, a chimeric gene containing the *ndh*D editing site was expressed in tobacco plastids. As for *psb*L, editing of the *ndh*D site requires a depletable *trans*-factor. However, the *ndh*D *trans*-factor is distinct from that required for *psb*L editing. Distinct *cis-sequences* and *trans*-factor requirements for the *psb*L and *ndh*D editing sites indicates an individual recognition

mechanism for the editing of plastid initiation codons.

A. Defining the *cis*-sequences directing *psb*L editing

As mentioned previously, the psbL gene is part of the *psb*E operon which contains the *psb*E, *psb*F, *psb*L and *psbJ* reading frames (Carillo et al., 1986). In the earlier examples (Chaudhuri et al., 1995) the editing of the *psb*L translation initiation site in a chimeric mRNA containing a 98 nt Δ*psb*F/Δ*psb*L fragment was described (-63/+34 in plasmid pSC2, Figure 7A). In the chimeric construct of Figure 7A, the first open reading frame is a truncated *psb*F (Δ*psb*F) gene containing 40 nt of the C-terminus. The second open reading frame contained 36 nt of the N-terminus of *psb*L (Δ*psb*L) translationally fused with the bacterial kanamycin resistance (*kan*) gene to yield the Δ*psb*L/kan fusion protein. The two open reading frames are separated by 22 nt of intergenic region. See Figure 7A.

To identify the sequences required for *psb*L editing, deletion derivatives of the 98 nt Δ*psb*F/Δ*psb*L fragment were tested for editing *in vivo*. As before, the *psb*L deletion derivatives were fused N-terminally to bacterial kanamycin resistance gene (*kan*), and cloned in the plastid *Prrn*/T*rps*16 expression cassette to create chimeric genes. See Figure 7A. Thus, for all the constructs, translation of Δ*psb*L/kan was made dependent on editing of the *psb*L A<u>C</u>G codon to A<u>U</u>G codon. Editing therefore could be tested by the kanamycin resistance phenotype. The only exception is the chimeric gene containing -2/+34 fragment (in plasmid pSC10) where the initiation codon for the translation of Δ*psb*L/kan reading frame was provided by *Prrn*. The *psb*L deletion derivatives were introduced into the tobacco plastid genome by

linkage to a selectable spectinomcyin resistance gene (Chaudhuri et al., 1995).

The upstream deletion series included constructs with 5'-ends at positions -63, -51, -39, -27, -16, -10 and -2 nucleotides relative to the editing site (position 0). The downstream deletion series included constructs with 3'-ends at positions +34, +22, +10, +5 and +1 nucleotides relative to the editing site. The editing efficiency of the chimeric mRNAs was determined by direct sequencing and phosphoimager analysis of PCR-amplified cDNAs. Editing in the deletion derivatives was maintained as long as the constructs contained 16 nt of upstream and 5 nt of the downstream sequence relative to the editing site as shown in Figures 7A and B. Interestingly, in the deletion series, the % of the chimeric mRNA that is edited (editing efficiency) was either similar to that of the full size 98 nt Δ*psb*F/Δ*psb*L fragment (about 50%-70%), or barely detectable (~0%). Expression of kanamycin resistance was also a reliable qualitative marker of editing in all transformants in which translation of the chimeric mRNA was dependent on editing. See Figure 7A. The exception were plants obtained by transformation with plasmid pSC10 in which kanamycin resistance is expressed from the translation initiation codon contained in the *Prrn* promoter fragment. Construction of the deletion derivatives is set forth below.

The *psb*L deletion derivatives and the *ndh*D gene fragment were generated by PCR amplification with 5' primers carrying *Nco*I restriction site and 3' primers carrying NheI restriction site using total cellular DNA from tobacco (cv. Petit Havana). The following primer pairs were used: plasmid pSC2, 023 and 029; plasmid pSC3, 023 and 030; plasmid pSC4, 023 and 031; plasmid pSC5, 023 and 032; plasmid pSC6, 024 and 029;

plasmid pSC7, 025 and 029; plasmid pSC8, 026 and 029; plasmid pSC9, 027 and 029; plasmid pSC10, 028 and 029; plasmid pSC18, 027 and 031; plasmid pSC19, 027 and 034; plasmid pSC20, 033 and 031; pSC23, 037 and 038. The PCR products were digested with *Nco*I and NheI restriction enzymes.

To introduce suitable restriction sites at the 5'-end of the kan coding region, kan was PCR amplified from pTNH32 (Carrer et al., 1993) using 5' primer (021) carrying *Nco*I and NheI restriction sites in tandem and 3' primer (022) carrying *Xba*I restriction site. The PCR product was cloned in NcoI/*Xba*I digested pUC120 to generate plasmid pSC1.

The chimeric genes were constructed by N-terminal fusion of PCR amplified sequences from tobacco psbL and *ndh*D genes (*Nco*I/NheI fragments) to bacterial kan gene lacking the initiation codon (NheI/*Xba*I fragments). The chimeric genes were then cloned in NcoI/*Xba*I digested plasmid pLAA24A (Zoubenko et al., 1994). Plasmid pLAA24 is a derivative of plastid transformation vector pPRV111A, (Gene Bank Accession No. U12812) which has a selectable spectinomycin resistance gene, and a uidA reporter gene in the Prrn/T*rps*16 expression cassette (Zoubenko et al., 1994). The *Prrn* 5'-regulatory region consists of the plastid rRNA operon promoter and a ribosome binding site and is on an *Sac*I/*Nco*I fragment. The T*rps*16 fragment includes the *rps*16 gene 3'-regulatory region between nucleotides 5,087 to 4,939 in the ptDNA (Shinozaki et al., 1986) and is contained within an *Xba*I/*Hind*III fragment. Digestion of plasmid pLAA24A with *Nco*I/*Xba*I restriction enzymes removes the uidA coding region from the expression cassette, which is then replaced with the chimeric constructs, also an NcoI/*Xba*I fragment.

To construct the chimeric genes of the subsequent examples, the following procedures were

used. Plastid transformation and plant regeneration were performed as described in Example I. PCR amplification and DNA sequencing were also performed as described in Example I above. The sequencing gels
5 were subjected to phosphoimager analysis (Molecular Dynamics) for quantitation of editing efficiency. Radioactivity in bands corresponding to nucleotides was determined. The values were normalized for sample loading and labeling efficiency against other
10 bands in the same lanes. mRNA editing efficiency (%) = [corrected edited signal/(corrected edited + corrected unedited signal)] X 100. The primers used were as follows:

01: 5'-CAATATCAGCAATGCAGTTCATCC-3'
15 04: 5'-CACGACGAGATCCTCGCCG-3'
05: 5'-GAATAGCCTCTCCACCCA-3'
06: 5'-GGAATCCTTCCAGTAGTATCGGCC-3'
07: 5'-GGAAAATAAAACAGCAAGTAC-3'
017: 5'-AATTCGAAGCGCTTGGATACAGTTGTAGGGA-3'
20 018: 5'-GTAAGAGATGTGAATCCGCCTGT-3'
019: 5'-GCATAAGTCGTTAGAAGGAG-3'
020: 5'-GAAGAAAGAAAATTAAGGAACC-3'
021: 5'-CATGCCATGGCTAGCATTGAACAAGATGGATTGCACG-3'
022: 5'-GTACTCTAGACCCGCTCAGAAGAACTCG-3'
25 023: 5'-CTAGCCATGGCTTTGGGATCAATATCAGCAATG-3'
024: 5'-CTAGCCATGGCATCAGCAATGCAGTTCATCC-3'
025: 5'-CTAGCCATGGCGTTCATCCAACGATAAACTTAA-3'
026: 5'-CTAGCCATGGCATAAACTTAATCCGAATTATAGAG-3'
027: 5'-CTAGCCATGGCCGAATTATAGAGCTACGACAC-3'
30 028: 5'-CTAGCCATGGCTACGACACAATCAAACCCGA-3'
029: 5'-CTAGCTAGCTTCAACATTTTGTTCGTTCGG-3'
030: 5'-CTAGCTAGCTTCGTTCGGGTTTGATTGTG-3'
031: 5'-CTAGCTAGCTGATTGTGTCGTAGCTCTATA-3'
032: 5'-CTAGCTAGCCGTAGCTCTATAATTCGGATT-3'
35 033: 5'-CTAGCCATGGTATAGAGCTACGACAC-3'
034: 5'-CTAGCTAGCAAGTGTCGTAGCTCTATA-3'
035: 5'-AATTATAGAGCTCCGACACAATC-3'

036: 5'-AATTATAGAGCTACCACACAATC-3'

037: 5'-CTAGCCATGGTATTTTGAGCACGGGTTTTTCTGGTCC-3'

038: 5'-CTAGCTAGCTGGAAAAACTACAATTATTGTTAACC-3'

5 **B.** Mutation of the nucleotides <ins>flanking the *psb*L editing site.</ins>

The edited A<ins>C</ins>G codon to C<ins>C</ins>G and ACC in the efficiently edited 98 nt Δ*psb*F/Δ*psb*L fragment were
10 altered to address the following issues: (1) Whether the flanking nucleotides are critical for editing. (2) Whether the fidelity of editing the correct C is maintained when one of the flanking nucleotides is changed to a C. (3) Whether translation initiation at
15 this site is required for editing, since changing the ACG codon to CCG and ACC would eliminate the possibility of translation initiation at the edited codon.

