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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 March 2010 (11.03.2010)

PCT

(10) International Publication Number
WO 2010/028331 A1

(51) International Patent Classification:
C12N 15/82 (2006.01)

(21) International Application Number:
PCT/US2009/056165

(22) International Filing Date:
8 September 2009 (08.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/095,134 8 September 2008 (08.09.2008) US

(71) Applicant *(for all designated States except US)*: **ATHENIX CORPORATION** [US/US]; P.O. Box 110347, Research Triangle Park, NC 27709 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HAMMER, Philip, E.** [US/US]; 215 Covington Square Drive, Cary, NC 27513 (US). **BEILINSON, Vadim** [EE/US]; 4029 Remington Oak Circle, Cary, NC 27519 (US). **HINSON, Todd, K.** [US/US]; 3510 Moriah Road, Rougemont, NC 27572 (US).

(74) Agents: **DAVENPORT, Destiny, M.** et al.; Alston & Bird LLP, Bank of America Plaza, 101 South Tryon Street, Suite 4000, Charlotte, NC 28280-4000 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2010/028331 A1

(54) **Title**: COMPOSITIONS AND METHODS FOR EXPRESSION OF A HETEROLOGOUS NUCLEOTIDE SEQUENCE IN PLANTS

(57) **Abstract**: The present invention provides compositions and methods for targeting polypeptides to the chloroplasts of higher plants. Compositions include expression cassettes having a nucleotide sequence encoding a chloroplast targeting peptide (CTP) operably linked to a nucleotide sequence of interest, wherein the CTP is derived from Chlamydomonas sp. Plant transformation vectors, plants and plant cells having the CTP sequences are also encompassed, as well as variants and fragments of the CTP sequences. Methods for expressing a heterologous nucleotide sequence in a plant using the CTP sequences disclosed herein are also provided.

COMPOSITIONS AND METHODS FOR EXPRESSION OF A HETEROLOGOUS NUCLEOTIDE SEQUENCE IN PLANTS

FIELD OF THE INVENTION

The present invention relates to the field of plant molecular biology, more particularly to the identification and use of regulatory elements in plants.

BACKGROUND OF THE INVENTION

Chloroplast biogenesis in plants is dependent upon the coordinated activities of two independent genetic systems localized in the chloroplast and the nucleus (see Cline and Henry (1996), Annu. Rev. Cell Dev. Biol. 12, 1-26). The vast constituent chloroplast proteins are encoded by the nuclear genes and are synthesized cytoplasmically- as precursor forms which contain N-terminal extensions known as transit peptides. The transit peptide is instrumental for specific recognition of the chloroplast surface and in mediating the post-translational translocation of pre-proteins across the chloroplast envelope and thence to the various different subcompartments within the chloroplast (e.g. stroma, thylakoid and thylakoid membrane).

Genes reported to have naturally encoded transit peptide sequences at their N-terminus include the chloroplast small subunit of ribulose-1,5-bisphosphate carboxylase (RuBisCo), de Castro Silva Filho et al. (1996) Plant Mol. Biol. 30: 769-780; Schnell, D. J. et al. (1991) J. Biol. Chem. 266 (5): 3335-3342; 5-(enolpyruvyl) shikimate-3-phosphate synthase (EPSPS), Archer et al. (1990) J. Bioenerg. and Biomemb. 22 (6):789-810; tryptophan synthase. Zhao, J. et al. (1995) J. Biol. Chem. 2 70 (11):6081-6087; plastocyanin, Lawrence et al. (1997) J. Biol. Chem. 272 (33):20357-20363; chorismate synthase, Schmidt et al. (1993) J. Biol. Chem. 268 (36):27477-27457; and the light harvesting chlorophyll a/b binding protein (LHBP), Lamppa et al. (1988) J. Biol. Chem. 263: 14996-14999, although not all of these sequences have been useful in the heterologous expression of chloroplast-targeted proteins in higher plants.

SUMMARY OF INVENTION

Compositions and methods for chloroplast targeting of polypeptides in a plant are provided. Compositions comprise expression cassettes comprising a nucleotide sequence encoding a chloroplast targeting peptide (or chloroplast transit peptide, "CTP") sequence derived from an algal organism operably linked to nucleotide sequence of interest. These expression constructs are useful for expression and proper targeting of the nucleotide sequence of interest in a monocot or a dicot plant. The invention further provides vectors comprising the expression cassettes, and plants and plant cells having stably incorporated or transiently expressed into their genomes an expression cassette described above. Additionally, compositions include transgenic seed of such plants.

Methods are also provided for expressing a nucleotide sequence in a plant or plant cell, as well as methods for identifying algal CTP sequences for use in a plant.

DESCRIPTION OF FIGURES

Figure 1 demonstrates TagGFP expression in tobacco protoplasts.

Figure 2 demonstrates expression and processing of *Chlamydomonas* EPSPS chloroplast transit peptide in maize cells. Lane 1: Non-transgenic maize line Hi-II; Lanes 2 – 6: Individual T_o events transformed with *Chlamydomonas* EPSPS CTP/GRG23(ace3)(R173K) construct; Lane 7: Protein molecular weight marker; Lane 8: Purified GRG23(ace3)(R173K) protein (4 ng).

Figure 3 demonstrates the calculation of the molecular weight of the processed *Chlamydomonas* EPSPS – GRG-23(ace3)(R173K) protein expressed in maize. A linear regression of the plot of Log Molecular Weight vs. Distance Migration of the protein molecular weight standards from Figure 2 was used to calculate the apparent molecular weight of the GRG23(ace3)(R173K) protein standard and the processed *Chlamydomonas* EPSPS – GRG-23(ace3)(R173K) detected in plant extract.

DETAILED DESCRIPTION

In the production of transgenic plants it is often useful to direct foreign proteins to specific subcellular locations, e.g., the chloroplast, vacuole, mitochondria, or ER. Previous workers have fused DNA sequences encoding transit peptides from various plant genes to the genes of interest. When the gene is translated the resulting

protein has the plant transit peptide fused to the amino terminus of the protein of interest, and thus the protein is directed, with varying efficiency, to the desired subcellular compartment.

Thus, the present invention is drawn to compositions and methods for chloroplast targeting of polypeptides in higher plants or plant cells. The compositions of the present invention comprise expression cassettes comprising a nucleotide sequence encoding a chloroplast transit peptide (CTP) derived from an algal organism operably linked to a nucleotide sequence of interest. In one embodiment, the CTP is derived from *Chlamydomonas sp.* In another embodiment, the CTP comprises the amino acid sequence set forth in SEQ ID NO:3, 5, or 7 or an amino acid sequence encoded by SEQ ID NO:1, 2, 4, or 6, as well as variants, fragments, and derivatives thereof. In addition, transformed plants, plant cells, and seeds are provided.

The CTP-encoding sequences of the invention, when assembled within a DNA construct such that the CTP-encoding sequence is operably linked to a nucleotide sequence of interest, facilitate co-translational or post-translational transport of the peptide of interest to the chloroplast of a plant cell stably transformed with this DNA construct. Methods for expressing a nucleotide sequence in a plant comprise introducing into plant cells an expression cassette comprising a CTP-encoding nucleotide sequence of the invention operably-linked to a nucleotide sequence of interest, and regenerating a transformed plant from the plant cell.

The articles "a" and "an" are used herein to refer to one or more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one or more elements.

As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

Chloroplast transit peptides

Chloroplasts are organelles found in plant cells and eukaryotic algae that conduct photosynthesis. The chloroplast is a complex cellular organelle composed of three membranes: the inner envelope membrane, the outer envelope membrane, and the thylakoid membrane. The membranes together enclose three aqueous

compartments termed the intermediate space, the stroma, and the thylakoid lumen. While chloroplasts contain their own circular genome, many constituent chloroplast proteins are encoded by the nuclear genes and are cytoplasmically-synthesized as precursor forms which contain N-terminal extensions known as chloroplast transit peptides (CTPs). The CTP is instrumental for specific recognition of the chloroplast surface and in mediating the post-translational translocation of pre-proteins across the chloroplast envelope and into the various different subcompartments within the chloroplast (e.g. stroma, thylakoid and thylakoid membrane).

At least two distinct functional domains have been identified in chloroplast transit peptides: the stromal targeting domain (STD) and the lumen targeting domain (LTD). STDs govern access to the general import pathway and are both necessary and sufficient for import of the passenger protein to the stroma. Stromal protein precursors possess transit peptides that contain only an STD, whereas thylakoid lumenal protein precursors have both an STD and an LTD.

STDs range in length from about 30 to 120 residues and are rich in hydroxylated residues and deficient in acidic residues. They tend to share several compositional motifs: an amino terminal 10-15 residues devoid of Gly, Pro and charged residues; a variable middle region rich in Ser, Thr, Lys and Arg; and a carboxy-proximal region with loosely conserved sequence (Ile/Val-X-Ala/Cys-Ala; SEQ ID NO:17) for proteolytic processing. However, there are no extensive blocks of sequence conservation, nor any conserved secondary structural motifs. Theoretical analyses suggest that STDs adopt predominantly random coil conformations.

Genes reported to have naturally encoded transit peptide sequences at their N-terminus include the chloroplast small subunit of ribulose-1,5-bisphosphate carboxylase (RuBisCo), de Castro Silva Filho et al. (1996) Plant Mol. Biol. 30: 769-780; Schnell, D. J. et al. (1991) J. Biol. Chem. 266 (5): 3335-3342; 5-(enolpyruvyl) shikimate-3-phosphate synthase (EPSPS), Archer et al. (1990) J. Bioenerg. and Biomemb. 22 (6):789-810; tryptophan synthase. Zhao, J. et al. (1995) J. Biol. Chem. 2 70 (11):6081-6087; plastocyanin, Lawrence et al. (1997) J. Biol. Chem. 272 (33):20357-20363; chorismate synthase, Schmidt et al. (1993) J. Biol. Chem. 268 (36):27477-27457; and the light harvesting chlorophyll a/b binding protein (LHBP), Lamppa et al. (1988) J. Biol. Chem. 263: 14996-14999. Although several CTPS have been described, only a few have been utilized successfully in attempts to target chimeric molecules to chloroplasts in higher plants.

