



(86) **Date de dépôt PCT/PCT Filing Date:** 2012/12/04
(87) **Date publication PCT/PCT Publication Date:** 2014/06/12
(45) **Date de délivrance/Issue Date:** 2019/06/25
(85) **Entrée phase nationale/National Entry:** 2015/06/03
(86) **N° demande PCT/PCT Application No.:** US 2012/067788
(87) **N° publication PCT/PCT Publication No.:** 2014/088560
(30) **Priorité/Priority:** 2012/12/04 (US13/693,585)

(51) **Cl.Int./Int.Cl. C12N 15/55** (2006.01),
C12N 1/13 (2006.01), **C12N 15/113** (2010.01),
C12N 15/62 (2006.01), **C12N 15/79** (2006.01),
C12N 15/87 (2006.01), **C12N 9/16** (2006.01)
(72) **Inventeurs/Inventors:**
SCHNEIDER, JANE C., US;
LIEBERMAN, SOYAN, US;
LIU, BO, US;
MOELLERING, ERIC R., US;
VERRUTO, JOHN H., US;
EDWARDS, AMANDA, US
(73) **Propriétaire/Owner:**

(54) **Titre : PROMOTEURS DE TETRASELMIS ET TERMINATEURS DESTINES A ETRE UTILISES DANS DES CELLULES
EUCARYOTES**
(54) **Title: TETRASELMIS PROMOTERS AND TERMINATORS FOR USE IN EUKARYOTIC CELLS**

(57) **Abrégé/Abstract:**

The present invention provides novel promoter and terminator sequences for use in gene expression in eukaryotic cells, such as algal cells. The invention further provides expression cassettes comprising a promoter, as described herein, operably linked to a gene. The invention further provides expression vectors and host eukaryotic cells, such as algal cells, for expressing a protein encoded by the gene; and methods for stably transforming eukaryotic algae such as Tetraselmis with transgenes.

(73) **Propriétaires(suite)/Owners(continued):**EXXONMOBIL RESEARCH AND ENGINEERING COMPANY, US

(74) **Agent:** BORDEN LADNER GERVAIS LLP

- (51) **International Patent Classification:**
C12N 15/113 (2010.01) *C12N 15/79* (2006.01)
- (21) **International Application Number:**
PCT/US2012/067788
- (22) **International Filing Date:**
4 December 2012 (04.12.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
13/693,585 4 December 2012 (04.12.2012) US
- (71) **Applicant:** EXXONMOBIL RESEARCH AND ENGINEERING COMPANY [—/US]; 1545 Route 22 East, P.O. Box 900, Annandale, NJ 08801-0900 (US).
- (72) **Inventors:** SCHNEIDER, Jane, C.; 2621 32nd St., San Diego, CA 92104 (US). LIEBERMAN, Soyan; 629 Barbara Ave., Solana Beach, CA 92075 (US). LIU, Bo; 9270 Pipilo St., San Diego, CA 92129 (US). MOELLERING, Eric, R.; 3216 Caminto Eastbluff #6s, La Jolla, CA 92037 (US). VERRUTO, John, H.; 4065 Georgia St., San Diego, CA 92103 (US). SKAGGS, Amanda; 8148 Genesee Ave. #141, San Diego, CA 92122 (US).
- (74) **Agent:** RAKERS, Leanne, M.; Harness, Dickey & Pierce, P.L.C., 7700 Bonhomme Avenue, Suite 400, St. Louis, MO 63105 (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, BN, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
- with international search report (Art. 21(3))
 - with sequence listing part of description (Rule 5.2(a))

(54) **Title:** TETRASELMIS PROMOTERS AND TERMINATORS FOR USE IN EUKARYOTIC CELLS

(57) **Abstract:** The present invention provides novel promoter and terminator sequences for use in gene expression in eukaryotic cells, such as algal cells. The invention further provides expression cassettes comprising a promoter, as described herein, operably linked to a gene. The invention further provides expression vectors and host eukaryotic cells, such as algal cells, for expressing a protein encoded by the gene; and methods for stably transforming eukaryotic algae such as *Tetraselmis* with transgenes.

WO 2014/088560 A1

**TETRASELMIS PROMOTERS AND TERMINATORS
FOR USE IN EUKARYOTIC CELLS**

REFERENCE TO A SEQUENCE LISTING

[0001]

FIELD

[0002] The present invention relates to gene regulatory elements for use in heterologous gene expression in eukaryotic cells, particularly algal cells.

BACKGROUND

[0003] Algal cells are a promising source of biofuels (Wijffels & Barbosa (2010) *Science* 329:796–99). Their ability to harness solar energy to convert carbon dioxide into carbon-rich lipids already exceeds the abilities of oil-producing agricultural crops, with the added advantage that algae grown for biofuel do not compete with oil-producing crops for agricultural land (Wijffels & Barbosa, 2010). *Tetraselmis* phytoplankton are unicellular marine green algae that can be cultured easily, rapidly, and economically. In order to maximize algal fuel production, new algal strains will need to be engineered for growth and carbon fixation at an industrial scale (Wijffels & Barbosa, 2010).

[0004] Modern recombinant strain development requires robust and efficient tools for expressing transgenes to alter cellular metabolism and physiology in desired ways. An essential component of any “toolkit” is a suite of functional promoter elements to drive transgene-expression. As such, there is a need for endogenous promoters, cloned and verified, from the strains for which recombinant DNA technology is being developed. There is a comparative lack of knowledge relating to techniques suitable for transforming *Tetraselmis* cells with transgenes, particularly of constitutive promoters and terminators derived from *Tetraselmis* species.