Mutation of the upstream nucleotide (ACG to CCG;
20 NtpSC14 line) resulted in the loss of editing (~0%). See Figure 8. Mutation of the downstream nucleotide (ACG to ACC; Nt-pSC15 line) allowed editing at the correct C, but at a significantly reduced efficiency, ~20%. See Figure 8. The mutational analysis
25 therefore indicated that the A residue directly upstream of the edited C is appears to be essential for editing while mutation of the downstream G residue to C is compatible with editing but is required for optimal efficiency. In addition,
30 editing of the correct C in the mutated codon ACC points to a high fidelity mechanism of the editing apparatus in the choice of the editing site. Furthermore, editing of the ACC codon suggests that translation initiation at this codon is not required
35 for editing. Construction of the chimeric genes and introduction into plants was carried out as described in section A for the Δ*psb*L/*kan* derivatives.

The *psb*L derivatives with a point mutation were obtained by the megaprimer method of PCR (Sarkar and Sommer, 1990) using plasmid pSC2 as the template. These were also designed as *Nco*I and NheI fragments. The primers used were the following: plasmid pSC14, step I, 035 and 029, step II, 023; plasmid pSC15, step I, 036 and 029, step II, 023.

C. Identification of *psb*L mRNA sequences which interact with a *psb*L-specific editing factor (*psb*L-SEF).

In previous examples it has been shown that the editing efficiency of the endogenous *psb*L transcript is reduced in plastids expressing the chimeric *psb*L mRNA. Reduced editing efficiency was due to competition of the 98 nt Δ*psb*F/Δ*psb*L fragment with the endogenous *psb*L mRNA for a site-specific editing factor (*psb*L-SEF) present in limiting amounts (Chaudhuri et al., 1995).

Testing *psb*L editing efficiency in plastids expressing the chimeric Δ*psb*F/Δ*psb*L deletion derivatives, shown in Figure 7, was used to further define *psb*L sequences which interact with *psb*L-SEF. Out of the 22 nucleotides minimally required for editing, only the segment upstream of the editing site was able to compete with endogenous *psb*L for *psb*L-SEF. The 16 nt *psb*L-SEF binding site (boxed) within the 22 nt *psb*L editing recognition sequence is shown in Figure 9A. Sequences between nucleotides -16/-10 are critical for competition since competition is abolished in plastids containing the pSC20 construct which lacks this sequence. See Figure 9B. Interestingly, the plants expressing the pSC14 construct with the A to C mutation at position -1 also maintained competition, although this mutation completely abolished editing. The *psb*L editing efficiency data for the critical constructs

are shown in Figures 9A and 9B. While in the wildtype plants *psb*L mRNA is >99% edited, competition in the transgenic lines lead to accumulation of a significant amount of unedited *psb*L transcript.

D. Editing of the *ndh*D initiation codon in chimeric mRNA

Sequence analysis of *ndh*D and the corresponding cDNA by Neckermann et al (1994) has established that the *ndh*D translation initiation codon is created by editing of an ACG codon to an AUG codon in tobacco, spinach and snapdragon. The following experiments were designed to test whether *ndh*D editing requires a depletable *trans*-factor as found for *psb*L, and whether this *trans*-factor is utilized for the editing of both initiation codon sites. For this purpose, an 89 nucleotide fragment (-48/+40) spanning the *ndh*D editing site was translationally fused with the *kan* coding region and cloned in a *Prrn*/T*rps*16 expression cassette. See Figure 10 A and B. The chimeric gene were constructed by N-terminal fusion of PCR amplified sequences from the tobacco *ndh*D gene (*Nco*I/NheI fragments) to bacterial *kan* gene lacking the initiation codon (NheI/*Xba*I fragments). For a detailed description of the construction of the chimeric gene see Example II, section A. The chimeric gene was introduced into the tobacco plastid genome by linkage to a spectinomycin resistance gene. In the chimeric gene, expression of the Δ*ndh*D/kan fusion protein was dependent on the editing of the *ndh*D site. To prevent translation from an upstream AUG, a point mutation was introduced 26 nt upstream of the editing site changing an A to a C, underlined in Figure 10 A.

Nt-pSC23 plants expressing the Δ*ndh*D/kan protein were resistant to kanamycin indicating editing of the *ndh*D site. Direct sequencing of PCR amplified

ΔndhD/kan revealed a very low efficiency (~7%) of editing, shown in Figure 10C. The *ndh*D transcript in the wild-type plants is edited at significantly higher efficiency (~45 %), which is reduced in the Nt-pSC23 plants to ~ 20 %. See Figure 10C. The reduction in the editing efficiency of the endogenous *ndh*D transcript in the transgenic plants indicates that increasing the demand for *ndh*D editing leads to the depletion of an editing factor which is present in limiting amounts. However, the efficiency of editing of the *psb*L transcript in the transgenic Nt-pSC23 plants was comparable to the wild-type levels, >99 %, shown in Figure 10C. Since *psb*L editing in the Nt-pSC23 plants is unaffected, the depleted editing factor is *ndh*D-specific, and is not required for *psb*L editing.

*ndh*D editing in plants expressing the chimeric kan gene fused with the 98 nt Δ*psb*F/Δ*psb*L fragment, Nt-pSC2, Figure 7, was also examined. In such plants reduced editing of the endogenous *psb*L mRNA due to competition for *psb*L-SEF has been shown (Chaudhuri et al., 1995; Figure 10C). However, in the same plants the endogenous *ndh*D editing is unaffected, see Figure 10C, indicating that *psb*L-SEF is not involved in editing the *ndh*D site.

DISCUSSION

The above examples describe the analysis of the *cis*-element requirements for mRNA editing in plastids. The data show that the C to U conversion in the *psb*L mRNA is directed by a 22 nucleotide sequence which encompasses 16 nucleotides upstream and 5 nucleotides downstream of the edited C at position 0. The 22 nt sequence is conserved in tobacco, spinach (Kudla et al., 1992) and bell pepper (Kuntz et al., 1992), species in which editing of the *psb*L translation initiation codon has been reported.

The role of nucleotides directly flanking the editing site was tested by mutating them in the 98 nt Δ*psb*F/Δ*psb*L fragment which is efficiently edited. Changing the upstream A at -1 to a C completely abolished editing of the correct C. However, changing the G at +1 to a C allowed editing of the correct C, although at a reduced efficiency. Editing of the correct C in the mutated ACC codon indicates the high fidelity of nucleotide selection for editing. This is consistent with the observation that specific C nucleotides are edited within flanking C sequences (Kossel et al., 1993; Maier et al., 1995). Furthermore, editing of the ACC codon suggests that translation initiation is not required for editing to occur, providing direct evidence for the lack of linkage between translation and editing. This finding is consistent with mRNA editing in plastids lacking ribosomes (Zeltz et al., 1993) and with editing of unspliced plastid mRNAs which are not translatable (Freyer et al., 1993).

The *psb*L translation initiation codon is only one of the approximately 25 editing sites found in the plastids of higher plants (Maier et al., 1995). Further studies will be required to determine how typical is the close proximity of cis-sequences to the editing sites in plastids found for *psb*L. In this regard, the *ndh*D initiation codon appears to be similar since all information required for editing is contained in a relatively small (98 nt) RNA segment. However, editing of sites II and III in the tobacco *ndh*B gene (Maier et al., 1992) requires sequences further away than 150 nucleotides (S.C. and P.M., unpublished). Therefore, localization of editing cis sequences is not uniform, in line with the proposed individual recognition mechanism for each of the ~25 plastid editing sites.

Individual recognition of the editing sites is consistent with the finding that site-specific trans-factors are depleted by over-expression of the *psb*L and *ndh*D target RNAs. While ACG to AUG editing in both transcripts creates a translation initiation codon, over-expression of either of the target RNAs affects the editing efficiency of only the source mRNA.

It is likely that C to U editing in plastids involves cytidine deamination, as shown for plant mitochondria (Yu and Schuster, 1995). Editing therefore minimally involves either a single polypeptide containing both a site-specific recognition domain and a deaminase domain, or a complex containing at least two components, one providing site-specific recognition and the other with cytidine deaminase activity. Such a multi-component complex consisting of cytidine deaminase (APOBEC-1) and auxiliary proteins has been shown to be involved in C to U editing of the mammalian nuclear apolipoprotein B mRNA. In addition to the common occurrence of C to U editing, close clustering of the cis-sequences around the editing site is an additional feature shared by the plastid *psb*L and the mammalian nuclear apolipoprotein B editing systems. Editing of apolipoprotein B is directed by an 11 nucleotide recognition sequence located four nucleotides downstream of the editing site. In addition, sequences upstream are required for efficient editing (reviewed in Innerarity et al., 1996). However, in contrast to editing of *psb*L, recognition specificity of the apolipoprotein B editing process is relaxed, since cytosines introduced adjacent to the edited nucleotide may also be modified (Chen et al., 1990).

EXAMPLE III

Editing based Δ*rpl2/kan* selectable marker gene

The chimeric Δ*rpl2/kan* gene was constructed by N-terminal fusion of PCR amplified sequences from the maize rpl2 gene (*Nco*I/*Nhe*I fragment; 5'primer: 5'-CTAGCCATGGAAACGAACTAAAGGAGAATAC-3'; 3' primer: 5'-CTAGCTAGCCGGGATAGGTGTTTTGTATAAA-3') to a bacterial *kan* gene lacking the initiation codon (*Nhe*I/*Xba*I fragments). See Figure 11A and 11B. The chimeric genes were then cloned in *Nco*I/*Xba*I digested plasmid pLAA24A (Zoubenko et al., 1994), as described for the construction of Δ*psb*L/*kan* genes and introduced into the tobacco plastid genome (Chaudhuri and Maliga, 1996). The chimeric mRNA was transcribed in tobacco plastids. In tobacco, no editing of the maize *rpl2* translation initiation codon was found. Also, the transformed plants were sensitive to kanamycin. However, editing of this chimeric gene will occur in rice, maize and other cereals in which the *rpl2* translation initiation codon is created by editing.