The present invention discloses the use of CTPs derived from algal species, particularly *Chlamydomonas sp.*, in higher plants. For the purposes of the present invention, "higher plants" are considered members of the subkingdom Embryophytae. In one embodiment, the CTP useful in the methods and compositions disclosed herein is derived from *Chlamydomonas*. In another embodiment, the CTP is set forth in SEQ ID NO:3, 5, or 7, or is encoded by SEQ ID NO:1, 2, 4, or 6, including variants, fragments, and derivatives thereof. However, one of skill in the art would understand how to identify chloroplast transit peptides other than the ones disclosed herein. For example, a number of CTPs (or protein sequences comprising CTPs) are listed in GENBANK®.

The CTPs disclosed herein are useful for targeting a polypeptide to the chloroplast of a plant cell. In one embodiment, the CTPs disclosed herein provide improved translocation compared to CTPs derived from, for example, higher plant organisms. The CTPs disclosed herein may result in an at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100%, or greater, or at least about 2-fold, at least about 3-fold, at least about 4-fold, or greater improvement in translocation of the polypeptide into the chloroplast when compared to a reference CTP. An improvement can be measured in terms of the amount of polypeptide that gets translocated into the chloroplast, the amount of active polypeptide that gets translocated into the chloroplast, or both. An improvement can also be measured in terms of an improvement in the phenotype of an organism transformed with the chloroplast-targeted protein of interest. For example, where the CTP of the invention is used to target an herbicide resistance protein to the chloroplast of the plant, an improvement in activity can be measured in terms of an improvement in herbicide resistance.

Expression cassettes

The CTP-encoding sequences of the invention may be provided in an expression cassette that allows it to drive expression and localization of a polypeptide encoded by the nucleotide sequence of interest into the chloroplast of plant cells. By "expression cassette" is intended a DNA construct that is capable of resulting in the expression of a protein from an open reading frame in a cell. The cassette will include in the 5′-3′ direction of transcription, a transcriptional initiation region

preferably comprising a promoter suitable for expression in a plant cell of interest, operably-linked to a CTP-encoding sequence of the invention, which is further operably linked to a nucleotide sequence of interest, and a translational and transcriptional termination region (i.e., termination region) functional in plants. The CTP-encoding nucleotide sequence and the nucleotide sequence of interest may be separated from one another by nucleotide sequences encoding one or more "linker" amino acids as discussed elsewhere herein.

The cassette may additionally contain at least one additional gene to be cotransformed into the organism, such as a selectable marker gene. Alternatively, the additional gene(s) can be provided on multiple expression cassettes. Such an expression cassette is provided with a plurality of restriction sites for insertion of the nucleotide sequence of interest to be under the transcriptional regulation of the regulatory regions.

The expression cassette may further comprise 3′ and/or 5′ untranslated region(s). By "3′ untranslated region" is intended a nucleotide sequence located downstream of a coding sequence. Polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting the addition of polyadenylic acid tracts to the 3′ end of the mRNA precursor are 3′ untranslated regions. By "5′ untranslated region" is intended a nucleotide sequence located upstream of a coding sequence. Other upstream or downstream untranslated elements include enhancers. Enhancers are nucleotide sequences that act to increase the expression of a promoter region. Enhancers are well known in the art and include, but are not limited to, the SV40 enhancer region and the 35S enhancer element.

The termination region may be native with the CTP-encoding nucleotide sequence of the present invention, may be native with the nucleotide sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acid Res.* 15:9627-9639.

The expression cassettes described herein may further comprise one or more regulatory elements other than CTP, as well as additional CTPs known in the art. By

"regulatory element" or "regulatory region" is intended a portion of nucleic acid found upstream or downstream of a gene, that may be comprised of either DNA or RNA, or both DNA and RNA and that is involved in gene expression. Regulatory elements may be capable of mediating organ specificity, or controlling developmental or temporal gene activation and include promoter elements, core promoter elements, elements that are inducible in response to an external stimulus, elements that are activated constitutively, transcriptional terminators, polyadenylation signals, and elements that decrease or increase promoter activity such as negative regulatory elements or transcriptional enhancers, respectively. By "cis-acting" is intended a sequence that is physically contiguous with the transcribed sequence. Cis-acting sequences typically interact with proteins or other molecules to carry out (turn on/off, regulate, modulate, etc.) transcription. By "transcriptional enhancer" is intended a nucleic acid sequence that, when positioned proximate to a promoter and present in a transcription medium capable of supporting transcription, confers increased transcription activity compared to that resulting from the promoter in the absence of the enhancer. Enhancers may function upstream, within, or downstream of a gene, even as far away as 50 kilobases from the transcriptional initiation site. Enhancers may also function independently of their orientation. By "transcriptional terminator" is intended a DNA sequence that includes a nucleotide base pair sequence necessary for reducing or eliminating transcription. By "polyadenylation signal" is intended a sequence that controls the termination of transcription and translation.

In one aspect of the invention, synthetic DNA sequences are designed for a given polypeptide, such as the chloroplast-targeted polypeptides useful in the methods disclosed herein. Expression of the open reading frame of the synthetic DNA sequence in a cell results in production of the polypeptide. Synthetic DNA sequences can be useful to simply remove unwanted restriction endonuclease recognition sites, to facilitate DNA cloning strategies, to alter or remove any potential codon bias, to alter or improve GC content, to remove or alter alternate reading frames, and/or to alter or remove intron/exon splice recognition sites, polyadenylation sites, Shine-Delgarno sequences, unwanted promoter elements and the like that may be present in a native DNA sequence. It is also possible that synthetic DNA sequences may be utilized to introduce other improvements to a DNA sequence, such as introduction of an intron sequence, creation of a DNA sequence that in expressed as a protein fusion

to organelle targeting sequences, such as chloroplast transit peptides, apoplast/vacuolar targeting peptides, or peptide sequences that result in retention of the resulting peptide in the endoplasmic reticulum. Synthetic genes can also be synthesized using host cell-preferred codons for improved expression, or may be synthesized using codons at a host-preferred codon usage frequency. See, for example, Campbell and Gowri (1990) *Plant Physiol.* 92:1-11; U.S. Patent Nos. 6,320,100; 6,075,185; 5,380,831; and 5,436,391, U.S. Published Application Nos. 20040005600 and 20010003849, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

The nucleic acids of interest to be targeted to the chloroplast may also be optimized for expression in the chloroplast to account for differences in codon usage between the plant nucleus and this organelle. In this manner, the nucleic acids of interest may be synthesized using chloroplast-preferred codons. See, for example, U.S. Patent No. 5,380,831, herein incorporated by reference.

Variants and fragments

Nucleic acid molecules that are fragments of the disclosed CTP sequences are also encompassed by the present invention. By "fragment" is intended a portion of the CTP sequence. A fragment of a nucleotide sequence may be biologically active and hence be capable of facilitating the translocation of a polypeptide of interest into the chloroplast of a plant, or it may be a fragment that can be used as a hybridization probe or PCR primer using methods disclosed below. Assays to determine whether such fragments have CTP activity are well known in the art.

Nucleic acid molecules that are fragments of a CTP-encoding nucleotide sequence disclosed herein may comprise at least about 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, contiguous nucleotides, or up to the number of nucleotides present in a full-length CTP sequence disclosed herein (for example, 306 nucleotides for SEQ ID NO:1) depending upon the intended use. By "contiguous" nucleotides is intended nucleic acid residues that are immediately adjacent to one another. Biologically active fragments of the CTP-encoding sequences of the present invention will encode a CTP that retains activity. By "retains CTP activity" is intended that the fragment will direct the translocation into the chloroplast of at least about 30%, at least about 50%, at least about 70%, or at least about 80% of the polypeptide encoded by the nucleotide sequence of interest. In one embodiment, a fragment of a CTP-encoding

nucleotide sequence disclosed herein may comprise one or more deletions of SEQ ID NO:1, 2, 4, or 6, including up to about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 12, about 15, about 18, about 21, about 24, about 27, about 30 or more deletions. In another embodiment, a fragment of a CTP-encoding nucleotide sequence disclosed herein may encode an amino acid comprising one or more deletions of SEQ ID NO:3, 5, or 7, including up to about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10 or more amino acid deletions.

A biologically active portion of a CTP can be prepared by isolating a portion of one of the CTP sequences of the invention and assessing the activity of that portion of the CTP. Methods for measuring CTP activity are well known in the art. See the section entitled "Evaluation of CTP Activity" for examples of suitable methods.

Variants of the CTP-encoding nucleotide sequences or the CTP amino acid sequences disclosed herein are also encompassed. By "variant" is intended a sufficiently identical sequence, or a sequence that differs by at least one amino acid from a native chloroplast transit peptide. CTP-encoding sequences encompassed by the present invention are sufficiently identical to the nucleotide sequence of SEQ ID NO:1, 2, 4, or 6. CTP sequences encompassed herein are sufficiently identical to the amino acid sequence of SEQ ID NO:3, 5, or 7. By "sufficiently identical" is intended a nucleotide sequence that has at least about 70% or 75%, about 80% or 85% sequence identity, about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity compared to a reference sequence using one of the alignment programs as described herein.

In one embodiment, the variants disclosed herein include nucleotide or amino acid substitutions, deletions, truncations, and insertions of one or more nucleotides of SEQ ID NO:1, 2, 4, or 6, or one or more amino acids of SEQ ID NO:3, 5, or 7, including up to about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 15, about 20, about 25, about 30 or more amino acid substitutions, deletions or insertions.

Naturally occurring variants can be identified with the use of well-known molecular biology techniques, such as polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant nucleotide sequences also include synthetically derived nucleotide sequences that have been generated, for example, by using site-directed mutagenesis but which still have CTP activity as defined herein.

Variants encompassed by the present invention are biologically active, that is they continue to possess the desired biological activity of the native sequence, that is, retaining CTP activity (i.e., facilitating translocation of the expressed polypeptide to the chloroplast). By "retains CTP activity" is intended that the variant will direct the translocation to the chloroplast of at least about 30%, at least about 50%, at least about 70%, or at least about 80% of the polypeptide encoded by the nucleotide sequence of interest. Methods for measuring CTP activity are well known in the art. See the section entitled "Evaluation of CTP Activity" for examples of suitable methods.