[0005] Transformation of *Tetraselmis subcordiformis* by glass bead treatment of protoplasts has been reported. See Cui *et al.* (2010) *World J. Microbiol. Biotechnol.* 26:1653–57. However, this method demonstrated only transient detection of the transformed gene—a green fluorescent protein controlled by a cytomegalovirus promoter—and did not make use of a selectable marker.

[0006] Further, while it will be necessary to manipulate the *Tetraselmis* genome in order to maximize biofuel output, this has proven difficult to date because of the thick cell wall that surrounds *Tetraselmis* cells. Better procedures for genetic manipulation of these organisms are urgently needed.

SUMMARY

[0007] This section provides a general summary of the disclosure, and is not a comprehensive disclosure of its full scope or all of its features.

[0008] Stable expression of genes, *e.g.* heterologous genes, requires the use of gene regulatory elements such as promoters and terminators. Novel promoters and terminators for use in driving gene expression in eukaryotic species such as but not limited to heterokont, fungal, and microalgal species, including *Tetraselmis*, are provided herein. Transformed heterokont, fungal, or algal cells can be used, for example, for synthesis of various products including lipids. In one aspect, an isolated DNA molecule is provided in which the isolated DNA molecule comprises a sequence selected from the group consisting of a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:1; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:2; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:3; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:4; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:5; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:6; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:7; or a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:8. Additionally, the isolated DNA molecule having at least 80% identity to

at least 100 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8 can have promoter activity in a eukaryotic cell.

[0009] For example, an isolated DNA molecule as provided herein can include any or any combination of: a nucleic acid sequence that has at least 85% or at least 90% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:1, a nucleic acid sequence that has at least 85% or at least 90% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:2, a nucleic acid sequence that has at least 85% or at least 90% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:3, a nucleic acid sequence that has at least 85% or at least 90% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:4, a nucleic acid sequence that has at least 85% or at least 90% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:5, a nucleic acid sequence that has at least 85% or at least 90% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:6, a nucleic acid sequence that has at least 85% or at least 90% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:7, a nucleic acid sequence that has at least 85% or at least 90% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:8. The isolated DNA molecule having at least 85% or at least 90% identity to at least 100 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8 can have promoter activity, for example, in an algal, fungal, or heterokont cell.

[0010] In some examples, an isolated DNA molecule is provided where the isolated DNA molecule includes a nucleic acid sequence that has at least 95% identity to at least 100 contiguous nucleotides of SEQ ID NO:1, a nucleic acid sequence that has at least 95% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:2, a nucleic acid sequence that has at least 95% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:3, a nucleic acid sequence that has at least 95% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:4, a nucleic acid sequence that has at least 95% nucleotide sequence identity to at least 100

contiguous nucleotides of SEQ ID NO:5, a nucleic acid sequence that has at least 95% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:6, a nucleic acid sequence that has at least 95% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:7, or a nucleic acid sequence that has at least 95% nucleotide sequence identity to at least 100 contiguous nucleotides of SEQ ID NO:8. The DNA molecule can have promoter activity, for example, in an algal, fungal, or heterokont cell.

[0011] In some examples an isolated DNA molecule as provided herein can be selected from the group consisting of: an isolated DNA molecule that comprises a nucleotide sequence comprising at least 100 contiguous nucleotides of SEQ ID NO:1; an isolated DNA molecule that comprises a nucleotide sequence comprising at least 100 contiguous nucleotides of SEQ ID NO:2; an isolated DNA molecule that comprises a nucleotide sequence comprising at least 100 contiguous nucleotides of SEQ ID NO:3; an isolated DNA molecule that comprises a nucleotide sequence comprising at least 100 contiguous nucleotides of SEQ ID NO:4; an isolated DNA molecule that comprises a nucleotide sequence comprising at least 100 contiguous nucleotides of SEQ ID NO:5; an isolated DNA molecule that comprises a nucleotide sequence comprising at least 100 contiguous nucleotides of SEQ ID NO:6; an isolated DNA molecule that comprises a nucleotide sequence comprising at least 100 contiguous nucleotides of SEQ ID NO:7; and an isolated DNA molecule that comprises a nucleotide sequence comprising at least 100 contiguous nucleotides of SEQ ID NO:8. The DNA molecule can have promoter activity in a eukaryotic cell, for example, in an algal, fungal, or heterokont cell.

[0012] For example, an isolated DNA molecule as provided herein can comprise a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:2; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500

contiguous nucleotides of SEQ ID NO:3; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:4; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:5; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:6; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:7; or a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:8. The DNA molecule can have promoter activity in a eukaryotic cell, for example, in an algal, fungal, or heterokont cell.

[0013] In further examples provided herein is an isolated DNA molecule comprising a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:1; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:2; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:3; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:4; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:5; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:6; a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:7; or a nucleotide sequence having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:8. An isolated DNA molecule as disclosed herein can find use, for example, as a sequence that when operably linked to a nucleic acid sequence that encodes a polypeptide or that can be transcribed into a functional RNA can affect expression of the nucleic acid sequence encoding a polypeptide or functional RNA.

[0014] Also provided herein is a promoter comprising a nucleic acid sequence having at least 80% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8. For example, a promoter as provided herein can comprise a nucleic acid sequence having at least 85% or at least 90% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:2, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:3, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:4, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:5, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:6, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:7, or at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:8.

[0015] In some examples, a promoter can comprise a sequence having at least 95% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:2, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:3, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:4, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:5, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:6, at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:7, or at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:8. In various examples, a promoter as provided herein can comprise at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8. In additional examples, a promoter as provided herein can be selected

from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8.