EXAMPLE IV

CONVERSION OF INTERNAL EDITING SITES TO EDITED TRANSLATION INITIATION CODONS

The translation initiation codon is created by conversion of an ACG codon to an AUG codon in the *psb*L, *ndh*D and *rpl2* plastid mRNAs. The *psb*L (Kudla et al., 1992) and *ndh*D (Neckermann et al., 1994) editing sites are present in a few but not all dicotyledonous species, whereas the *rpl2* site is edited in most but not all cereals (Hoch et al., 1991). The maize *rpl2* site is not edited in tobacco (Chauduri and Maliga, see Example III). Therefore, the *psb*L, and *ndh*D editing sites are useful to create

editing-based marker genes in some dicots, and *rpl2*-based chimeric genes are useful in most monocots.

There are many more examples for the editing of internal codons than for editing of translation initiation codons. However, translation initiation is not required for editing of codons in the *psb*L sequence context (Chaudhuri and Maliga, 1996). Based on these results, internal codons may also serve as translation initiation codons as long as editing creates a translatable mRNA. There is a high frequency of Ser to Phe, Ser to Leu and Pro to Leu transitions, and a lower extent of Thr to (F)Met transitions. Given that U and A are relatively frequent at the first nucleotide position, editing of UCG codons will be maintained in most editing contexts even if the first nucleotide is changed to A to create a codon which may be edited to a translation initiation codon by C to U conversion. Good candidates for such mutagenesis are editing Sites I and II of the rpoB mRNA, which are widely edited in both dicots and monocots (Zeltz et al., 1993).

RNA sequences required to direct editing may be contained within a short segment adjacent to the editing site as in the case of the *psb*L gene or may be at a distance as in case of Sites II and III of the *ndh*B gene (Chaudhuri and Maliga, 1996). Editing has been tested in *ndh*B and *rpo*B minigenes to identify editing sites that are useful for the construction of chimeric genes.

*ndh*B and rpoB editing sites have been identified for which the relevant cis-sequences are within a short segment. These short gene segments have been incorporated in chimeric genes, expressed in tobacco plastids, and tested for editing by direct sequencing of the PCR-amplified transgene cDNAs. The editing

sites in the source genes are listed in Table IV. The map of the *ndh*B and *rpo*B minigene derivatives is shown in Figure 12.

TABLE IV

RNA editing in *ndh*B and rpoB minigenes.

Editing site	Codon No	Codon (amino acid) Unedited / Edited	Reference
*rpo*B site I	158	[a]TCG (Ser) to TTG (Leu)	Zeltz et al., 1993
*rpo*B site II	184	TCA (Ser) to TTA (Leu)	Zeltz et al., 1993
*ndh*B site I	156	CCA (Pro) to CTA (Leu)	Maier et al., 1992
*ndh*B site II	196	CAT (His) to TAT (Tyr)	Maier et al., 1992
*ndh*B site III	204	TCA (Ser) to TTA (Leu)	Maier et al., 1992
*ndh*B site IV	246	CCA (Pro) to CTA (Leu)	Maier et al., 1992
*ndh*B site X	249	TCT (Ser) to TTT (Phe)	Kossel, H. (personal comm.)

[a]In tobacco, a TCA codon is edited to a TTA codon (Chaudhuri et al., 1995).

The *ndh*B minigene contains an *ndh*B fragment which is 369 nucleotide in size (between nucleotides 143,174 and 144,042 of the plastid genome, Shinozaki et al., 1986). It contains some of the first exonic sequence including five editing sites, named as sites I, II, III, IV and X. Sites I, II, III and IV are edited in maize, rice and tobacco (Maier et al., 1992). Therefore, marker genes based on the editing of these sites will be useful in a broad range of crops, including monocots and dicots. Site X is edited in tobacco only (Kossel, H., personal communication), therefore this site is less useful for the construction of chimeric marker genes.

In the *ndh*B minigene, the truncated coding region is expressed in the original reading frame, in the *Prrn*-*Trps*16 cassette. This was achieved by introducing an *Nco*I site at the 5'-end of the truncated reading frame, which includes the translation initiation codon (CCATGG) from which the minigene RNA can be translated from translation signals contained in the cassette. The minigene

contains the DNA sequence ATGGCAGCTACT downstream of the translation initiation codon; nucleotide C at position five corresponds to nucleotide 143,674 in the plastid genome. In addition, during PCR amplification, an in-frame stop codon was introduced at the 3'-end of the truncated coding region. (5' PCR primer: 5'-CTAGCCATGGCAGCTACTCTAGGGGGAATG-3'; 3' PCR primer: 5'-CTAGTCTAGACGTATACGTCAGGAGTCCA-3'). The minigene was physically linked to a selectable spectinomcyin resistance (*aad*A) gene in a suitable plastid targeting vector and the vector DNA was introduced into tobacco leaf chloroplasts by the biolistic process. Transplastomes with the integrated, linked transgenes were selectively amplified by incubating the bombarded leaf segments on a spectinomycin medium, on which transgenic shoots were directly regenerated. The protocols for plastid transformation have been described in Svab and Maliga, 1993 and Zoubenko et al., 1994.

Out of the five sites, Sites I, IV and X were highly edited in the minigene. This finding indicates, that the cis sequences required for editing are located relatively close to the editing sites, as was shown for the *psb*L translation initiation codon (Chaudhuri and Maliga, 1996). Furthermore, cis sequences for Sites I and IV are suitable for inclusion in marker genes with utility in both dicots and monocots, since the capacity for editing is present in these widely divergent taxonomic groups. Interestingly, *ndh*B Sites II and III were not edited in the minigene, indicating that the cis sequences required for editing are further away than +/- 150 nucleotides from the editing site. Therefore, cis sequences required for editing are not uniformly positioned relative to the editing site.

The *rpo*B minigene contains a 281 nucleotide fragment of the rpoB gene, encoding the RNA

polymerase β-subunit. The fragment contains two editing sites (I, II, see Table IV and Fig. 12; based on Zeltz et al., 1993). Both editing sites are present in maize, rice, barley, spinach and tobacco (Maier et al., 1992). The rpoB minigene was constructed and introduced into plastids as described for the *ndh*B minigene. The minigene contains the DNA sequence ATGGTCCCGGT downstream of the translation initiation codon; nucleotide G at position four corresponds to the complement of nucleotide 27120 in the plastid genome (Zhinozaki et al., 1986). The rpoB fragment for the minigene construction was obtained by PCR amplification (5' PCR primers: 5'-CTAGCCATGGGTCCCGGTATTTATTACCG-3'; 3' PCR primer: 5'-CTAGGTCGACTTAGGCATTTTCTTTTGACCCAAT-3'). Transgenic plants representing several independently transformed lines were obtained and assayed for editing. Complete editing of both of the sites was found in the minigenes by sequencing PCR-amplified cDNAs. Given the presence of these sites in both monocots and dicots, marker genes based on the editing of either of these sites could be used in a wide variety of crops.

EXAMPLE V

RNA EDITING FOR TISSUE-SPECIFIC REGULATION OF FOREIGN GENE EXPRESSION

RNA editing in plastids was discussed assuming that editing is constitutive, and facilitates expression of marker genes in all tissue types. It is known, however, that environmental and developmental conditions significantly affect editing efficiency (Bock et al., 1993; Hirose et al., 1996). Tissue specific differences in editing efficiency facilitate

the design of chimeric genes the translation of which is dependent on tissue type due to tissue-specific conversion of ACG codons to a translation initiation codon. Such chimeric genes are useful when accumulation of an economically useful protein, such as an insecticidal endotoxin is desired only in specific tissue types, such as leaves, root hairs, root cortex or epidermis cells.

Alternatively, desired tissue-specific expression of economically useful genes may be obtained when editing tissue specifically creates a translation termination (stop) codon. Formation of stop codons by editing may be readily obtained by engineering in plastids. For example, a stop codon in plastids is created when changing the reading frame of the psbL editing site. Normally, the *psb*L translation initiation codon is created by C to U conversion in the ACGA sequence. Moving the reading frame by one nucleotide, editing creates the TGA stop codon. Editing of the first C nucleotide of a codon is also known, such as Site II of the *ndh*B transcript (Maier et al., 1992). Therefore, C to U editing of the CAA codon will create the TAA translation termination codon.

Most plastid genes are organized in polycistronic transcription units. Therefore, polycistronic transcription units may be built for simultaneous expression of multiple proteins. An example for a dicistronic transcription unit, form which two proteins are simultaneously expressed was obtained by engineering of the plastid genome (Staub and Maliga, 1995). Tissue-specific expression of such polycistronic mRNAs may be obtained by making the translation dependent on RNA editing, either through creation of a translation initiation codon, or by terminating translation.

REFERENCES

1. Araya, A., Domec, C., Begu, D., and Litvak, J. (1992) an *in vitro* system for editing of ATP synthase subunit mRNA using wheat mitochondrial extracts. *Proc. Natl. Acad. Sci. USA*, 89, 1040-1044.