The skilled artisan will further appreciate that changes to the CTP can be introduced by mutation into the nucleotide sequence encoding the CTPs of the invention without altering the ability of the CTP to drive translocation of a polypeptide in the chloroplast of a plant cell. Thus, variant isolated nucleic acid molecules can be created by introducing one or more nucleotide substitutions, additions, or deletions into the corresponding nucleotide sequence disclosed herein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Such variant nucleotide sequences are also encompassed by the present invention.

Alternatively, variant nucleotide sequences can be made by introducing mutations randomly along all or part of the CTP sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for ability to drive translocation of an operably linked polypeptide sequence into the chloroplast a plant cell.

By "operably linked" is intended a functional linkage between a regulatory element (e.g., a CTP) and a second sequence, wherein the CTP sequence directs the translocation of the polypeptide of interest to the chloroplast of a plant cell. Generally, but not always, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

To determine the percent identity of two nucleic acids, the sequences are aligned for optimal comparison purposes. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., percent identity = number of identical positions/total number of positions (e.g., overlapping positions) x 100). In one embodiment, the two sequences are the same

length. In another embodiment, the comparison is across the entirety of the reference sequence (e.g., SEQ ID NO:1, 2, 4, or 6). The percent identity between two sequences can be determined using techniques similar to those described below, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A nonlimiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the BLASTN program of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403. BLAST nucleotide searches can be performed with the BLASTN program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to sequences of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389. Alternatively, PSI-Blast can be used to perform an iterated search that detects distant relationships between molecules. *See* Altschul *et al.* (1997) *supra*. When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (e.g., BLASTN) can be used. See, www.ncbi.nlm.nih.gov. Another non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the ClustalW algorithm (Higgins *et al.* (1994) *Nucleic Acids Res.* 22:4673-4680). ClustalW compares sequences and aligns the entirety of the DNA sequence, and thus can provide data about the sequence conservation of the entire nucleotide sequence. The ClustalW algorithm is used in several commercially available DNA analysis software packages, such as the ALIGNX module of the vector NTi Program Suite (Informax, Inc). A non-limiting example of a software program useful for analysis of ClustalW alignments is GeneDoc™. Genedoc™ (Karl Nicholas) allows assessment of DNA similarity and identity between multiple genes. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0), which is part of the GCG sequence alignment software package (available from Accelrys, Inc., 9865 Scranton Rd., San Diego, California, USA).

Unless otherwise stated, GAP Version 10, which uses the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48(3):443-453, will be used to determine sequence identity or similarity using the following parameters: % identity and % similarity for a nucleotide sequence using GAP Weight of 50 and Length Weight of 3, and the nwsgapdna.cmp scoring matrix; % identity or % similarity for an amino acid sequence using GAP weight of 8 and length weight of 2, and the BLOSUM62 scoring program. Equivalent programs may also be used. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10.

Using methods such as PCR, hybridization, and the like, corresponding sequences from other organisms, particularly other algal organisms, can be identified, such sequences having substantial identity to the sequences of the invention. *See*, for example, Sambrook J., and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) and Innis, *et al.* (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, NY). Sequences identified by their identity to the CTP sequences set forth herein are encompassed by the present invention.

In another embodiment, CTPs can be identified based on the identification of sequences known to comprise CTPs. For example, chloroplast-targeted sequences can be identified based on similarity to the chloroplast-targeted proteins disclosed herein or known in the art (e.g., acetolactate synthase (AHAS), small subunit (SSU), and EPSPS). The CTP sequence from these targeted proteins can be identified using methods known in the art. See, for example, Emanuelsson and von Heijne (2001) *Biochimica et Biophysica Acta* 1541:114-119; Nielson *et al.* (1997) *Protein Eng.* 10:1-6; and, Nielson and Krogh (1998) *Intell. Syst. Mol. Biol.* 6:122-130, each of which is herein incorporated by reference in its entirety. A variety of computer programs are also available for identifying. See, for example, ChloroP (which can be found at the internet address www.cbs.dtu.dk/services/ChloroP/); Predotar (which can be found at the internet address www.inra.fr/Internet/Produits/Predotar/); and, SignalP (which can be found at the internet address www.cbs.dtu.dk/services/SignalP/).

Oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA from an organism of interest. Methods for designing PCR primers and PCR cloning are generally known in the art and are disclosed in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). See also Innis *et al.*, eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis and Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); and Innis and Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, New York). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, and partially-mismatched primers.

In a hybridization method, all or part of a known nucleotide sequence can be used to screen cDNA or genomic libraries. Methods for construction of such cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook and Russell, 2001, *supra*. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labeled with a detectable group such as ^{32}P, or any other detectable marker, such as other radioisotopes, a fluorescent compound, an enzyme, or an enzyme co-factor. Probes for hybridization can be made by labeling synthetic oligonucleotides based on the known CTP-encoding sequence disclosed herein or primers to the known chloroplast targeted protein. Degenerate primers designed on the basis of conserved nucleotides in the nucleotide sequence can additionally be used. The probe typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, at least about 20, at least about 25, 30, 35, 40, 45, 50, 55, 60, 75, 100, 125, 150, 175, 200, 250, 300, 350, or 400 consecutive nucleotides of the CTP-encoding sequence of the invention, a nucleotide sequence encoding a chloroplast targeted protein, or a fragment or variant thereof. Preparation of probes for hybridization is generally known in the art and is disclosed in Sambrook and Russell, 2001, supra, herein incorporated by reference.

For example, the entire CTP-encoding sequence disclosed herein (or coding sequence for chloroplast-targeted protein), or one or more portions thereof, may be used as a probe capable of specifically hybridizing to corresponding CTP-like sequences. To achieve specific hybridization under a variety of conditions, such

probes include sequences that are unique and are at least about 10 nucleotides in length, or at least about 20 nucleotides in length. Such probes may be used to amplify corresponding CTP-encoding sequences from a chosen organism by PCR. This technique may be used to isolate additional coding sequences from a desired organism or as a diagnostic assay to determine the presence of coding sequences in an organism. Hybridization techniques include hybridization screening of plated DNA libraries (either plaques or colonies; see, for example, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Hybridization of such sequences may be carried out under stringent conditions. By "stringent conditions" or "stringent hybridization conditions" is intended conditions under which a probe will hybridize to its target sequence to a detectably greater degree than to other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences that are 100% complementary to the probe can be identified (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, or less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulfate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. Optionally, wash buffers may comprise about

0.1% to about 1% SDS. Duration of hybridization is generally less than about 24 hours, usually about 4 to about 12 hours. Optionally, wash buffers may comprise about 0.1% to about 1% SDS.

Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl (1984) *Anal. Biochem.* 138:267-284: $T_m = 81.5°C + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\%\ form) - 500/L$; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with $\geq$90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T_m, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T_m of less than 45°C (aqueous solution) or 32°C (formamide solution), the SSC concentration can be increased so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes,* Part I, Chapter 2 (Elsevier, New York); and Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Chapter 2 (Greene Publishing and Wiley-Interscience, New York). See Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Isolated sequences that have CTP activity and which hybridize under stringent conditions to the CTP sequences disclosed herein, or to fragments thereof, are encompassed by the present invention.

Methods of Use

Methods of the present invention are directed to the proper expression, translocation, and processing of chloroplast-targeted sequences in higher plants and plant cells under the control of the CTP sequences of the present invention. For the purposes of the present invention, a "processed" chloroplast targeted protein is one in which the CTP has been removed. At the time of translocation of a chloroplast targeted protein into the chloroplast of a plant cell, the CTP is removed from targeted protein by cleavage at a particular "cleavage site" between the CTP and the mature protein. The cleavage site can be determined experimentally, or may be predicted based on sequence structure (e.g., by alignment of the unprocessed protein with chloroplast targeted proteins in which the cleavage site is known, by analyzing the sequence for the presence of characteristic CTP domains, and the like) or by using one or more algorithms for cleavage site prediction as discussed elsewhere herein (e.g., SignalP).

The transgenic plants may have a change in phenotype, including, but not limited to, an altered pathogen or insect defense mechanism, an increased resistance to one or more herbicides, an increased ability to withstand stressful environmental conditions, a modified ability to produce starch, a modified level of starch production, a modified oil content and/or composition, a modified ability to utilize, partition and/or store nitrogen, and the like. These results can be achieved through the expression and targeting of a polypeptide of interest to chloroplasts in plants, wherein the polypeptide of interest functions in the chloroplast. The CTP sequences of the invention are useful for targeting native sequences as well as heterologous (non-native) sequences in higher plants. For the purposes of the present invention, "higher plants" are considered members of the subkingdom Embryophytae. In one embodiment, the plant is a monocotyledon. In another embodiment, the plant is a dicotyledon.

Generally, the nucleotide sequence encoding the CTP of the invention is provided in an expression cassette with a nucleotide sequence of interest for expression in the plant of interest. In one embodiment, the CTP-encoding sequences

of the invention are useful for the improved translocation of native sequences in a plant. In other embodiments, the CTP-encoding sequences are useful for expression and translocation of polypeptides encoded by heterologous nucleotide sequences. By "heterologous nucleotide sequence" is intended a sequence that is not naturally operably-linked with the CTP-encoding sequence of the invention, including non-naturally occurring multiple copies of a naturally occurring DNA sequence. While this nucleotide sequence is heterologous to the CTP-encoding sequence, it may be homologous, or "native," or heterologous, or "foreign," to the plant host. In some cases, the transformed plant may have a change in phenotype. "Heterologous" generally refers to the nucleic acid sequences that are not endogenous to the cell or part of the native genome in which they are present, and have been added to the cell by infection, transfection, microinjection, electroporation, microprojection, or the like.

Any nucleotide sequence of interest may be used with the CTP-encoding sequences of the invention, so long as the polypeptide encoded by the nucleotide sequence of interest (i.e., the "polypeptide of interest") is functional in a chloroplast. Such nucleotide sequences include, but are not limited to, herbicide-tolerance coding sequences, insecticidal coding sequences, nematicidal coding sequences, antimicrobial coding sequences, antifungal coding sequences, antiviral coding sequences, abiotic and biotic stress tolerance coding sequences, or sequences modifying plant traits such as yield, grain quality, nutrient content, starch quality and quantity, nitrogen fixation and/or utilization, and oil content and/or composition. More specific genes of interest for the present invention include, but are not limited to, genes that improve crop yield, genes that improve desirability of crops, genes encoding proteins conferring resistance to abiotic stress, such as drought, temperature, salinity, toxic metals or trace elements, or those conferring resistance to toxins such as pesticides and herbicides, or to biotic stress, such as attacks by fungi, viruses, bacteria, insects, and nematodes, and development of diseases associated with these organisms. It is recognized that any gene of interest can be operably linked to the CTP-encoding sequences of the invention and expressed in a plant, so long as the polypeptide encoded by the gene is functional in chloroplasts.