[0016] A promoter as provided herein can be a constitutive promoter, and may be active in a host cell cultured under conditions in which one or more nutrients are deficient as well as in culture conditions in which nutrients are sufficient for proliferation and/or growth of the culture. For example, a promoter provided herein may mediate transcriptions of an operably linked nucleic acid sequence in nitrogen replete as well as in nitrogen deficient conditions.

[0017] Also provided is an isolated DNA molecule comprising a sequence selected from the group consisting of a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:9 or a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:10. The isolated DNA molecule can have terminator activity. For example, an isolated DNA molecule can comprise a nucleotide sequence having at least 85% or at least 90% to at least 100 contiguous nucleotides of SEQ ID NO:9 or SEQ ID NO:10. In some examples, an isolated DNA molecule can comprise a sequence having at least 95% to at least 100 contiguous nucleotides of SEQ ID NO:9 or SEQ ID NO:10. For example, an isolated DNA molecule as provided herein can comprise a nucleotide sequence having at least 80%, at least 85%, or at least 90% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:9 or SEQ ID NO:10. For example, an isolated DNA molecule as provided herein can be selected from the group consisting of: an isolated DNA molecule comprising a nucleotide sequence having at least 95% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:9 and an isolated DNA molecule comprising a nucleotide sequence having at least 95% identity to at least 100, at least 200, at least 300, or at least 400 contiguous nucleotides of SEQ ID NO:10. For example, an isolated DNA molecule as provided herein can be selected from the group consisting of: an isolated DNA molecule comprising at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:9 and an isolated DNA molecule comprising at least 100, at least 200, at least 300, or at least 400 contiguous nucleotides of SEQ ID

NO:10. The isolated DNA molecule can comprise a terminator. The DNA molecule can find use, for example, as a sequence that when operably linked to a nucleic acid sequence that encodes a polypeptide or that can be transcribed into a functional RNA can affect expression of the nucleic acid sequence encoding a polypeptide or functional RNA.

[0018] Also provided herein is an expression cassette. The expression cassette comprises a promoter as disclosed herein and a heterologous gene operably linked to the promoter. The gene can encode, for example, a polypeptide or a functional RNA. The expression cassette can further optionally comprise a nucleic acid sequence comprising a terminator sequence, such as but not limited to, any disclosed herein. In various examples, the expression cassette can include a heterologous gene that encodes (a) a protein associated with lipid biosynthesis, (b) a protein having lipolytic activity, (c) a protein that participates in photosynthesis, (d) a protein associated with carbon fixation, (e) a transporter protein, (f) a dehydrogenase, (g) a transcription factor, (h) a protein involved in cell signaling, or (i) a functional RNA. A functional RNA can be, for example, an antisense sequence, a micro RNA, a shRNA, or a ribozyme. The expression cassette can be provided in a vector, *e.g.*, an expression vector, which can optionally include one or more of an origin of replication, sequences mediating recombination into a host genome, or a selectable marker.

[0019] Also provided herein is an expression cassette that includes a promoter as provided herein operably linked to a selectable marker gene or reporter gene. For example, an expression cassette can include a promoter as disclosed herein operably linked to a gene encoding an antibiotic-resistance-conferring protein and in particular nonlimiting examples can comprise any of SEQ ID NOs:1–8.

[0020] Also provided herein is a vector for eukaryotic cell transformation that comprises an expression cassette as disclosed herein. Additionally, a vector herein can include any promoter disclosed herein operably linked to a selectable marker gene or reporter gene. The transformation or expression vector can further include a terminator, such as but not limited to any disclosed herein, operably linked to the selectable marker or reporter gene. The gene encoding a selectable marker can be a gene encoding a polypeptide that confers resistance to an antibiotic, a polypeptide that confers tolerance to

an herbicide, a gene encoding an auxotrophic marker, or any other gene product that can allow for selection of transformants. A gene encoding a reporter protein can, for example, encode a fluorescent protein or an enzyme that can produce a detectable product.

[0021] Also provided herein is a method for transforming a eukaryotic cell. The method comprises introducing a transformation or expression vector that includes a promoter as provided herein operably linked to a heterologous gene, a selectable marker, and/or a reporter gene; and selecting for a transformant. For example, the eukaryotic cell can be transformed by means of a biolistic procedure. The eukaryotic cell can be selected from the group consisting of fungi, heterokonts, algae, and plants, and in a particular embodiment is an algal cell.

[0022] Also provided herein is a method for co-transforming a eukaryotic cell. The method comprises introducing both a transformation or expression vector that includes a promoter as provided herein operably linked to a nucleic acid sequence encoding a selectable marker protein, and further includes a nucleic acid sequence encoding a polypeptide of interest, and selecting for a transformant. The nucleic acid sequence encoding a polypeptide of interest can encode any polypeptide, including, without limitation, a second selectable marker, a reporter protein, an enzyme, a structural protein, a transcriptional regulator, a transporter, *etc.* The nucleic acid sequence encoding a polypeptide can also optionally be linked to a promoter, such as but not limited to a promoter as disclosed herein. The eukaryotic cell can be co-transformed, for example, by means of a biolistic procedure, by electroporation, by the use of agrobacteria or other biological vectors, *etc.* The eukaryotic cell can be selected from the group consisting of fungi, heterokonts, algae, and plants, and in a particular embodiment is an algal cell.