2. Benne, R. (1994) RNA editing in trypanosomes. *Eur. J. Biochem.* 221, 9-23.

3. Bachmann,B., Lueke,W. and Hunsmann,G. (1990) Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Res.*, 18, 1309.

4. Bock,R,, Hagemann,R., Kossel,H and Kudla,J (1993) Tissue- and stage-specific modulation of RNA editing of the *psb*F and *psb*L transcript from spinach plastids - a new regulatory mechanism? *Mol. Gen. Genet.*, 240, 238-244.

5. Bock,R., and Maliga,P. (1995) In vivo testing of a tobacco plastid DNA segment for guide RNA function in *psb*L editing. *Mol. Gen. Genet.*, in press.

6. Bock,R., Kossel,H. and Maliga,P. (1994) Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. *EMBO J.*, 13, 4623-4628.

7. Bonnard, G., Gualberto, J.M. (1992) RNA editing in plant mitochondria. *Critical Reviews in Plant Sciences*, **10**, 503-524.

8. Carillo,N., Seyer,P., Tyagi,A. and Herrmann,R.G. (1986) Cytochrome b-559 genes from Oenothera hookeri and Nicotiana tobacco show a remarkably high degree of conservation as compared to spinach. The enigma of cytochrome b-559: Highly conserved genes and proteins but no known function. *Curr. Genet.*, 10, 619-624.

9. Carrer,H., Hockenberry,T.N., Svab,Z. and Maliga,P. (1993) Kanamycin resistance as a selectable marker for plastid transformation in tobacco. *Mol. Gen. Genet.*, 241, 49-56.

10. Carrer,H., Staub, J.M. and Maliga,P. (1991) Gentamicin resistance in Nicotiana conferred by AAC(3)-I, a narrow substrate specificity acetyltransferase. *Plant Mol. Biol.*, 17, 301-303.

11. Chan,L. (1993) RNA editing: exploring one mode with apolipoprotein B mRNA. *BioEssays* 15, 33-41.

12. Chaudhuri, S. and Maliga, P. (1996) Sequences directing to C to U editing of the plastid *psbL* mRNA are located within a 22 nucleotide segment spanning the editing site. *EMBO J.*, accepted, pending on revisions.

13. Chaudhuri, S., Carrer, H. and Maliga, P. (1995) Site-specific factor involved in the editing of the *psbL* mRNA in tobacco plastids. *EMBO J.* 14, 2951-2957.

14. Cornelissen M, Vandewiele M. (1989) Nuclear transcriptional activity of the tobacco plastid *psbA* promoter. *Nucleic Acids Res.* 17:19-29.

15. Freyer, R., Hock, B., Neckermann, K., Maier, R.M. and Kössel, H. (1993) RNA editing in maize chloroplasts is a processing step independent of splicing and cleavage to monocistronic mRNAs. *The Plant J.*, 4, 621-629.

16. Gray,M.W. and Covello,P.S. (1993) RNA editing in plant mitochondria and chloroplasts. *FASEB J.*, 7, 64-71.

17. Hanson, M.R., Sutton, C.A. and Lu,B. (1995) Plant organelle gene expression: altered by RNA editing. *Trends in plant Sciences*, **1**, 57-64.

18. Higuchi,M., Single,F.N., Kohler,M., Sommer,B., Sprengel,R. and Seeburg, P.H. (1993) RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell*, 75, 1361-1370.

19. Hirose, T, Fan, H, Suzuki, Y.Y., Wakasugi, T., Tsudzuki, T., Kossel, H. and Sugiura, M. (1996) Occurrence of silent RNA editing in chloroplasts: its species specificity and the influence of environmental and developmental conditions. *Plant Mol. Biol.* 30:667-672.

20. Hock, B., Maier, R.M., Appel, K., Igloi, G.L. and Kössel, H. (1991) Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353: 178-180.

21. Hodges,P. and Scott,J. (1993) Editing of mammalian apolipoprotein B mRNA by site-specific RNA deamination. In Benne,R. (ed.), *RNA editing. The alteration of protein coding sequences of RNA.* Ellis Horwood, New York, NY., pp. 125-151.

22. Innerarity, T.L., Boren, J., Yamanaka, S. and Olofsson, S.O. (1996) Biosynthesis of apolipoprotein

B48-containing lipoptotiens. *J. Biol. Chem.*, 271, 2353-2356.

5 23. Koncz, C.S., Martini, N., Mayerhofer, R., Konca-Kalman, Z.S., Korber H., Redei, G.Y., Schell, J., (1989) High-frequence T-DNA-mediated gene tagging in plants. *Proc. Natl Acad. Sci. USA* 86:8467-8471.

10 24. Kossel,H., Hoch,B., Maier,R.M., Igloi,G.L., Kudla,J., Zeltz,P., Freyer,R., Neckermann,K. and Ruf,S. (1993) RNA editing in chloroplasts. In Kueck, U. and Brennicke,A. (eds.), *Plant Mitochondria*. VCH Publishers, Weinheim, Germany, pp. 93-102.

15 25. Kudla,J., Igloi,G.L., Metzlaff,M., Hagemann,R. and Kossel, H. (1992) RNA editing in tobacco chloroplasts leads to the formation of a translatable *psb*L messenger RNA by a C-to-U substitution within the initiation codon. *EMBO J.*, 11, 1099-1103.

20 26. Kuntz, M., Camara, B., Weil, J.H. and Schantz, R. (1992) the *psbL* gene from bell pepper (*Capsicum annuum*): plastid RNA editing occurs in non-photosynthetic chromoplasts. *Plant. Mol. Biol*. 20, 25 1185-1188.

27. Maier,R.M., Neckermann,K., Hoch,B., Akhmedov,N.B. and Kossel,H (1992) Identification of 30 editing positions in the *ndh*B transcript from maize chloroplasts reveals sequence similarities between editing sites of chloroplasts and plant mitochondria. *Nucl. Acids Res.*, 20, 6189-6194.

35 28. Maier, R.M., Neckermann, K., Igloi, G.L. and Kossel, H. (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol*. 251, 614-628.

40 29. Maliga,P (1995) Biolistic transformation of tobacco cells with nuclear drug resistance genes. In Maliga,P., Klessig,D., Cashmore,A., Gruissem,W. and Varner,J. (eds), Methods in Plant Molecular Biology - 45 A Laboratory Manual. Cold Spring Harbor Press, pp. 73-54.

30. Maliga,P. (1993) Towards plastid transformation in flowering plants. *Trends Biotechnol.*, 11, 101-106.

50 31. Mettler,I.J. (1987) A simple and rapid method for minipreparation of DNA from tissue cultured plant cells. *Plant Mol. Biol. Rep.*, 5, 346-349.

32. Murashige,T. and Skoog,F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol. Plant.*, 15, 473-497.

33. Neckermann, K., Zeltz, P., Igloi, G.L., Kössel, H., and Maier, R.M. (1994) The role of RNA editing in convervation of start codons in chloroplast genomes. *Gene*, 146, 177-182.

34. Sarkar, G. and Sommer, S.S. (1990) The "megaprimer" method of site-directed mutagensis. *Biotechniques* **8**, 404-407.

35. Schuster,W. and Brennicke,A. (1994) The plant mitochondrial genome: physical structure, information content, RNA editing, and gene migration. *Annu. Rev. Plant Physiol. Plant Mol. Biol*, 45, 61-78.

36. Schuster, W. and Brennicke, A. (1994) The plant mitochondrial genome: physical structure, information content, RNA editing, and gene migration. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **45**, 61-78.

37. Scott, J. (1995) A place in the world for RNA editing. *Cell,* **81**, 833-836.

38. Shinozaki,K. et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and gene expression. *EMBO J.*, 5, 2043-2049.

39. Simpson,L., Maslov,D.A. and Blum,B. (1993) RNA editing in Leishmania mitochondria. In Benne,R. (ed.), RNA editing. The alteration of protein coding sequences of RNA. Ellis Horwood, New York, NY., pp. 53-85.

40. Staub,J. and Maliga,P. (1994) Translation of *psb*A mRNA is regulated by light via the 5'-untranslated region in tobacco plastids. *Plant J.*, 6, 547-553.

41. Sutton, C.A., Zoubenko, O.V., Hanson, M.R. and Maliga, P. (1995) A plant mitochondrial sequence transcribed in transgenic tobacco chloroplasts is not edited. *Mol. Cell. Biol.* **15**, 1377-1381.

42. Svab,Z. and Maliga,P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aad*A gene. *Proc. Natl. Acad. Sci. USA,* 90, 913-917.

43. Teng,B.B., Burant,C.F., Davidson, N.O. (1993) Molecular cloning of the apolipoprotein B messenger RNA editing protein. *Science*, 260, 1816-1819.

44. Yu, W. and Schuster, W. (1995) Evidence for a site-specific cytidine deamination reaction involved in C to U editing of plant mitochondria. *J. Biol. Chem.,* 270, 18227-18233.

5

45. Zeltz,P., Hess,W.R., Neckermann,K., Boerner,T. and Kossel,H. (1993) Editing of the chloroplast *rpo*B transcript is independent of chloroplast translation and shows different patterns in barley and maize.
10 *EMBO J.*, 12, 4291-4296.