These nucleotide sequences of interest may encode proteins involved in providing disease or pest resistance. By "disease resistance" or "pest resistance" is intended that the plants avoid the harmful symptoms that are the outcome of the plant-pathogen interactions. Disease resistance and insect resistance genes such as

lysozymes or cecropins for antibacterial protection, or proteins such as defensins, glucanases or chitinases for antifungal protection, or *Bacillus thuringiensis* endotoxins, protease inhibitors, collagenases, lectins, or glycosidases for controlling nematodes or insects are all examples of useful gene products. Examples of genes of interest may be found, for example, at www.nbiap.vt.edu/cfdocs/fieldtests2.cfm.

"Pest" includes, but is not limited to, insects, fungi, bacteria, viruses, nematodes, mites, ticks, and the like. Insect pests include insects selected from the orders *Coleoptera, Diptera, Hymenoptera, Lepidoptera, Mallophaga, Homoptera, Hemiptera, Orthroptera, Thysanoptera, Dermaptera, Isoptera, Anoplura, Siphonaptera, Trichoptera,* etc., particularly *Coleoptera, Lepidoptera, and Diptera*. Viruses include but are not limited to tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, maize dwarf mosaic virus, etc. Nematodes include but are not limited to parasitic nematodes such as root knot, cyst, and lesion nematodes, including *Heterodera* spp., *Meloidogyne* spp., and *Globodera* spp.; particularly members of the cyst nematodes, including, but not limited to, *Heterodera glycines* (soybean cyst nematode); *Heterodera schachtii* (beet cyst nematode); *Heterodera avenae* (cereal cyst nematode); and *Globodera rostochiensis* and *Globodera pailida* (potato cyst nematodes). Lesion nematodes include but are not limited to *Pratylenchus* spp. Fungal pests include those that cause leaf, yellow, stripe and stem rusts.

An "herbicide resistance protein" or a protein resulting from expression of an "herbicide resistance-encoding nucleic acid molecule" includes proteins that confer upon a cell the ability to tolerate a higher concentration of an herbicide than cells that do not express the protein, or to tolerate a certain concentration of an herbicide for a longer period of time than cells that do not express the protein. Herbicide resistance traits may be introduced into plants by genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides, genes coding for resistance to herbicides that act to inhibit the action of glutamine synthase, such as phosphinothricin or basta (*e.g.*, the *bar* gene), glyphosate (e.g., the EPSP synthase gene and the GAT gene) or other such genes known in the art.

Genes that improve crop yield include dwarfing genes, such as Rht1 and Rht2 (Peng *et al.* (1999) *Nature* 400:256-261), and those that increase plant growth, such as ammonium-inducible glutamate dehydrogenase. Genes that improve desirability of crops include, for example, those that allow plants to have a reduced saturated fat

content, those that boost the nutritional value of plants, and those that increase grain protein. Genes that improve salt tolerance are those that increase or allow plant growth in an environment of higher salinity than the native environment of the plant into which the salt-tolerant gene(s) has been introduced.

Plant transformation vectors

Typically the plant expression cassette will be inserted into a "plant transformation vector." By "transformation vector" is intended a DNA molecule that is necessary for efficient transformation of a cell. Such a molecule may consist of one or more expression cassettes, and may be organized into more than one "vector" DNA molecule. For example, binary vectors are plant transformation vectors that utilize two non-contiguous DNA vectors to encode all requisite cis- and trans-acting functions for transformation of plant cells (Hellens and Mullineaux (2000) *Trends in Plant Science* 5:446-451). "Vector" refers to a nucleic acid construct designed for transfer between different host cells. "Expression vector" refers to a vector that has the ability to incorporate, integrate and express heterologous DNA sequences or fragments in a foreign cell. By "introducing" is intended to present to the organism being transformed the nucleotide construct in such a manner that the construct gains access to the interior of at least one cell of the organism.

This plant transformation vector may be comprised of one or more DNA vectors needed for achieving plant transformation. For example, it is a common practice in the art to utilize plant transformation vectors that are comprised of more than one contiguous DNA segment. These vectors are often referred to in the art as 'binary vectors'. Binary vectors as well as vectors with helper plasmids are most often used for *Agrobacterium*-mediated transformation, where the size and complexity of DNA segments needed to achieve efficient transformation is quite large, and it is advantageous to separate functions onto separate DNA molecules. Binary vectors typically contain a plasmid vector that contains the cis-acting sequences required for T-DNA transfer (such as left border and right border), a selectable marker that is engineered to be capable of expression in a plant cell, and a "gene of interest" (a gene engineered to be capable of expression in a plant cell for which generation of transgenic plants is desired). Also present on this plasmid vector are sequences required for bacterial replication.

The cis-acting sequences are arranged in a fashion to allow efficient transfer into plant cells and expression therein. For example, the selectable marker gene and the gene of interest are located between the left and right borders. Often a second plasmid vector contains the trans-acting factors that mediate T-DNA transfer from *Agrobacterium* to plant cells. This plasmid often contains the virulence functions (Vir genes) that allow infection of plant cells by *Agrobacterium*, and transfer of DNA by cleavage at border sequences and vir-mediated DNA transfer, as in understood in the art (Hellens and Mullineaux (2000) *Trends in Plant Science*, 5:446-451). Several types of *Agrobacterium* strains (e.g. LBA4404, GV3101, EHA101, EHA105, etc.) can be used for plant transformation. The second plasmid vector is not necessary for transforming the plants by other methods such as microprojection, microinjection, electroporation, polyethylene glycol, etc.

Plant Transformation

Methods of the invention involve introducing a nucleotide construct into a plant. By "introducing" is intended to present to the plant the nucleotide construct in such a manner that the construct gains access to the interior of a cell of the plant. The methods of the invention do not require that a particular method for introducing a nucleotide construct to a plant is used, only that the nucleotide construct gains access to the interior of at least one cell of the plant. Methods for introducing nucleotide constructs into plants are known in the art including, but not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

Transformation of plant cells can be accomplished by one of several techniques known in the art. By "plant" is intended whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds, plant cells, propagules, embryos and progeny of the same. Plant cells can be differentiated or undifferentiated (e.g. callus, suspension culture cells, protoplasts, leaf cells, root cells, phloem cells, pollen). "Transgenic plants" or "transformed plants" or "stably transformed" plants or cells or tissues refer to plants that have incorporated or integrated exogenous nucleic acid sequences or DNA fragments into the plant cell. By "stable transformation" is intended that the nucleotide construct introduced into a plant integrates into the genome of the plant and is capable of being inherited by progeny thereof.

In general, plant transformation methods involve transferring heterologous DNA into target plant cells (e.g. immature or mature embryos, suspension cultures, undifferentiated callus, protoplasts, etc.), followed by applying a maximum threshold level of appropriate selection (depending on the selectable marker gene) to recover the transformed plant cells from a group of untransformed cell mass. Explants are typically transferred to a fresh supply of the same medium and cultured routinely. Subsequently, the transformed cells are differentiated into shoots after placing on regeneration medium supplemented with a maximum threshold level of selecting agent. The shoots are then transferred to a selective rooting medium for recovering rooted shoot or plantlet. The transgenic plantlet then grow into mature plant and produce fertile seeds (e.g. Hiei *et al.* (1994) *The Plant Journal* 6:271-282; Ishida *et al.* (1996) *Nature Biotechnology* 14:745-750). A general description of the techniques and methods for generating transgenic plants are found in Ayres and Park (1994) *Critical Reviews in Plant Science* 13:219-239 and Bommineni and Jauhar (1997) *Maydica* 42:107-120. Since the transformed material contains many cells; both transformed and non-transformed cells are present in any piece of subjected target callus or tissue or group of cells. The ability to kill non-transformed cells and allow transformed cells to proliferate results in transformed plant cultures. Often, the ability to remove non-transformed cells is a limitation to rapid recovery of transformed plant cells and successful generation of transgenic plants. Molecular and biochemical methods may be used to confirm the presence of the integrated heterologous gene of interest in the genome of transgenic plant.

Generation of transgenic plants may be performed by one of several methods, including, but not limited to, introduction of heterologous DNA by *Agrobacterium* into plant cells (*Agrobacterium*-mediated transformation), bombardment of plant cells with heterologous foreign DNA adhered to particles, and various other non-particle direct-mediated methods (e.g. Hiei *et al.* (1994) *The Plant Journal* 6:271-282; Ishida *et al.* (1996) *Nature Biotechnology* 14:745-750; Ayres and Park (1994) *Critical Reviews in Plant Science* 13:219-239; Bommineni and Jauhar (1997) *Maydica* 42:107-120) to transfer DNA.

Methods for transformation of chloroplasts are known in the art. See, for example, Svab *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530; Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917; Svab and Maliga (1993) *EMBO J.* 12:601-606. The method relies on particle gun delivery of DNA containing

a selectable marker and targeting of the DNA to the plastid genome through homologous recombination. Additionally, plastid transformation can be accomplished by transactivation of a silent plastid-borne transgene by tissue-preferred expression of a nuclear-encoded and plastid-directed RNA polymerase. Such a system has been reported in McBride *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:7301-7305.

The cells that have been transformed may be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having constitutive expression of the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved. In this manner, the present invention provides transformed seed (also referred to as "transgenic seed") having a nucleotide construct of the invention, for example, an expression cassette of the invention, stably incorporated into their genome.

Plants

The present invention may be used for transformation of any higher plant species, including, but not limited to, monocots and dicots. In one embodiment, the CTP encompassed herein is active in both monocots and dicots. In another embodiment, the CTP is active only in monocots or only in dicots. Examples of plants of interest include, but are not limited to, corn (maize), sorghum, wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, and oilseed rape, *Brassica* sp., alfalfa, rye, millet, safflower, peanuts, sweet potato, cassava, coffee, coconut, pineapple, citrus trees, cocoa, tea, banana, avocado, fig, guava, mango, olive, papaya, cashew, macadamia, almond, oats, vegetables, ornamentals, and conifers.