[0023] For example, an algal cell that can be transformed or co-transformed with a vector as provided herein can be selected from the group consisting of species of *Achnanthes*, *Amphiprora*, *Amphora*, *Ankistrodesmus*, *Asteromonas*, *Boekelovia*, *Bolidomonas*, *Borodinella*, *Botrydium*, *Botryococcus*, *Bracteococcus*, *Chaetoceros*, *Carteria*, *Chlamydomonas*, *Chlorococcum*, *Chlorogonium*, *Chlorella*, *Chroomonas*, *Chrysosphaera*, *Cricosphaera*, *Cryptothecodinium*, *Cryptomonas*, *Cyclotella*, *Dunaliella*, *Ellipsoidon*, *Emiliana*, *Eremosphaera*, *Ernodesmius*, *Euglena*, *Eustigmatos*, *Franceia*,

Fragilaria, Gloeothamnion, Haematococcus, Halocafeteria, Heterosigma, Hymenomonas, Isochrysis, Lepocinclis, Micractinium, Monoraphidium, Nannochloris, Nannochloropsis, Navicula, Nechochloris, Nephrochloris, Nephroselmis, Nitzschia, Ochromonas, Oedogonium, Oocystis, Ostreococcus, Pavlova, Parachlorella, Pascheria, Pelagomonas, Phaeodactylum, Phagus, Picochlorum, Platymonas, Pleurochrysis, Pleurococcus, Prototheca, Pseudochlorella, Pseudoneochloris, Pseudostaurastrum, Pyramimonas, Pyrobotrys, Scenedesmus, Schizochlamydeella, Skeletonema, Spyrogyra, Stichococcus, Tetrachlorella, Tetraselmis, Thalassiosira, Tribonema, Vaucheria, Viridiella, Vischeria, and Volvox. In one example, the algal cell can be of the *Chlorophyceae* class, such as a species of *Chlamydomonas*, *Dunaliella*, *Hematococcus*, or *Scenedesmus*. In other examples, the algal cell can be of the *Trebouxiophyceae* class, such as a species of *Chlorella*, *Pseudochlorella*, or *Nannochloris*, for example. In some examples, the algal cell can be an algal species of the *Prasinophyceae* class, for example, an *Ostreococcus* or a *Tetraselmis* cell.

[0024] Also provided herein is a eukaryotic host cell transformed with an expression vector or transformation vector as provided herein. The eukaryotic host cell can be a microorganism such as a microalga, such as but not limited to for example, a *Tetraselmis* cell.

[0025] In a further aspect, provided herein is a nucleic acid molecule that comprises a nucleotide sequence encoding a chloroplast transit peptide. The chloroplast transit peptide can direct a polypeptide to which it is operably linked to the stroma of a chloroplast, such as the chloroplast of a microalga. Provided herein is a nucleic acid molecule that comprises a nucleotide sequence encoding a chloroplast transit peptide having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:36 or a functional fragment thereof. The nucleotide sequence encoding a chloroplast transit peptide as provided herein can be operably linked to a polypeptide-encoding nucleotide sequence such that on expression of the chloroplast transit peptide-polypeptide construct, the synthesized polypeptide is localized to the chloroplast of a eukaryotic photosynthetic organism, such as an algal chloroplast. Further provided is a nucleic acid molecule that comprises a nucleotide sequence encoding a chloroplast transit

peptide having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:36, operably linked to a nucleotide sequence encoding a polypeptide of interest.

[0026] In a related aspect, provided herein is a method for directing a polypeptide of interest to the chloroplast of a eukaryotic photosynthetic organism. The method includes transforming a cell of a eukaryotic photosynthetic organism with a construct that includes a nucleotide sequence encoding a chloroplast transit peptide having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:36 operably linked to a nucleotide sequence encoding a polypeptide of interest, and providing conditions under which the construct is expressed, whereby the polypeptide of interest is directed to the chloroplast of the of eukaryotic photosynthetic organism. The eukaryotic photosynthetic organism can be a plant or alga, and in a particular example is a microalga, such as any provided herein, and may be, for example, a microalga of the *Chlorophyte* phylum. The microalga can be, for example, a microalga of the class *Chlorophyceae*, *Trebouxiophyceae*, or *Prasinophyceae*, and in some examples is a *Tetraselmis* species.

[0027] Further areas of applicability will become apparent from the description provided herein. The description and specific examples in this summary are intended for purposes of illustration only and are not intended to limit the scope of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] The drawings described herein are for illustrative purposes only of selected embodiments, not all possible implementations, and are not intended to limit the scope of the present disclosure.

[0029] **Figure 1** is a plasmid map for p05320.

[0030] **Figure 2** is a diagram of the wild type parent (WT-105) genome sequence for the actin gene. Bent labeled arrows indicate primers used to amplify the upstream and downstream regions labeled “WT-105 actin P” and “WT-105 actin T”. The actin open reading frame is shown labeled as “WT-105 actin”.

[0031] **Figure 3** is a colony PCR screen for *Tetraselmis* transformation. PCR was performed on colonies of the Zeocin-resistant isolates (C2–C10), the wild type parent (wt), Tris EDTA buffer (TE), water (H₂O), or genomic DNA from the wild type parent (gDNA). The Na⁺/H⁺ transporter (upper panel) is a likely single-copy endogenous gene that was used as a sensitivity control. The *ble* gene (lower panel) is present on the p05320 plasmid. A 2-log ladder (NEB, cat. no. N3200S) is included as a size standard.