46. Zoubenko,O.V., Allison,L.A., Svab,Z. and Maliga,P. (1994) Efficient targeting of foreign genes into the tobacco plastid genome. *Nucleic Acids Res.*,
15 22, 3819-3824.

While certain preferred embodiments of the
20 present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present
25 invention, as set forth in the following claims.

What is claimed is:

1. A recombinant chimeric DNA construct useful for selection of plastid transformants comprising an edited plastid gene segment translationally fused to the coding region of a selectable marker gene, said selectable marker gene being expressible following RNA editing of said plastid gene segment.

2. A vector comprising the construct of claim 1, containing homologous DNA sequences necessary for plastid directed transformation.

3. The DNA construct of claim 1, wherein said edited plastid gene segment is selected from the group consisting of *psb*L, *ndh*D, *rpo*B, *ndh*B or *rpl*2.

4. The DNA construct of claim 1, wherein said selectable marker gene is selected from the group consisting of the *kan* gene, the *aaaD* gene, or any other selectable marker gene as set forth in Table I.

5. The chimeric DNA construct of claim 1, which includes Δ*psb*L/*kan*.

6. The chimeric DNA construct of claim 1, which includes Δ*psb*L/*aad*A.

7. The chimeric DNA construct of claim 1, which includes Δ*ndh*D/*kan*.

8. The chimeric DNA construct of claim 1, which includes Δ*rpl2*/*kan*.

9. A chimeric DNA construct comprising an edited segment selected from *ndh*B sites I, IV or V cloned upstream from a selectable marker gene.

10. A chimeric DNA construct comprising an edited segment selected from rpoB sites I or II cloned upstream from a selectable marker gene.

5

11. A method for selection of transplastomic lines comprising:

a) transforming plastids in a sample with a DNA construct comprising an edited gene segment translationally fused to the coding region of a
10 selectable marker gene, said selectable marker gene being expressible following RNA editing of said gene segment;

b) culturing said samples in medium containing a selective agent facilitating
15 identification of transformed plastids;

c) selecting and propagating cells expressing said selectable marker; and

d) regenerating a plant from said cells expressing said seletable marker.

20

12. The method of claim 11, wherein said chimeric DNA construct is incorporated into a vector containing homologous DNA sequences necessary for plastid directed transformation.

25

13. The method of claim 12, wherein said vector further comprises a foreign gene of interest to beneficially augment the phenotype of said regenerated plant.

30

14. The method of claim 13, wherein said edited segment is edited in a tissue specific manner such that said foreign gene of interest is expressed in a tissue specific manner.