Vegetables include, but are not limited to, tomatoes, lettuce, green beans, lima beans, peas, and members of the genus *Curcumis* such as cucumber, cantaloupe, and musk melon. Ornamentals include, but are not limited to, azalea, hydrangea, hibiscus, roses, tulips, daffodils, petunias, carnation, poinsettia , and chrysanthemum. Preferably,

plants of the present invention are crop plants (for example, maize, sorghum, wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, oilseed rape, etc.).

This invention is particularly suitable for any member of the monocot plant family including, but not limited to, maize, rice, barley, oats, wheat, sorghum, rye, sugarcane, pineapple, yams, onion, banana, coconut, and dates.

Evaluation of Plant Transformation

Following introduction of heterologous foreign DNA into plant cells, the transformation or integration of heterologous DNA in the plant genome is confirmed by various methods such as analysis of nucleic acids or proteins and metabolites associated with the integrated DNA.

PCR analysis is a rapid method to screen transformed cells, tissue or shoots for the presence of incorporated DNA at the earlier stage before transplanting into the soil (Sambrook and Russell, 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). PCR is carried out using oligonucleotide primers specific to the gene of interest or *Agrobacterium* vector background, etc.

Plant transformation may be confirmed by Southern blot analysis of genomic DNA (Sambrook and Russell, 2001, *supra*). In general, total DNA is extracted from the transformant, digested with appropriate restriction enzymes, fractionated in an agarose gel and transferred to a nitrocellulose or nylon membrane. The membrane or "blot" then is probed with, for example, radiolabeled ^{32}P target DNA fragment to confirm the integration of introduced DNA in the plant genome according to standard techniques (Sambrook and Russell, 2001, *supra*).

In Northern blot analysis, RNA is isolated from specific tissues of transformant, fractionated in a formaldehyde agarose gel, blotted onto a nylon filter according to standard procedures that are routinely used in the art (Sambrook and Russell, 2001, *supra*). Expression of RNA encoded by a heterologous gene operably linked to the CTP-encoding sequence is then tested by hybridizing the filter to a radioactive probe derived from the heterologous gene, by methods known in the art (Sambrook and Russell, 2001, *supra*).

Evaluation of CTP Activity

Assays to determine the efficiency by which the CTP sequences of the invention target a protein of interest to a chloroplast are known. See, for example, Mishkind *et al.* (1985) *J of Cell Biol* 100:226-234, which is herein incorporated by reference in its entirety. A reporter gene such as-glucuronidase (GUS), chloramphenicol acetyl transferase (CAT), or green fluorescent protein (GFP) is operably linked to the CTP sequence. This fusion is placed behind the control of a suitable promoter, ligated into a transformation vector, and transformed into a plant or plant cell. Following an adequate period of time for expression and localization into the chloroplast, the chloroplast fraction is extracted and reporter activity assayed. The ability of the isolated sequences to target and deliver the reporter protein to the chloroplast can be compared to other known CTP sequences. See de Castro Silva Filho et al. (1996) Plant Mol. Biol. 30: 769-780. Protein import can also be verified *in vitro* through the addition of proteases to the isolated chloroplast fraction. Proteins which were successfully imported into the chloroplast are resistant to the externally added proteases whereas proteins that remain in the cytosol are susceptible to digestion. Protein import can also be verified by the presence of functional protein in the chloroplast using standard molecular techniques for detection, or by evaluating the phenotype resulting from expression of a chloroplast targeted protein.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

Example 1. *Chlamydomonas* AHAS CTP.

The *Chlamydomonas* AHAS CTP (SEQ ID NO:3) was identified from comparison of the full-length *Chlamydomonas* AHAS ("ChlamyAHAS") peptide (GENBANK accession AF022816; SEQ ID NO:8) with several AHAS peptides from plants, bacteria, yeast and fungi. This alignment shows that the ChlamyAHAS CTP shows very little sequence conservation with plant CTPs, whereas the mature proteins show conserved elements. Thus, by comparing the multiple AHAS proteins with the ChlamyAHAS protein, it was inferred that amino acids 1-99 of the *Chlamydomonas*

AHAS protein comprised the CTP. The predicted processing site for this CTP upon import into the *Chlamydomonas* chloroplast is at approximately amino acid 92 of SEQ ID NO:8.

Example 2. *Chlamydomonas* RuBisCo Small Subunit (SSU) CTP.

The *Chlamydomonas* SSU CTP (SEQ ID NO:5) was identified from comparison of the full–length *Chlamydomonas* SSU ("ChlamySSU") peptide (GENBANK accession CAA28160; SEQ ID NO:9) with several plant RuBisCo small subunit peptides. This alignment shows that the ChlamySSU CTP shows very little sequence conservation with plant CTPs, whereas the mature proteins show conserved elements. Thus, by comparing the known processing sites of several plant RuBisCo proteins with the ChlamySSU protein, it was inferred that amino acids 1-45 of the *Chlamydomonas* SSU protein (SEQ ID NO:9) comprised the CTP.

The cleavage site for the *Chlamydomonas* RuBisCO small subunit precursor ("ChlamySSU") has been determined empirically to be between residues 44 and 45 of SEQ ID NO:9 (Schmidt et al 1979, Journal of Cell Biology 83:615-662).

Example 3. *Chlamydomonas* EPSPS CTP.

The *Chlamydomonas* EPSPS CTP (SEQ ID NO:7) was identified from comparison of the full-length *Chlamydomonas* EPSPS peptide (GENBANK accession XP_001702942; SEQ ID NO:10) with several plant and bacterial EPSPSs. This alignment showed that the ChlamyEPSPS CTP shows very little sequence conservation with plant CTPs, whereas the mature proteins show conserved elements. Thus, by comparing several EPSPS proteins with the ChlamyEPSPS protein, it was inferred that amino acids 1-75 of the *Chlamydomonas* EPSPS protein (SEQ ID NO:10) comprised the CTP. The predicted processing site for this CTP upon import into the *Chlamydomonas* chloroplast is at approximately amino acid 61 of SEQ ID NO:10.

Table 1. *Chlamydomonas* CTPs

Protein	SEQ ID NO:	Residues of Precursor Protein Used for CTP	Predicted Cleavage Site
ChlamyAHAS	8	1-99	92
ChlamySSU	9	1-45	Amino Acid 44 or Amino Acid 45
ChlamyEPSPS	10	1-75	61

Example 4. DNA constructs for using the Chlamydomonas CTPs for protein targeting

A DNA element utilizing the *Chlamydomonas* CTP, including the *Chlamydomonas* AHAS, *Chlamydomonas* SSU, or *Chlamydomonas* EPSPS CTPs to generate multiple constructs for targeting of proteins to chloroplasts can involve inclusion of convenient restriction endonuclease recognition sites (for example a *Bam* HI restriction site) as well as small peptide linkers between the chloroplast CTP and the protein (for example a Gly-Ser-Gly tripeptide CTP; SEQ ID NO:18). Furthermore, such DNA elements can be designed and made synthetically, in a way that the DNA sequence is varied from the original DNA, but encodes the identical peptide.

Alternatively, one can design DNA constructs such that no restriction enzyme sites are needed, and the CTP/protein fusion can be accomplished by total synthesis of the combined coding region, or by PCR based strategies, including "sewing PCR" and the like.

One can also design the CTP/protein fusion in a manner where some of either protein is truncated. For example, one can remove one or more amino acids from the N-terminus of a bacterially expressed protein, and still achieve a functional fusion to a *Chlamydomonas* CTP.

A cassette containing a synthetically designed DNA sequence encoding the Chlamydomonas AHAS CTP that incorporates a BamHI restriction site and a Gly-Ser–Gly linker was designed. These DNA constructs contain (from 5' to 3') (1) a *Pst* I cloning site, (2) the bases ACC to provide "Kozak" context for efficient translation, (3) the portion of the gene encoding the amino terminal methionine through the known transit peptide cleavage site of the ChlamyAHAS, and including a small DNA

region encoding the amino acids C-terminal to the cleavage site, (4) DNA bases encoding the residues Gly – Ser – Gly with an embedded *BamH* I cloning site, and, (5) the coding region of the gene of interest.

A cassette containing a synthetically designed DNA sequence encoding the Chlamydomonas SSU CTP that incorporates a BamHI restriction site and a Gly-Ser–Gly linker was designed. These DNA constructs contain (from 5' to 3') (1) a *Pst* I cloning site, (2) the bases ACC to provide "Kozak" context for efficient translation, (3) the portion of the gene encoding the amino terminal methionine through the known transit peptide cleavage site of the ChlamySSU, and including a small DNA region encoding the amino acids C-terminal to the cleavage site, (4) DNA bases encoding the residues Gly – Ser – Gly with an embedded *BamH* I cloning site, and, (5) the coding region of the gene of interest.

A cassette containing a synthetically designed DNA sequence encoding the Chlamydomonas EPSPS CTP that incorporates a BamHI restriction site and a Gly-Ser–Gly linker was designed. These DNA constructs contain (from 5' to 3') (1) a *Pst* I cloning site, (2) the bases ACC to provide "Kozak" context for efficient translation, (3) the portion of the gene encoding the amino terminal methionine through the known transit peptide cleavage site of the ChlamyEPSPS, and including a small DNA region encoding the amino acids C-terminal to the cleavage site, (4) DNA bases encoding the residues Gly – Ser – Gly with an embedded *BamH* I cloning site, and, (5) the coding region of the gene of interest.

Example 5. Fusion of a transit peptide from a non-plant species to a heterologous protein, and proper localization and cleavage in monocots: *Chlamydomonas* AHAS CTP functions in monocots.

DNA constructs were designed such that the resulting protein encoded the *Chlamydomonas* AHAS transit peptide ("ChlamyAHAS") at the N-terminus, followed by a protein fusion to a gene conferring herbicide resistance upon cells (GRG-1; U.S. Patent No. 7,405,347). For the ChlamyAHAS precursor, the transit peptide cleavage sites were inferred from alignments of the protein sequences to ALS proteins from bacteria, fungi and yeast.