[0032] **Figure 4** is a Southern blot depicting results for putative PH-925 transformants obtained with PDS-1000 biolistics. 5 µg of HindIII/BglII digested genomic DNA was probed with a DIG-labeled PCR *ble* open reading frame product. Transformant strains are labeled according to the corresponding plasmid used in transformation and transformant colony number (e.g. p05322-B1). Different isolates carrying the same construct are designated with a dash plus a number (e.g., p05322-B1, -K1, etc.). The *ble* PCR product was added as a control. A DIG labeled marker was included with the corresponding sizes indicated on the left.

[0033] **Figure 5** is a Southern blot identifying integration of the *ble* gene in the *Tetraselmis* genome. An agarose gel was loaded with (1) genomic DNA from WT-105 and digested with KpnI/MfeI, (2) genomic DNA from the p05320-8 and digested with KpnI/MfeI, (3) DIG-labeled DNA Molecular Weight Marker II (DIG-Marker), and (lanes 4 & 5) two levels of PCR product amplified from p05320 (*ble* PCR positive control). The gel was blotted and the blot was probed with a DIG-labeled PCR product that hybridized to the *ble* gene.

[0034] **Figure 6** is a western blot to detect expression of the Ble protein. An 18% gel was loaded with protein from the wild type parent (WT-105), the Magic Mark XP Western Protein Standard (Protein marker), protein from the pSGI05320-8, and protein from an *E. coli* strain engineered to express Sh-Ble. Blots were probed with an antibody to Ble. The black arrow indicates the expected molecular weight of Ble.

[0035] **Figure 7** is a schematic diagram depicting constructs used to transform WT-105 for promoter testing.

[0036] **Figure 8** is a schematic diagram depicting constructs used to transform WT-105 for promoter testing on non-selectable targets.

[0037] **Figure 9** is a western blot (upper panel) analysis of putative SBPase over-expressing lines of PH-925 generated by transformation with p05481. 30 µg total proteins from each sample were separated on 12% SDS-PAGE gels, which were processed for western blotting using the anti-FLAG peptide antibody (anti-FLAG pAb). After chemiluminescent detection, the blot was stained (lower panel) with Coomassie brilliant blue (CBB stain) to assess relative levels of protein loaded. Lanes 4, 7, 8, 10, 12, and 14 show FLAG-tagged protein. A protein molecular weight marker (Magic Mark II, Invitrogen) was included and marker masses are indicated at left.

DETAILED DESCRIPTION

A. Definitions

[0038] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. In case of conflict, the present application including the definitions will control. Unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

[0039] Although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present invention, suitable methods and materials are described below. The materials, methods and examples are illustrative only and are not intended to be limiting. Other features and advantages of the invention will be apparent from the detailed description and from the claims.

[0040] To facilitate an understanding of the present invention, a number of terms and phrases are defined below.

[0041] As used in the present disclosure and claims, the singular forms “a,” “an,” and “the” include plural forms unless the context clearly dictates otherwise.

[0042] Wherever embodiments are described herein with the language “comprising,” otherwise analogous embodiments described in terms of “consisting of” and/or “consisting essentially of” are also provided.

[0043] The term “and/or” as used in a phrase such as “A and/or B” herein is intended to include “A and B”, “A or B”, “A”, and “B”.

[0044] The terms, “cells”, “cell cultures”, “cell line”, “recombinant host cells”, “recipient cells” and “host cells” as used herein, include the primary subject cells and any progeny thereof, without regard to the number of transfers. It should be understood that not all progeny are exactly identical to the parental cell (due to deliberate or inadvertent mutations or differences in environment); however, such altered progeny are included in these terms, so long as the progeny retain the same functionality as that of the originally transformed cell.

[0045] The term “gene” is used broadly to refer to any segment of nucleic acid molecule (typically DNA, but optionally RNA) that encodes a protein or that can be transcribed into a functional RNA. Genes may include sequences that are transcribed but are not part of a final, mature, and/or functional RNA transcript, and genes that encode proteins may further comprise sequences that are transcribed but not translated, for example, 5’ untranslated regions, 3’ untranslated regions, introns, *etc.* Further, genes may optionally further comprise regulatory sequences required for their expression, and such sequences may be *e.g.*, sequences that are not transcribed or translated. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

[0046] The term “nucleic acid” or “nucleic acid molecule” refers to, *e.g.*, DNA or RNA (*e.g.*, mRNA). The nucleic acid molecules can be double-stranded or single-stranded; single stranded RNA or DNA can be the coding (sense) strand or the non-coding (antisense) strand.

[0047] The terms “coding sequence” or “coding region” as used herein, refer to a region of a DNA sequence that can be transcribed to produce an mRNA transcript that can be translated into an amino acid sequence, *e.g.*, of a peptide or polypeptide, or an RNA transcript that can be translated into an amino acid sequence, *e.g.*, of a peptide or a polypeptide. The term “non-coding sequence” or “non-coding region” refers to (1) a region of a DNA sequence that, if transcribed, is not translated into an amino acid sequence (*e.g.*, introns, untranslated regions, *etc.*); or (2) a region of an RNA sequence that is not translated into amino acids. For simplicity and brevity, a sequence that “encodes a polypeptide” refers to a DNA sequence that can be transcribed and translated to produce the polypeptide or an RNA sequence that can be translated to produce the polypeptide, whereas a sequence that “encodes a functional RNA” refers to a DNA sequence that when transcribed produces a functional RNA molecule. An RNA molecule that encodes a polypeptide or functional RNA can be further processed prior to, or concomittant with, translation into a polypeptide or formation of the mature functional RNA.