35

15. The method of claim 13, for use in monocots or dicots.

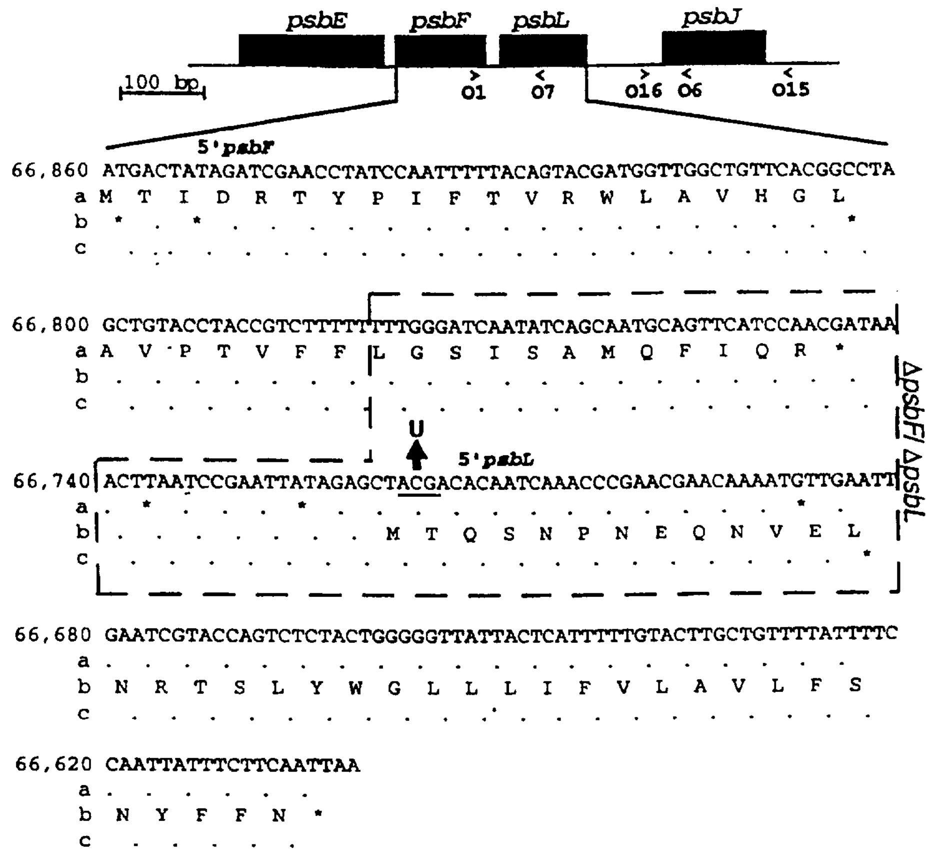


Figure 1

Fig. 2A

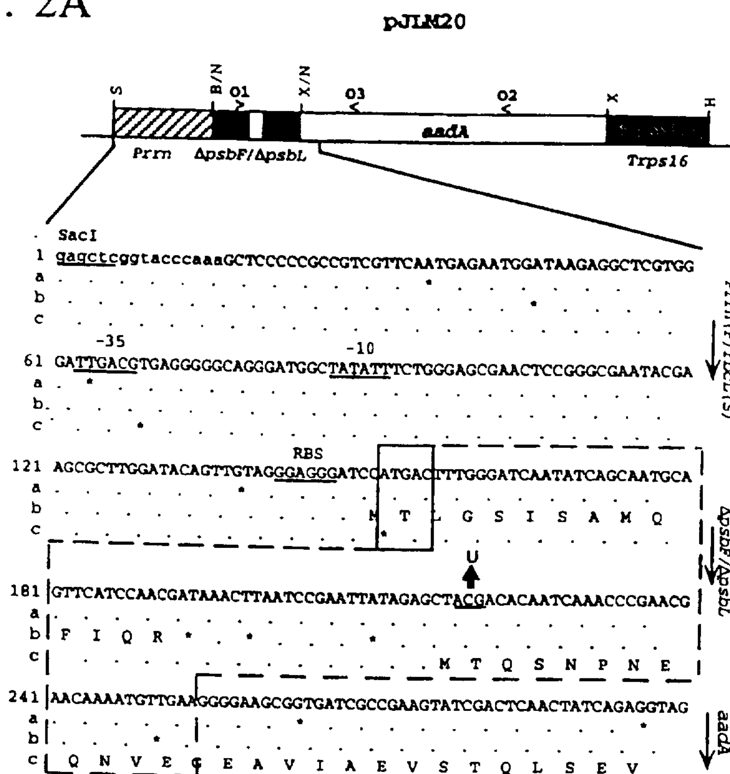


Fig. 2B

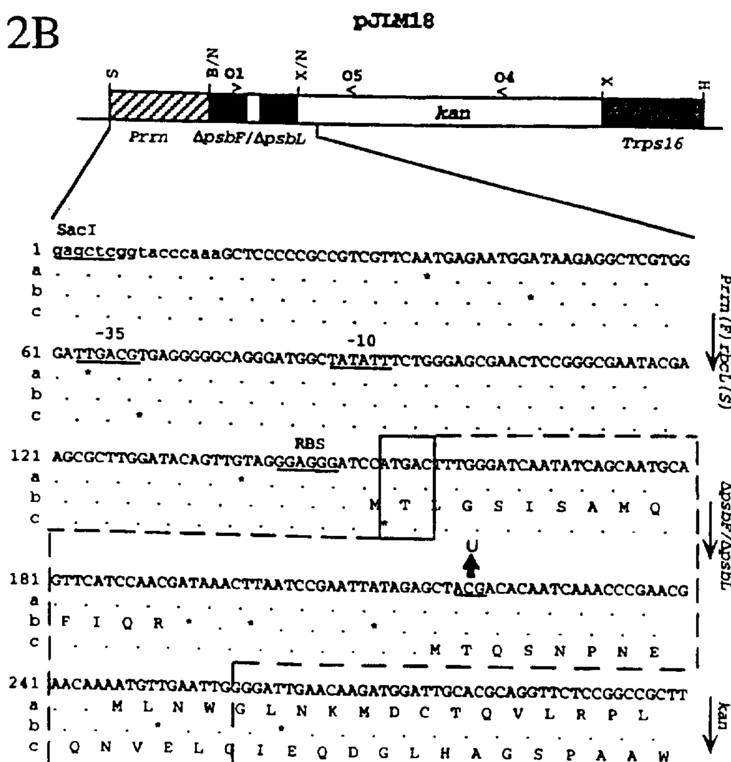


Fig. 3A

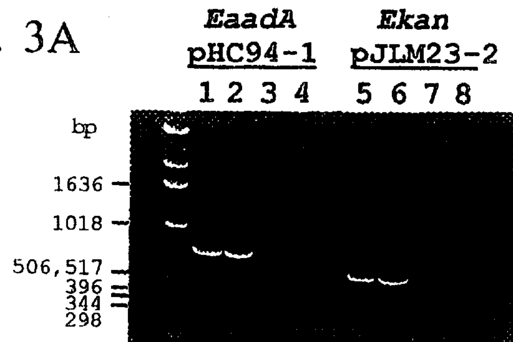


Fig. 3B

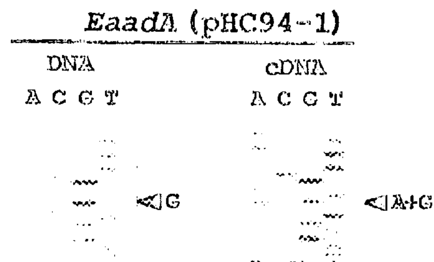


Fig. 3C

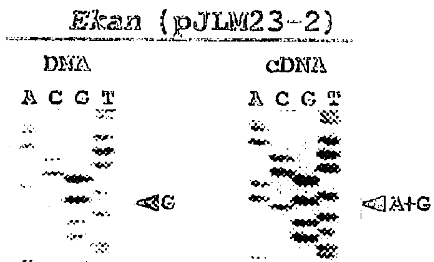
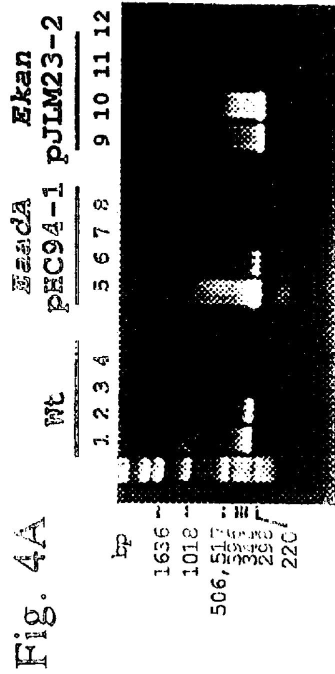