These DNA constructs contain (from 5' to 3') a *Pst* I cloning site, (2) the bases ACC to provide "Kozak" context for efficient translation, (3) the portion of the gene encoding the amino terminal methionine through the known transit peptide cleavage

site of the ChlamyAHAS CTP and including a small DNA region encoding the amino acids C-terminal to the cleavage site, (4) DNA bases encoding the residues Gly – Ser – Gly with an embedded *BamH* I cloning site, and, (5) the coding region of the gene of interest (in this case GRG-1).

These DNAs molecules were made synthetically (DNA 2.0 of Menlo Park, CA). The DNA sequence of the region containing this construct is provided as SEQ ID NO:11, and the resulting amino acid sequence is provided as SEQ ID NO:12.

A control construct (pAG250) was made, which contains GRG-1 expressed from the TrpPro5 promoter, wherein the GRG-1 protein does not have a chloroplast CTP.

This no CTP/GRG-1 construct was engineered into a vector for use in *Agrobacterium*-mediated transformation of maize embryos, and transgenic maize plants containing this construct were generated. T_0 transformed plants were analyzed by PCR to confirm presence of the construct in the maize lines, and these T_0 plants were then out-crossed to a non-transgenic line to generate hemizygous T_1 progeny. The resulting T_1 transgenic plants produce large amounts of GRG-1 protein. Nonetheless, plants transformed with pAG250 and expressing unlocalized GRG-1 are not resistant to glyphosate.

The algal CTP ChlamyAHAS /GRG-1 construct was engineered into a vector for use in *Agrobacterium*-mediated transformation of maize embryos, and transgenic maize plants containing this construct were generated.

T_0 transformed plants were analyzed by PCR to confirm presence of the construct in the maize lines, and these T_0 plants were then out-crossed to a non-transgenic line to generate hemizygous T_1 progeny. The resulting T_1 transgenic plants are resistant to spray applications of glyphosate (as compared to non-transgenic controls).

Western blots of leaf tissue from transgenic maize plants show that these plants express the GRG-1 protein. Furthermore, the size of the protein identified by Western blot is consistent with import of the protein into chloroplasts, and processing of the ChlamyAHAS/GRG-1 protein at or near the cleavage site.

Thus the ChlamyAHAS CTP is sufficient to target GRG-1 to the maize chloroplast, and result in a phenotype (herbicide resistance) that is not conferred by GRG-1 in the absence of targeting to the chloroplast.

Example 6. Fusion of a transit peptide from a non-plant species to a heterologous protein, and proper localization and cleavage in monocots: *Chlamydomonas* SSU CTP functions in monocots

To test if an algal chloroplast CTP can function in monocots, transgenic monocot plants were generated and expression and cleavage of an algal CTP was assessed by Western blot analysis.

DNA constructs were designed such that the resulting protein encoded the algal transit peptide at the N-terminus, fused to a protein conferring herbicide resistance upon cells (in this case the GRG-8 protein; U.S. Patent Publication No. 20060150270)

These DNA constructs contain (from 5' to 3') (1) a *Pst* I cloning site, (2) the bases ACC to provide "Kozak" context for efficient translation, (3) the portion of the gene encoding the amino terminal methionine through the known transit peptide cleavage site of the ChlamySSU, and including a small DNA region encoding the amino acids C-terminal to the cleavage site, (4) DNA bases encoding the residues Gly – Ser – Gly with an embedded *BamH* I cloning site, and, (5) the coding region of the gene of interest (in this case GRG-8).

The DNA sequence of the region containing this construct is provided as SEQ ID NO:13, and the resulting amino acid sequence is provided as SEQ ID NO:14.

This CTP/GRG-8 construct (pAG1675) was engineered into a vector for use in *Agrobacterium*-mediated transformation of maize embryos, and transgenic maize plants generated and identified.

T_0 plants transformed with pAG1675 were analyzed by PCR to confirm presence of the construct in the maize lines, and these T_0 plants were then out-crossed to a non-transgenic line to generate hemizygous T_1 progeny. T_1 transgenic plants exhibited resistance to spray applications of glyphosate compared to non-transgenic controls.

Example 7. Expression and processing of Chlamydomonas SSU chloroplast CTP fused to GRG-8 protein in maize cells

Western blots of leaf tissue from transgenic maize plants were found to express the CTP/GRG-8 protein. Total leaf protein was extracted from maize leaves (Pierce P-PER protein extract buffer) and separated on a 4-12% Bis-Tris gel. GRG-8 protein was visualized using goat anti-GRG8 polyclonal antibodies. A non-transgenic

maize extract was compared alongside (lane 3). To evaluate CTP processing, a HIS-tagged GRG-8 protein standard was purified from an *E. coli* strain. The size of the protein identified by Western blot is consistent with import of the protein into chloroplasts, and processing of the CTP/GRG-8 protein within the ChlamySSU CTP, at or near the predicted cleavage site.

Example 8. Evaluation of glyphosate tolerance of maize plant expressing *Chlamydomonas* SSU-GRG8 protein

The glyphosate spray tolerance of a transgenic maize event expressing the *Chlamydomonas* SSU-GRG8 protein was compared to several non-transgenic T_0 control plants. Individual plants were transferred to the greenhouse and grown in flats for 10 days. After 10 days, a glyphosate concentration that approximated a 1x field spray rate (7 mM supplemented with 0.1% Tween 20 as surfactant) was applied to the flats. The glyphosate was applied using a spray table to allow consistent application of the herbicide to individual plants. Plants were rated after 3 weeks to determine if the plants tolerated the glyphosate spray (mostly green leaf material: <50% damage) or did not tolerate the glyphosate spray (>75% damage, or plant death). The transgenic plant showed tolerance to glyphosate, whereas each of the control plants failed to show tolerance.

Example 9. Fusion of a transit peptide from a non-plant species to a heterologous protein, and proper localization and cleavage in monocots: *Chlamydomonas* EPSPS CTP functions in monocots

DNA and amino acid sequences for the *Chlamydomonas* EPSPS precursor were obtained from public databases. The transit peptide cleavage site was predicted based on alignments of the protein sequences with EPSPS proteins from bacteria, fungi and yeast. A synthetic gene was constructed which encoded the CTP from the amino terminal methionine through the predicted cleavage site. This DNA was ligated to create an in-frame fusion with the start codon of a synthetic *GRG-23(ace3)(R173K)* gene (U.S. Patent Application Publication No. 20080127372). This CTP-GRG-23(ace3)(R173K) cassette was then ligated into a plant transformation vector.

DNA constructs were designed such that the resulting protein encoded the *Chlamydomonas* EPSPS ("ChlamyEPSPS") transit peptide at the N-terminus, followed by fusion to a protein conferring herbicide resistance upon cells (GRG-23(ace3)(R173K)).

These DNAs molecules were made synthetically (DNA 2.0 of Menlo Park, CA). The DNA sequence of the region containing this construct is provided as SEQ ID NO:15, and the resulting amino acid sequence is provided as SEQ ID NO:16.

These DNA constructs contain (from 5' to 3'), (1) a *Pst* I cloning site, (2) the bases ACC to provide "Kozak" context for efficient translation, (3) the portion of the gene encoding the amino terminal methionine through the known transit peptide cleavage site of the ChlamyEPSPS and including a small DNA region encoding the amino acids C-terminal to the cleavage site, (4) the coding region of the gene of interest (in this case GRG-23(ace3)(R173K).

This *Chlamydomonas* EPSPS CTP/ GRG-23(ace3)(R173K) construct was engineered into a vector for use in *Agrobacterium*-mediated transformation of maize embryos, and transgenic events identified.

T_0 transformed plants were analyzed by PCR to confirm presence of the construct in the maize lines, and these T_0 plants were then out-crossed to a non-transgenic line to generate hemizygous T_1 progeny. The resulting T_1 transgenic plants are resistant to spray applications of glyphosate (as compared to non-transgenic controls). Western blots of leaf tissue from transgenic maize plants were found to express the *Chlamydomonas* EPSPS CTP/GRG-23(ace3)(R173K) protein. The protein detected in plant tissues is smaller than the full-length *Chlamydomonas* EPSPS – GRG-23(ace3)(R173K) protein, and is similar in size to the native GRG-23(ace3)(R173K) protein (Figure 2). The ability of the *Chlamydomonas* EPSPS CTP/GRG-23(ace3)(R173K) protein to confer herbicide resistance and the size of the resulting mature GRG-23(ace3)(R173K) protein in herbicide resistant plants are both consistent with import of the protein into chloroplasts, and processing of the *Chlamydomonas* EPSPS CTP/GRG-23(ace3)(R173K) protein at or near the cleavage site.

Example 10. Molecular weight analysis of *Chlamydomonas* EPSPS – GRG-23(ace3)(R173K) protein expressed in maize.

From a Western blot of transgenic maize lines expressing *Chlamydomonas* EPSPS – GRG-23(ace3)(R173K) protein, the distance migration of protein molecular weight standards was graphed to generate a linear plot of Log Molecular Weight vs. Distance Migration (Figure 3). The linear regression equation of this plot was used to calculate the apparent molecular weight of the GRG23(ace3)(R173K) protein standard (Figure 2, lane 8) and of the processed *Chlamydomonas* EPSPS – GRG-23(ace3)(R173K) protein detected in plant extract. By this method, the apparent molecular weight of the purified GRG23(ace3)(R173K) protein was determined to be 45,347 grams/mole, and the apparent molecular weight of the processed *Chlamydomonas* EPSPS – GRG-23(ace3)(R173K) (lane 6, distance migration = 48.5 mm) was determined as 46,227 grams/mole. The estimated molecular weights of the processed proteins are consistent with processing of the *Chlamydomonas* CTP several amino acids upstream of its junction with GRG-23(ace3)(R173K) (the molecular weight of GRG23(ace3)(R173K) is estimated to be 45,570 grams/mole). No protein of a size consistent with unprocessed *Chlamydomonas* EPSPS – GRG-23(ace3)(R173K) protein (MW = 52,012 grams/mole) is detected by Western blot. Therefore, the *Chlamydomonas* EPSPS CTP is processed in maize at a discrete recognition site within the *Chlamydomonas* CTP. Purification and N-terminal amino acid analysis of this protein by methods known in the art would allow unambiguous determination of the exact cleavage site within the *Chlamydomonas* CTP.