[0048] A “functional RNA molecule” is an RNA molecule that can interact with one or more proteins or nucleic acid molecules to perform or participate in a structural, catalytic, or regulatory function that affects the expression or activity of a gene or gene product other than the gene that produced the functional RNA. A functional RNA can be, for example, a transfer RNA (tRNA), ribosomal RNA (rRNA), anti-sense RNA (asRNA), microRNA (miRNA), short-hairpin RNA (shRNA), small interfering RNA (siRNA), small nucleolar RNAs (snoRNAs), piwi-interacting RNA (piRNA), or a ribozyme.

[0049] A nucleic acid may be “derived from” an indicated source, which includes the isolation (in whole or in part) of a nucleic acid segment from an indicated source or the purification of a polypeptide from an indicated source. A nucleic acid molecule may also be derived from an indicated source by, for example, direct cloning, PCR amplification, or artificial synthesis from the indicated polynucleotide source or based on a sequence associated with the indicated polynucleotide source. Genes or nucleic acid molecules derived from a particular source or species also include genes or nucleic acid molecules having sequence modifications with respect to the source nucleic acid molecules. For

example, a gene or nucleic acid molecule derived from a source (*e.g.*, a particular referenced gene) can incur one or more mutations with respect to the source gene or nucleic acid molecule that are unintended or that are deliberately introduced, and if one or more mutations, including substitutions, deletions, or insertions, are deliberately introduced the sequence alterations can be introduced by random or targeted mutation of cells or nucleic acids, by amplification or other molecular biology techniques, or by chemical synthesis. A gene or nucleic acid molecule that is derived from a referenced gene or nucleic acid molecule that encodes a functional RNA or polypeptide can encode a functional RNA or polypeptide having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the referenced or source functional RNA or polypeptide, or to a functional fragment thereof. For example, a gene or nucleic acid molecule that is derived from a referenced gene or nucleic acid molecule that encodes a functional RNA or polypeptide can encode a functional RNA or polypeptide having at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the referenced or source functional RNA or polypeptide, or to a functional fragment thereof.

[0050] The term “isolated” nucleic acid, such as an isolated protein or nucleic acid as used herein, refers to a biomolecule removed from the context in which the biomolecule exists in nature. An isolated biomolecule can be, in some instances, partially or substantially purified. For example, an isolated nucleic acid molecule can be a nucleic acid sequence that has been excised from the chromosome, genome, or episome into which it is integrated in nature.

[0051] A “purified” nucleic acid molecule or nucleotide sequence, or protein or polypeptide sequence, is substantially free of cellular material and cellular components. The purified nucleic acid molecule or protein may be free of chemicals beyond buffer or solvent, for example. “Substantially free” is not intended to mean that other components beyond the novel nucleic acid molecules are undetectable.

[0052] The terms “naturally-occurring” and “wild-type” (WT) refer to a form found in nature. For example, a naturally occurring or wild-type nucleic acid molecule,

nucleotide sequence, or protein may be present in, and isolated from, a natural source, and is not intentionally modified by human manipulation.

[0053] “Exogenous nucleic acid molecule” or “exogenous gene” refers to a nucleic acid molecule or gene that has been introduced (“transformed”) into a cell. A transformed cell may be referred to as a recombinant cell, into which additional exogenous gene(s) may be introduced. A descendent of a cell transformed with a nucleic acid molecule is also referred to as “transformed” if it has inherited the exogenous nucleic acid molecule. The exogenous gene may be from a different species (and so “heterologous”), or from the same species (and so “homologous”), relative to the cell being transformed. An “endogenous” nucleic acid molecule, gene or protein is a native nucleic acid molecule, gene or protein as it occurs in, or is naturally produced by, the host.

[0054] Further, the term “exogenous” as used herein in the context of a gene or protein, refers to a gene or protein that is not derived from the host organism species.

[0055] The term “transgene” as used herein, refers to an exogenous gene, that is, a gene introduced into a microorganism or a progenitor by human intervention.

[0056] The term “ortholog” of a gene or protein as used herein refers to its functional equivalent in another species.

[0057] The term “native” is used herein to refer to nucleic acid sequences or amino acid sequences as they naturally occur in the host. The term “non-native” is used herein to refer to nucleic acid sequences or amino acid sequences that do not occur naturally in the host, or are not configured as they are naturally configured in the host. A nucleic acid sequence or amino acid sequence that has been removed from a host cell, subjected to laboratory manipulation, and introduced or reintroduced into a host cell is considered “non-native.” Synthetic or partially synthetic genes introduced into a host cell are “non-native.” Non-native genes further include genes endogenous to the host microorganism operably linked to one or more heterologous regulatory sequences that have been recombined into the host genome, or genes endogenous to the host organism that are in a locus of the genome other than that where they naturally occur.

[0058] The term “heterologous” when used in reference to a polynucleotide, a gene, a nucleic acid, a polypeptide, or an enzyme refers to a polynucleotide, gene, a nucleic acid, polypeptide, or an enzyme not derived from the host species. For example, “heterologous gene” or “heterologous nucleic acid sequence” as used herein, refers to a gene or nucleic acid sequence from a different species than the species of the host organism it is introduced into. When referring to a gene regulatory sequence or to an auxiliary nucleic acid sequence used for maintaining or manipulating a gene sequence (*e.g.* a 5’ untranslated region, 3’ untranslated region, poly A addition sequence, intron sequence, splice site, ribosome binding site, internal ribosome entry sequence, genome homology region, recombination site, *etc.*) or to a nucleic acid sequence encoding a protein domain or protein localization sequence, “heterologous” means that the regulatory or auxiliary sequence or sequence encoding a protein domain or localization sequence is from a different source than the gene with which the regulatory or auxiliary nucleic acid sequence or nucleic acid sequence encoding a protein domain or localization sequence is juxtaposed in a construct, genome, chromosome or episome. Thus, a promoter operably linked to a gene to which it is not operably linked to in its natural state (*i.e.* in the genome of a non-genetically engineered organism) is referred to herein as a “heterologous promoter,” even though the promoter may be derived from the same species (or, in some cases, the same organism) as the gene to which it is linked. Similarly, when referring to a protein localization sequence or protein domain of an engineered protein, “heterologous” means that the localization sequence or protein domain is derived from a protein different from that into which it is incorporated by genetic engineering.