Fig. 4A



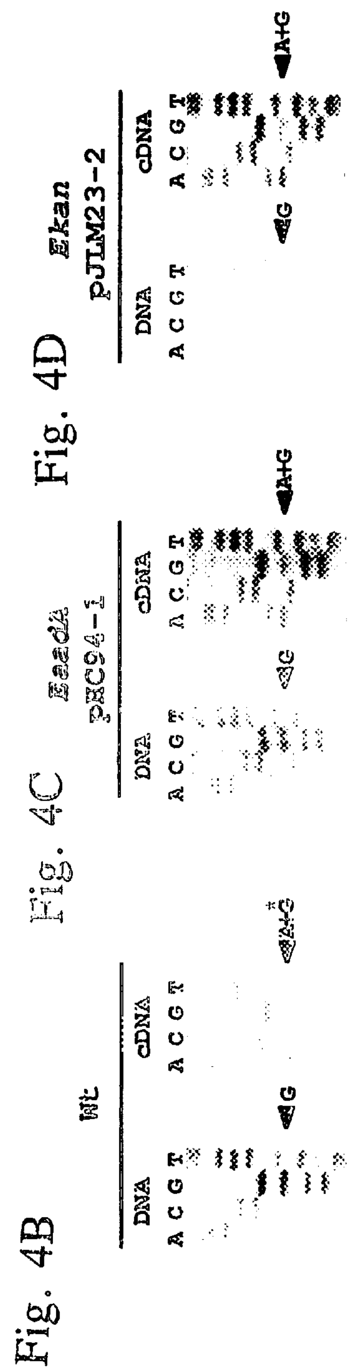


Fig. 4B

Fig. 4C

Fig. 4D

5/12

Fig. 5A

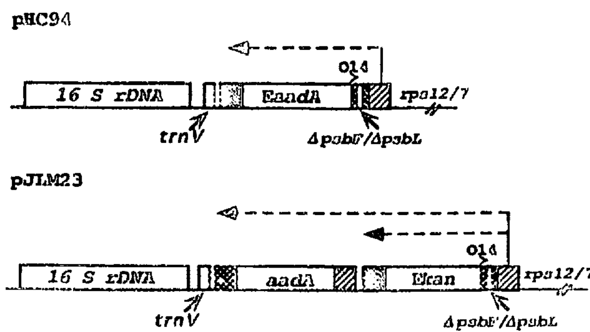
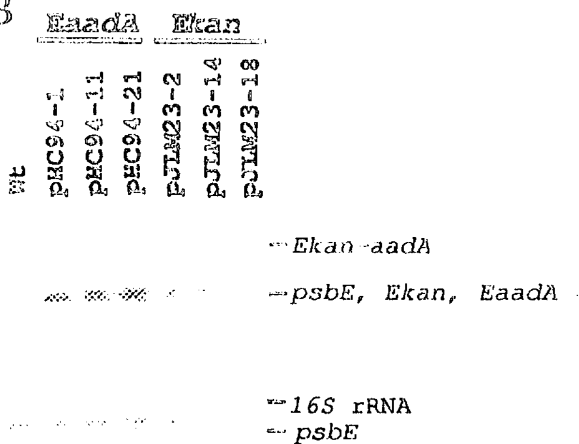


Fig. 5B



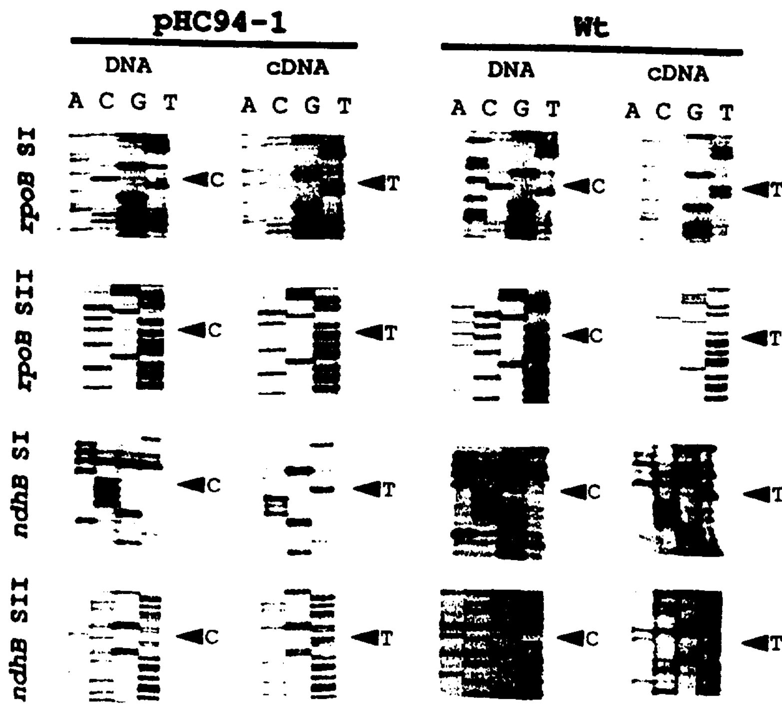


Figure 6

Fig. 7A

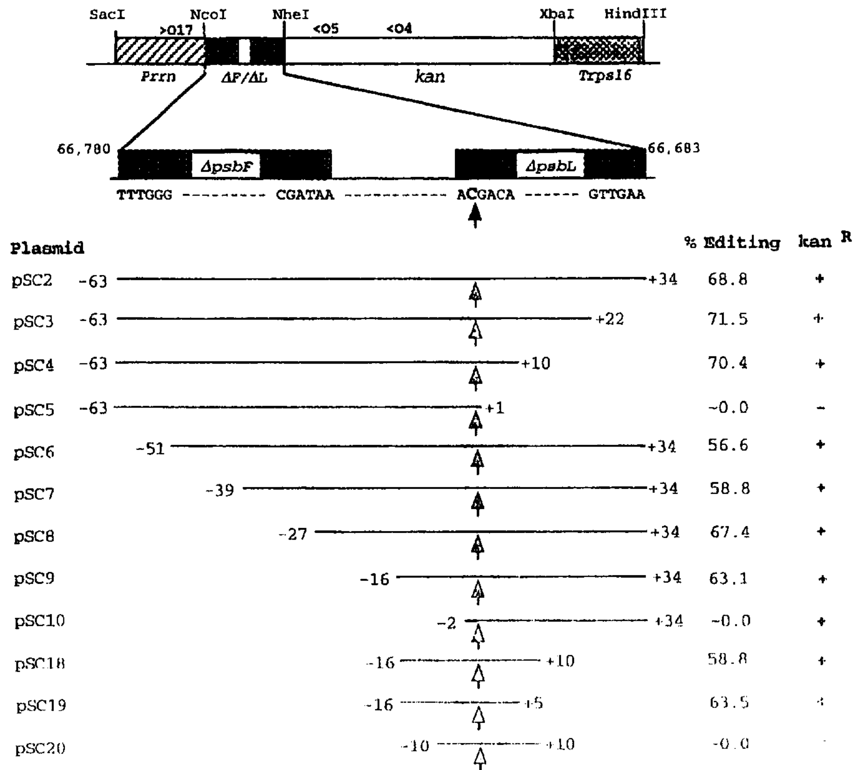
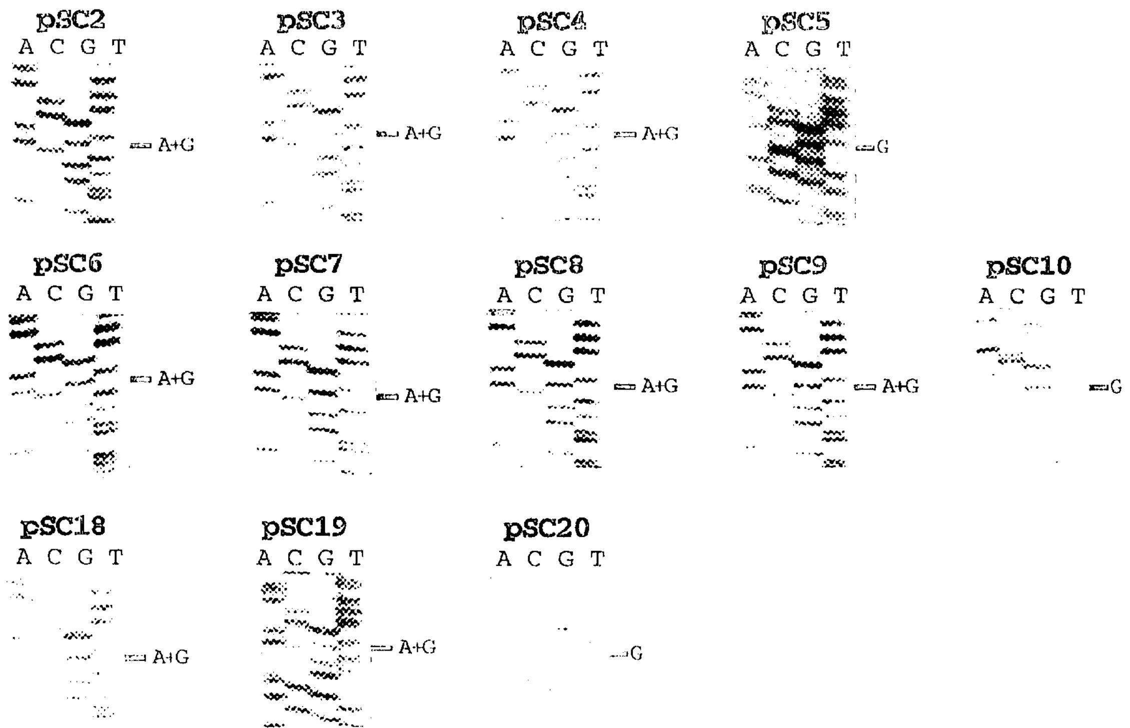
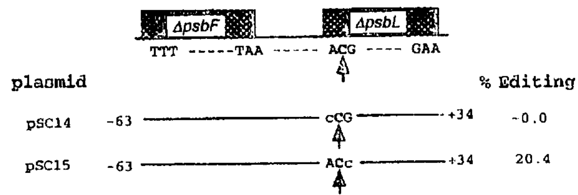