Example 11. Evaluation of algal CTP sequences in dicots

To assess the ability of algal chloroplast CTPs to function in dicot cells, the Chlamydomonas AHAS Chloroplast CTP (SEQ ID NO:3) was positioned in frame 5' of TagGFP gene (Evrogen, Moscow, Russia). A control vector contained the TagGFP gene without a chloroplast transit peptide. The ChlamyAHAS CTP construct (pAX3517) and control construct (pAX3521) were organized to initiate transcription from the *Arabidopsis* UBQ3 promoter (Norris et. al, 1993, Plant Molecular Biology 21:895-906). Constructs utilized either 35S or PinII transcriptional terminators.

Approximately twelve micrograms of each purified plasmid was used in polyethylene glycol-mediated tobacco protoplast transformation experiments. After

transformation, the protoplasts were incubated in a growth chamber at 25°C for 23 hours. Expression and localization of TagGFP protein following transient expression was monitored under an inverted fluorescent microscope.

The construct expressing the TagGFP without a chloroplast CTP was detected only in the cytoplasm of protoplasts. However, the *Chlamydomonas* AHAS CTP construct correctly delivered TagGFP into the chloroplast resulting in accumulation of florescence in the chloroplast of these protoplasts (Figure 1).

Example 12. Evaluation of an algal CTP sequence in soybean cells

To assess the ability of algal chloroplast CTPs to function in soybean cells, the ChlamyAHAS CTP construct (pAX3517), the ChlamyEPSPS construct (pAX4562), and a control construct (pAX3521) containing the TagGFP gene without a chloroplast transit peptide were used in polyethylene glycol-mediated transformation of soybean protoplasts. Expression and localization of TagGFP protein following transient expression was monitored under an inverted fluorescent microscope.

For the control construct lacking a chloroplast CTP, TagGFP fluorescence was observed only in the cytoplasm. Similarly, no expression in the chloroplasts was observed from two independent constructs expressing TagGFP with the ChlamyEPSPS chloroplast CTP. However, TagGFP was detected in the chloroplast of protoplasts that had been transformed with the ChlamyAHAS chloroplast CTP construct. Thus, this CTP functions in soybean and well as tobacco cells.

Table 2. Localization of algal chloroplast CTP-linked proteins in dicot cells.

Construct	CTP	Gene	Localization of TagGFP in tobacco Protoplasts	Localization of TagGFP in soybean Protoplasts
pAX3517	ChlamyAHAS	TagGFP	Chloroplast	Chloroplast
pAX4562	ChlamyEPSPS	TagGFP	*not tested*	Non-Chloroplast
pAX3521	None	TagGFP	Non-Chloroplast	Non-Chloroplast

Table 3. Summary of function of Chlamydomonas CTPs in Plant Cells

CTP	Function in Monocot cells	Function in Dicot Cells
ChlamySSU	++	*not tested*
ChlamyEPSPS	+++	-
ChlamyAHAS	+++	+++

All publications and patent applications mentioned in the specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

THAT WHICH IS CLAIMED:

1. A plant cell having stably incorporated into its genome an expression cassette comprising a nucleotide sequence encoding a chloroplast targeting peptide (CTP), wherein said nucleotide sequence encoding said CTP is operably linked to a nucleotide sequence of interest, and wherein said nucleotide sequence encoding said CTP is selected from the group consisting of:

a) the nucleotide sequence set forth in SEQ ID NO:1, 2, or 4;

b) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO:1, 2, or 4, wherein said nucleotide sequence encodes a chloroplast transit peptide;

c) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO:3 or 5; and,

d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO:3 or 5, wherein said amino acid sequence is a chloroplast transit peptide.

2. The plant cell of claim 1, wherein said plant cell is from a monocot.

3. The plant cell of claim 1, wherein said plant cell is from a dicot.

4. A plant comprising the plant cell of claim 1.

5. The plant of claim 4, wherein said plant is a monocot.

6. The plant of claim 4, wherein said plant is a dicot.

7. A seed derived from the plant of claim 7, wherein said seed comprises the nucleotide sequence encoding said chloroplast transit peptide.

8. The plant cell of claim 1, wherein the nucleotide sequence of interest encodes a gene product that confers herbicide, pathogen, or insect resistance.

9. A monocot plant cell having stably incorporated into its genome an expression cassette comprising a nucleotide sequence encoding a chloroplast targeting peptide (CTP), wherein said nucleotide sequence encoding said CTP is operably linked to a nucleotide sequence of interest, and wherein said nucleotide sequence encoding said CTP is selected from the group consisting of:

a) the nucleotide sequence set forth in SEQ ID NO:6;

b) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO:6, wherein said nucleotide sequence encodes a chloroplast transit peptide;

c) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO:7; and,

d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO:7, wherein said amino acid sequence is a chloroplast transit peptide.

10. A plant comprising the plant cell of claim 9.

11. A seed derived from the plant of claim 10, wherein said seed comprises the nucleotide sequence encoding said chloroplast transit peptide.

12. The plant cell of claim 9, wherein the nucleotide sequence of interest encodes a gene product that confers herbicide, pathogen, or insect resistance.

13. A method for expressing a nucleotide sequence of interest in a plant, said method comprising:

a) introducing into a plant cell an expression cassette comprising a nucleotide sequence encoding a chloroplast targeting peptide (CTP), wherein said nucleotide sequence encoding said CTP is operably linked to said nucleotide sequence of interest, and wherein said nucleotide sequence encoding said CTP is selected from the group consisting of

i) the nucleotide sequence set forth in SEQ ID NO:1, 2, or 4;

ii) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO:1, 2, or 4, wherein said nucleotide sequence encodes a chloroplast transit peptide;

iii) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO:3 or 5; and,

iv) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO:3 or 5, wherein said amino acid sequence is a chloroplast transit peptide; and,

b) regenerating a transformed plant from said plant cell; wherein said plant has stably incorporated into its genome said expression cassette.

14. The method of claim 13, wherein said plant is a dicot.

15. The method of claim 13, wherein said plant is a monocot.

16. The method of claim 13, wherein said nucleotide sequence of interest encodes a gene product that confers herbicide or pest resistance.

17. A method for expressing a nucleotide sequence of interest in a monocot plant, said method comprising:

a) introducing into a plant cell an expression cassette comprising a nucleotide sequence encoding a chloroplast targeting peptide (CTP), wherein said nucleotide sequence encoding said CTP is operably linked to said nucleotide sequence of interest, and wherein said nucleotide sequence encoding said CTP is selected from the group consisting of

i) the nucleotide sequence set forth in SEQ ID NO:6;

ii) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO:6, wherein said nucleotide sequence encodes a chloroplast transit peptide;

iii) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO:7; and,

iv) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO:7, wherein said amino acid sequence is a chloroplast transit peptide; and,

b) regenerating a transformed plant from said plant cell;

wherein said plant has stably incorporated into its genome said expression cassette.

18. The method of claim 17, wherein said nucleotide sequence of interest encodes a gene product that confers herbicide or pest resistance.

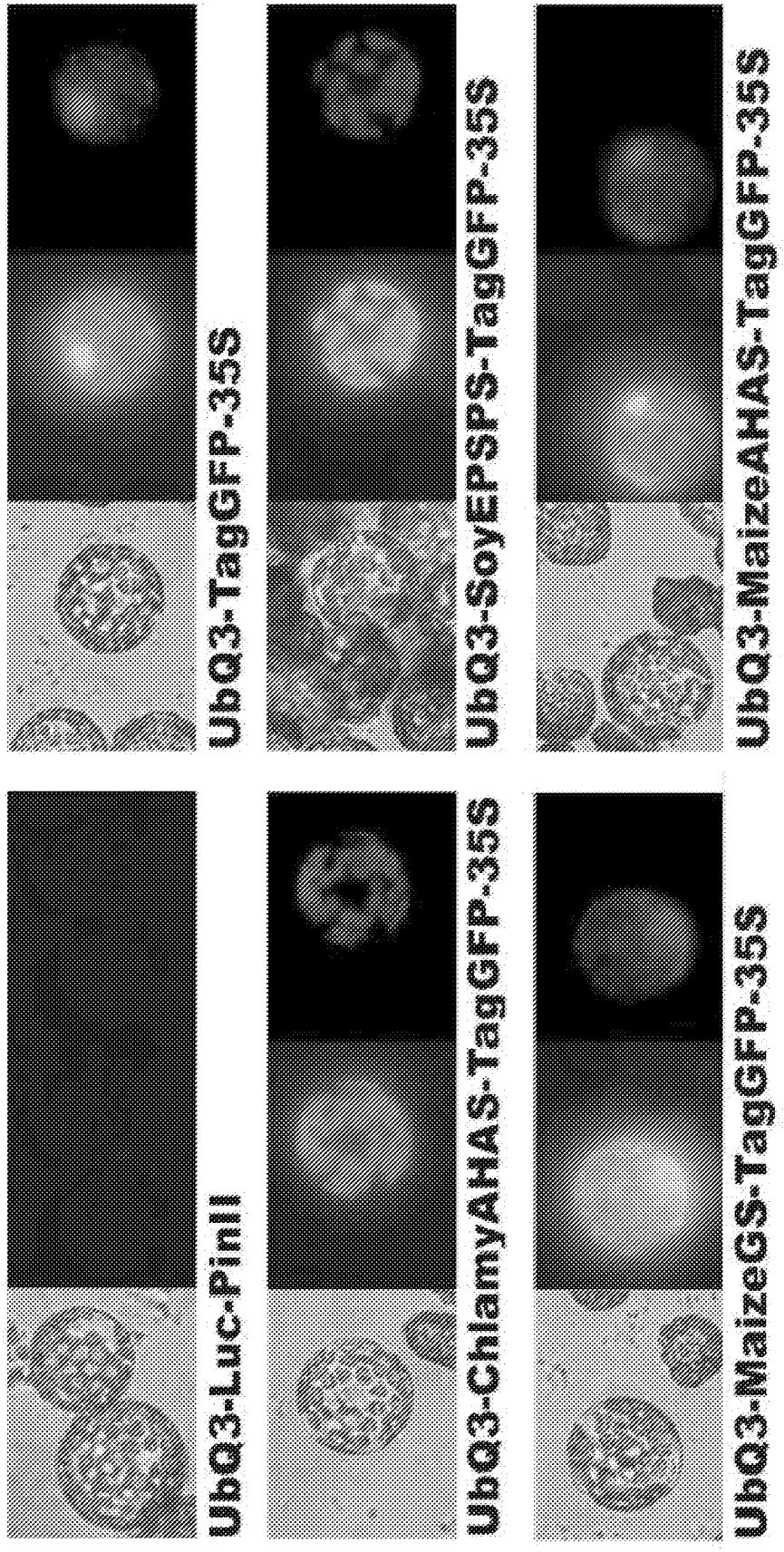

UbQ3-Luc-PinII
UbQ3-TagGFP-35S
UbQ3-ChlamyAHAS-TagGFP-35S
UbQ3-SoyEPSPS-TagGFP-35S
UbQ3-MaizeGS-TagGFP-35S
UbQ3-MaizeAHAS-TagGFP-35S