[0059] The term “recombinant” or “engineered” nucleic acid molecule as used herein, refers to a nucleic acid molecule that has been altered through human intervention. As non-limiting examples, a cDNA is a recombinant DNA molecule, as is any nucleic acid molecule that has been generated by *in vitro* polymerase reaction(s), or to which linkers have been attached, or that has been integrated into a vector, such as a cloning vector or expression vector. As non-limiting examples, a recombinant nucleic acid molecule: 1) has been synthesized or modified *in vitro*, for example, using chemical or enzymatic techniques (for example, by use of chemical nucleic acid synthesis, or by use of enzymes for the replication, polymerization, exonucleolytic digestion,

endonucleolytic digestion, ligation, reverse transcription, transcription, base modification (including, *e.g.*, methylation), or recombination (including homologous and site-specific recombination)) of nucleic acid molecules; 2) includes conjoined nucleotide sequences that are not conjoined in nature, 3) has been engineered using molecular cloning techniques such that it lacks one or more nucleotides with respect to the naturally occurring nucleic acid molecule sequence, and/or 4) has been manipulated using molecular cloning techniques such that it has one or more sequence changes or rearrangements with respect to the naturally occurring nucleic acid sequence. As non-limiting examples, a cDNA is a recombinant DNA molecule, as is any nucleic acid molecule that has been generated by *in vitro* polymerase reaction(s), or to which linkers have been attached, or that has been integrated into a vector, such as a cloning vector or expression vector.

[0060] The term “recombinant protein” as used herein refers to a protein produced by genetic engineering. The terms “peptide,” “polypeptide” and “protein” are used interchangeably herein, although “peptide” may be used to refer to a polypeptide having no more than about 100 amino acids, or no more than about 60 amino acids. The nucleic acid sequences according to some embodiments of the present invention encode cyanobacterial transcription factor domain proteins.

[0061] When applied to organisms, the terms “transgenic” or “recombinant” or “engineered” or “genetically engineered” refer to organisms that have been manipulated by introduction of an exogenous or recombinant nucleic acid sequence into the organism. Non-limiting examples of such manipulations include gene knockouts, targeted mutations and gene replacement, promoter replacement, deletion, or insertion, as well as introduction of transgenes into the organism. For example, a transgenic microorganism can include an introduced exogenous regulatory sequence operably linked to an endogenous gene of the transgenic microorganism. Recombinant or genetically engineered organisms can also be organisms into which constructs for gene “knock down” have been introduced. Such constructs include, but are not limited to, RNAi, microRNA, shRNA, antisense, and ribozyme constructs. Also included are organisms whose genomes have been altered by the activity of meganucleases or zinc finger

nucleases. A heterologous or recombinant nucleic acid molecule can be integrated into a genetically engineered/recombinant organism's genome or, in other instances, not integrated into a recombinant/genetically engineered organism's genome. As used herein, "recombinant microorganism" or "recombinant host cell" includes progeny or derivatives of the recombinant microorganisms of the invention. Because certain modifications may occur in succeeding generations from either mutation or environmental influences, such progeny or derivatives may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0062] The term "expression cassette" as used herein, refers to a nucleic acid construct that encodes a protein or functional RNA (*e.g.* a tRNA, a short hairpin RNA, one or more microRNAs, a ribosomal RNA, *etc.*) operably linked to expression control elements, such as a promoter, and optionally, any or a combination of other nucleic acid sequences that affect the transcription or translation of the gene, such as, but not limited to, a transcriptional terminator, a ribosome binding site, a splice site or splicing recognition sequence, an intron, an enhancer, a polyadenylation signal, an internal ribosome entry site, *etc.*

[0063] "Regulatory sequence", "regulatory element", or "regulatory element sequence" refers to a nucleotide sequence located upstream (5'), within, or downstream (3') of a coding sequence or functional RNA-encoding sequence. Transcription of the coding sequence or functional RNA-encoding sequence and/or translation of an RNA molecule resulting from transcription of the coding sequence are typically affected by the presence or absence of the regulatory sequence. These regulatory element sequences may comprise promoters, cis-elements, enhancers, terminators, or introns. Regulatory elements may be isolated or identified from UnTranslated Regions (UTRs) from a particular polynucleotide sequence. Any of the regulatory elements described herein may be present in a chimeric or hybrid regulatory expression element. Any of the regulatory elements described herein may be present in a recombinant construct of the present invention.