Fig. 7B





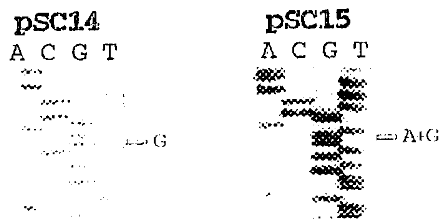


Figure 8

Fig. 9A

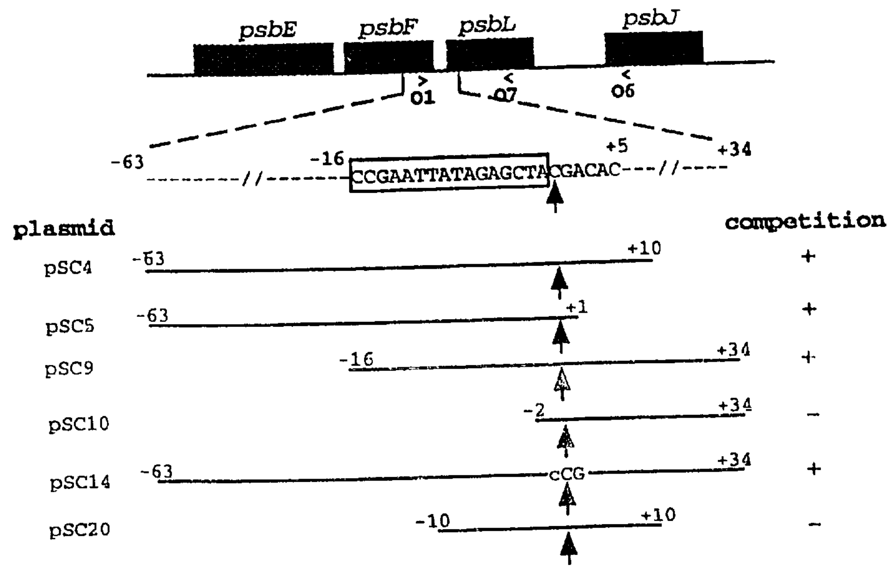


Fig. 9B

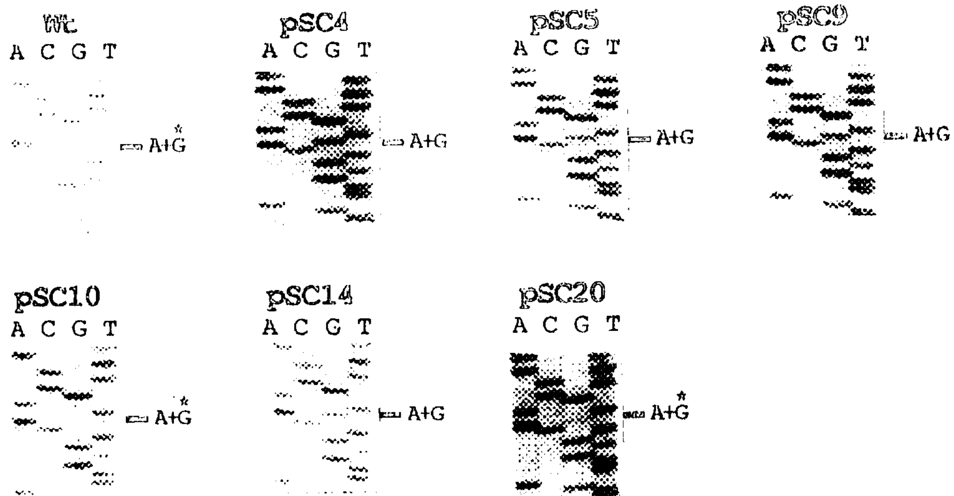


Fig. 10A

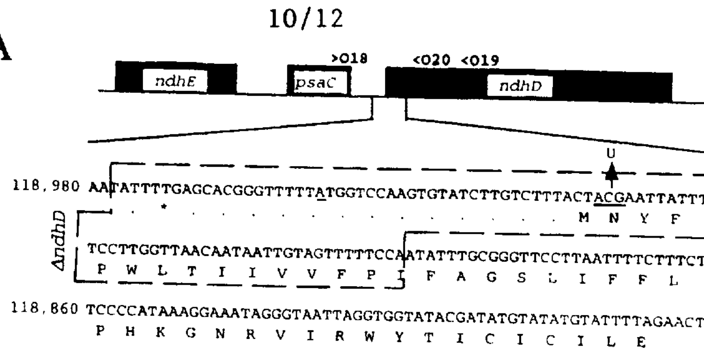


Fig. 10B

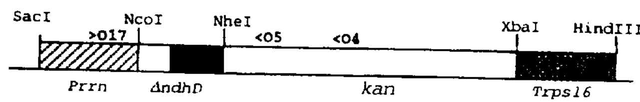


Fig. 10C

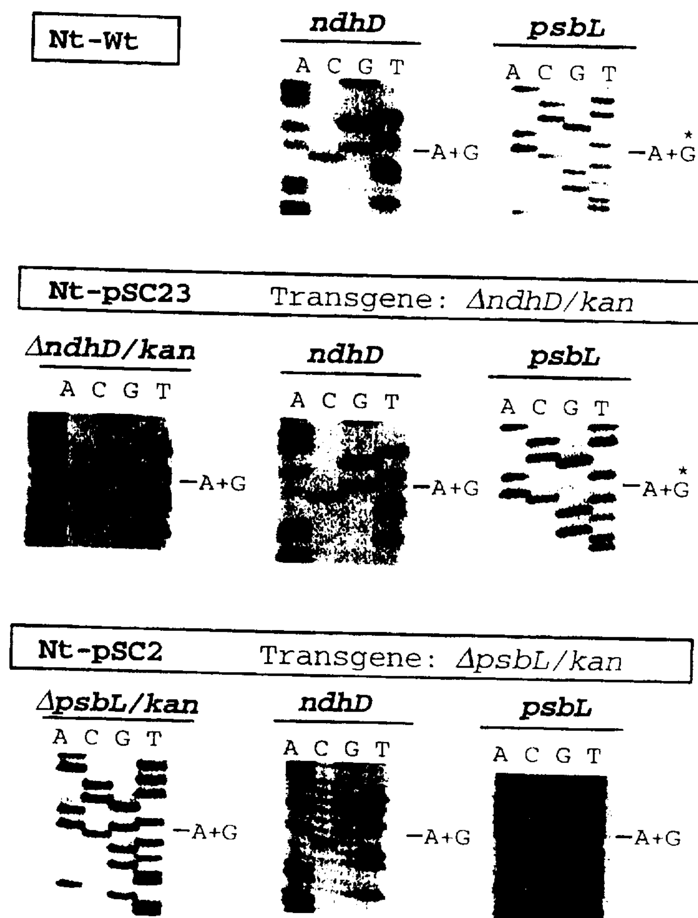


Fig. 11 A

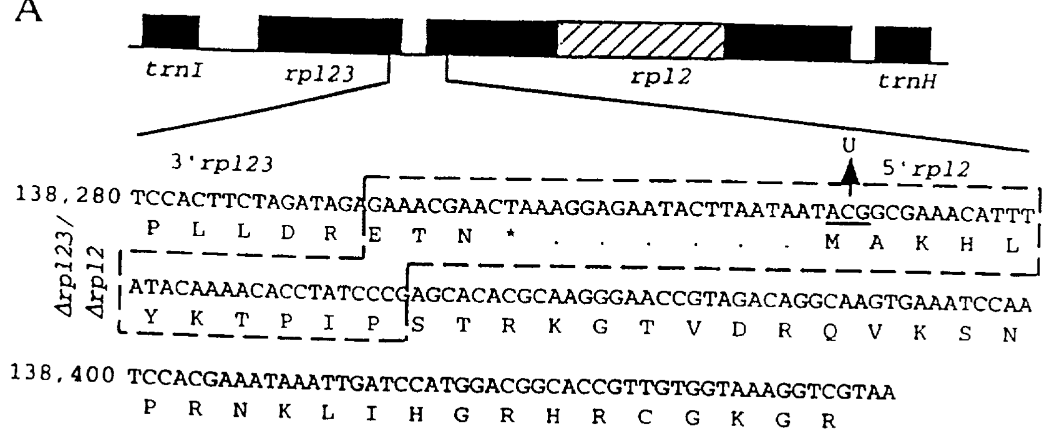
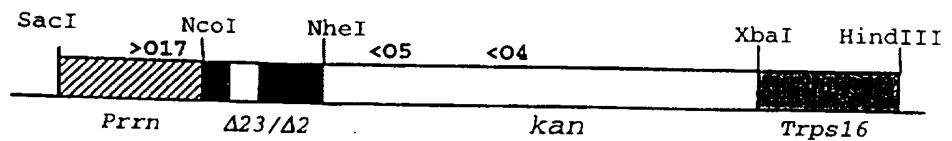


Fig. 11 B



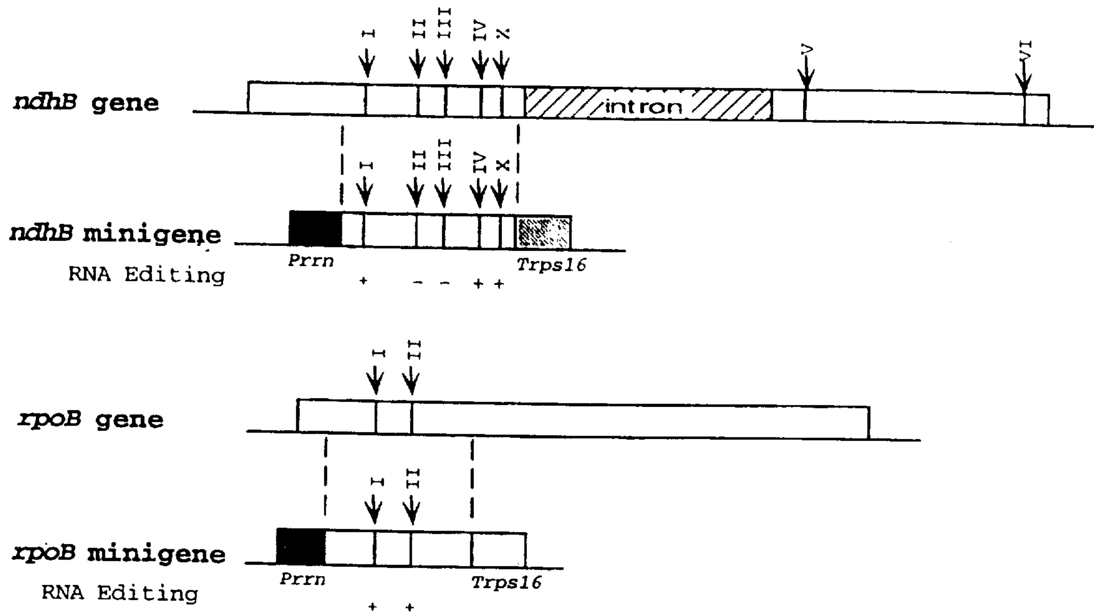
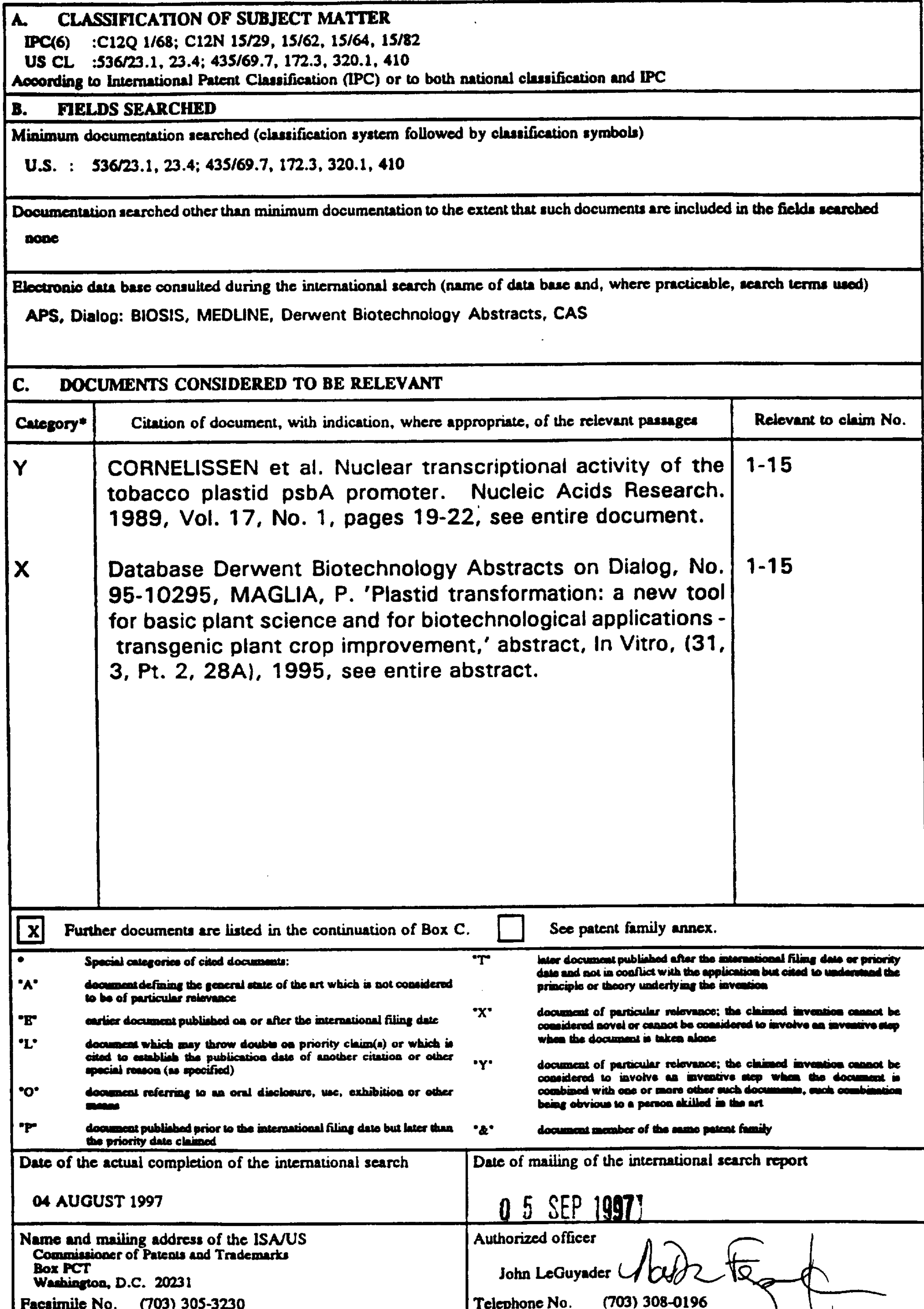


Figure 12



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/10318

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C12N 15/29, 15/62, 15/64, 15/82
US CL :536/23.1, 23.4; 435/69.7, 172.3, 320.1, 410
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 23.4; 435/69.7, 172.3, 320.1, 410

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

none

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Dialog: BIOSIS, MEDLINE, Derwent Biotechnology Abstracts, CAS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CORNELISSEN et al. Nuclear transcriptional activity of the tobacco plastid psbA promoter. Nucleic Acids Research. 1989, Vol. 17, No. 1, pages 19-22, see entire document.	1-15
X	Database Derwent Biotechnology Abstracts on Dialog, No. 95-10295, MAGLIA, P. 'Plastid transformation: a new tool for basic plant science and for biotechnological applications - transgenic plant crop improvement,' abstract, In Vitro, (31, 3, Pt. 2, 28A), 1995, see entire abstract.	1-15

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 AUGUST 1997	05 SEP 1997

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	John LeGuyader
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/10318

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database Derwent Biotechnology Abstracts on Dialog, No. 95-11453, CARRER et al. 'Targeted insertion of foreign genes into the tobacco plastid genome without physical linkage to the selectable marker gene - leaf chloroplast particle bombardment with plasmid pZS205 and plasmid pHC64 on tungsten microprojectile; transgenic plant construction with streptomycin- and kanamycin-resistance,' abstract, Bio/Technology (13, 8, 791-94), 1995, see entire abstract.	1-15