FIG. 1

1/3

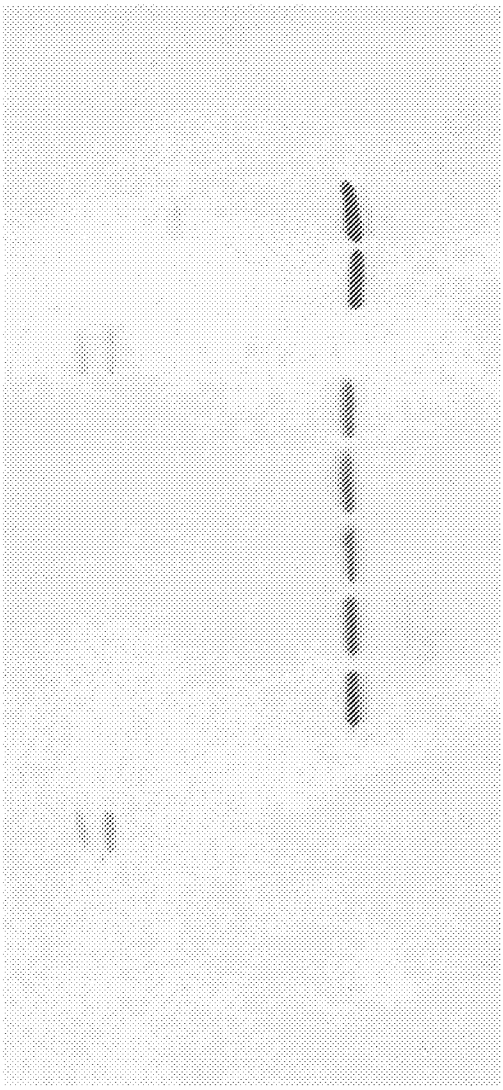


FIG. 2
2/3

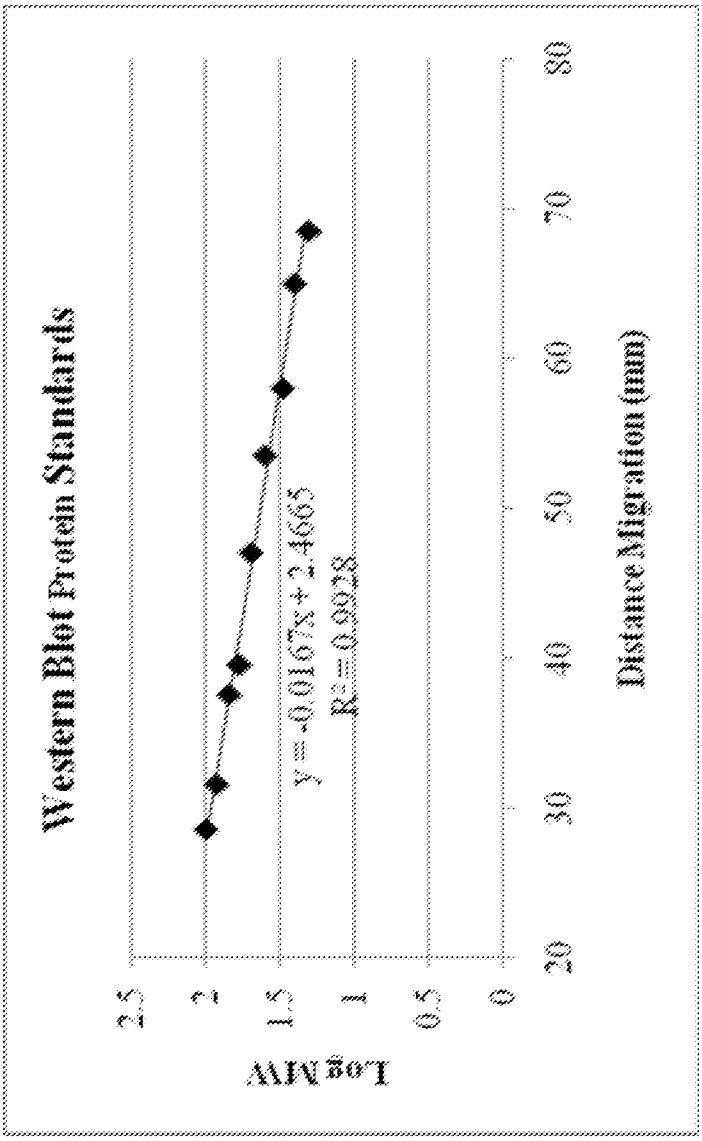

Western Blot Protein Standards
y = -0.0167x + 2.4665
R² = 0.9928
Log MW
Distance Migration (mm)
2.5
2
1.5
1
0.5
0
20
30
40
50
60
70
80

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2009/056165

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/82

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)
C12N

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 2008/039450 A2 (LEE JAMES WEIFU [US]) 3 April 2008 (2008-04-03) SEQ ID NO:24 is identical to present SEQ ID NO:5; page 7, line 1 - line 16; claims 20,21,47	1-7, 13-15 1-18
Y	DATABASE UniProt [Online] 1 January 1998 (1998-01-01), "SubName: Full=Acetolactate synthase;" XP002559552 retrieved from EBI accession no. UNIPROT:022547 Database accession no. 022547 shows 100% identity in 99 aa overlap to SEQ ID NO:3, 99,3% in 297 nt overlap to SEQ ID NO:1 and 98,3% in 297 nt overlap to SEQ ID NO:2; abstract; compound -/--	1-8, 13-16

[X] Further documents are listed in the continuation of Box C.

[X] 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 but 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
9 December 2009	30/12/2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Kania, Thomas

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2009/056165

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DATABASE UniProt [Online] 4 December 2007 (2007-12-04), "RecName: Full=3-phosphoshikimate 1-carboxyvinyltransferase; EC=<A HREF="http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?[enzyme-ECNumber:2.5.1.19]+-e">2.5.1.19</A>;" XP002559553 retrieved from EBI accession no. UNIPROT:A8JH48 Database accession no. A8JH48 abstract; compound	9-12, 17-18
A	ROESLER K R ET AL: "PRIMARY STRUCTURE OF CHLAMYDOMONAS-REINHARDTII RIBULOSE 1 5-BISPHOSPHATE CARBOXYLASE-OXYGENASE ACTIVASE AND EVIDENCE FOR A SINGLE POLYPEPTIDE" PLANT PHYSIOLOGY (ROCKVILLE), vol. 94, no. 4, 1990, pages 1837-1841, XP002559554 ISSN: 0032-0889	1-8, 13-16
A	KINDLE KAREN L: "Amino-terminal and hydrophobic regions of the Chlamydomonas reinhardtii plastocyanin transit peptide are required for efficient protein accumulation in vivo" PLANT MOLECULAR BIOLOGY, vol. 38, no. 3, October 1998 (1998-10), pages 365-377, XP002559555 ISSN: 0167-4412	1-8, 13-16
A	LEON R ET AL: "Metabolic engineering of ketocarotenoids biosynthesis in the unicelullar microalga Chlamydomonas reinhardtii" JOURNAL OF BIOTECHNOLOGY, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 130, no. 2, 15 June 2007 (2007-06-15), pages 143-152, XP025317922 ISSN: 0168-1656 [retrieved on 2007-05-25]	1-8, 13-16
A	SINGH B ET AL: "ARABIDOPSIS ACETOHYDROXYACID SYNTHASE EXPRESSED IN ESCHERICHIA-COLI IS INSENSITIVE TO THE FEEDBACK INHIBITORS" PLANT PHYSIOLOGY (ROCKVILLE), vol. 99, no. 3, 1992, pages 812-816, XP002559556 ISSN: 0032-0889	1-8, 13-16
A	WO 02/26995 A1 (SYNGENTA LTD [GB]; WARNER SIMON ANTHONY JAMES [GB]; HAWKES TIMOTHY ROB) 4 April 2002 (2002-04-04)	9-12, 17-18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2009/056165

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-8, 13-16(all partially)

   A plant cell having incorporated into its genome a nucleotide sequence encoding a CTP, wherein said nucleotide sequence encoding said CTP is operably linked to a nucleotide of interest, and wherein said nucleotide sequence encoding said CTP has the nucleotide sequence of SEQ ID NO:1,2 and encodes an amino acid sequence set forth in SEQ ID NO:3, plants and seeds derived from said plants, as well as methods for the expression of said nucleotide sequence in a plant.

   ---

2. claims: 1-8, 13-16(all partially)

   idem for the CTP according to SEQ ID NO:4 and 5

   ---

3. claims: 9-12, 17-18

   idem for the CTP according to SEQ ID NO:6 and 7 and monocot plant cells

   ---

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2009/056165

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008039450 A2	03-04-2008	US 2008176304 A1	24-07-2008
WO 0226995 A1	04-04-2002	AR 030813 A1	03-09-2003
		AU 8786201 A	08-04-2002
		AU 2001287862 B2	14-12-2006
		BR 0114322 A	15-06-2004
		CN 1541270 A	27-10-2004
		EP 1325136 A1	09-07-2003
		JP 2004528808 T	24-09-2004
		MX PA03002641 A	19-06-2003
		US 2003200560 A1	23-10-2003