[0064] The terms "promoter", "promoter region", or "promoter sequence" refer to a nucleic acid sequence capable of binding RNA polymerase to initiate transcription of a

gene in a 5' to 3' ("downstream") direction. A gene is "under the control of" or "regulated by" a promoter when the binding of RNA polymerase to the promoter is the proximate cause of said gene's transcription. The promoter or promoter region typically provides a recognition site for RNA polymerase and other factors necessary for proper initiation of transcription. A promoter may be isolated from the 5' untranslated region (5' UTR) of a genomic copy of a gene. Alternatively, a promoter may be synthetically produced or designed by altering known DNA elements. Also considered are chimeric promoters that combine sequences of one promoter with sequences of another promoter. Promoters may be defined by their expression pattern based on, for example, metabolic, environmental, or developmental conditions. A promoter can be used as a regulatory element for modulating expression of an operably linked transcribable polynucleotide molecule, *e.g.*, a coding sequence or functional RNA sequence. Promoters may contain, in addition to sequences recognized by RNA polymerase and, preferably, other transcription factors, regulatory sequence elements such as cis-elements or enhancer domains that affect the transcription of operably linked genes. An "algal promoter" is a native or non-native promoter that is functional in algal cells.

[0065] The term "constitutive" promoter as used herein, refers to a promoter that is active under most environmental and developmental conditions. A constitutive promoter is active regardless of external environment, such as light and culture medium composition. In some examples, a constitutive promoter is active in the presence and in the absence of a nutrient. For example, a constitutive promoter may be a promoter that is active (mediates transcription of a gene to which it is operably-linked) under conditions of nitrogen depletion as well as under conditions in which nitrogen is not limiting (nitrogen replete conditions). In contrast, an "inducible" promoter is a promoter that is active in response to particular environmental conditions, such as the presence or absence of a nutrient or regulator, the presence of light, *etc.*

[0066] The term "operably linked," as used herein, denotes a configuration in which a control sequence or localization sequence is placed at an appropriate position relative to a sequence that encodes a polypeptide or functional RNA such that the control sequence directs or regulates the expression or localization of the mRNA encoding the polypeptide,

the polypeptide, and/or the functional RNA). Thus, a promoter is in operable linkage with a nucleic acid sequence if it can mediate transcription of the nucleic acid sequence. When introduced into a host cell, an expression cassette that includes a control sequence can result in transcription of the gene to which it is operably linked and/or translation of an encoded RNA or polypeptide under appropriate conditions. Antisense or sense constructs that are not or cannot be translated are not excluded by this definition. In the case of both expression of transgenes and suppression of endogenous genes (*e.g.*, by antisense, or sense suppression) one of ordinary skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only substantially identical to a sequence of the gene from which it was derived. As explained herein, these substantially identical variants are specifically covered by reference to a specific nucleic acid sequence.

[0067] As used herein “attenuated” means reduced in amount, degree, intensity, or strength. Attenuated gene expression may refer to a significantly reduced amount and/or rate of transcription of the gene in question, or of translation, folding, or assembly of the encoded protein. As nonlimiting examples, an attenuated gene may be a mutated or disrupted gene (*e.g.*, a gene disrupted by partial or total deletion, insertional mutation, or meganuclease disruption), or a gene having decreased expression resulting from alteration of gene regulatory sequences. An attenuated gene may also be a gene that is targeted by a “gene knockdown” construct, such as, for example, a construct encoding an antisense RNA, a microRNA, a short hairpin RNA, or a ribozyme. In the case of both expression of transgenes and suppression of endogenous genes (*e.g.*, by antisense or sense suppression) one of ordinary skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only substantially identical to a sequence of the gene from which it was derived. As explained herein, these substantially identical variants are specifically covered by reference to a specific nucleic acid sequence.

[0068] The term “photosynthetic organism” as used herein is any prokaryotic or eukaryotic organism that can perform photosynthesis. Photosynthetic organisms include higher plants (*i.e.*, vascular plants), bryophytes, algae, and photosynthetic bacteria. The term “algae” includes, but is not limited to, a species of *Bacillariophyceae* (diatoms), *Bolidomonas*, *Chlorophyceae* (green algae), *Chrysophyceae* (golden algae),

Carophyceae, *Cyanophyceae* (cyanobacteria), *Eustigmatophyceae* (pico-plankton), *Glaucocystophytes*, *Pelagophytes*, *Phaeophyceae* (brown algae), *Prasinophyceae* (pico-plankton), *Raphidophytes*, *Rhodophyceae* (red algae), *Synurophyceae* and *Xanthophyceae* (yellow-green algae). The term “algae” includes microalgae. The term “microalgae” as used herein refers to microscopic, single-celled algae species including, but not limited to, eukaryotic single-celled algae of the *Bacillariophyceae*, *Chlorophyceae*, *Prasinophyceae*, *Trebouxiophyceae*, and *Eustigmatophyceae* classes. The term “photosynthetic bacteria” includes, but is not limited to, cyanobacteria, green sulfur bacteria, purple sulfur bacteria, purple non-sulfur bacteria, and green non-sulfur bacteria.

[0069] The term “selectable marker” or “selectable marker gene” as used herein includes any gene that confers a phenotype on a cell in which it is expressed to facilitate the selection of cells that are transfected or transformed with a nucleic acid construct of the invention. The term may also be used to refer to gene products that effectuate said phenotypes. Examples of selectable markers include:

- genes conferring resistance to antibiotics such as amikacin (*aphA6*), ampicillin (*amp^R*), blasticidin (*bls*, *bsr*, *bsd*), bleomicin or phleomycin (ZEOCIN™) (*ble*), chloramphenicol (*cat*), emetine (RBS14p or *cryI-I*), erythromycin (*ermE*), G418 (GENETICIN™) (*neo*), gentamycin (*aac3* or *aacC4*), hygromycin B (*aphIV*, *hph*, *hpt*), kanamycin (*nptII*), methotrexate (*DHFR mtx^R*), penicillin and other β -lactams (β -lactamases), streptomycin or spectinomycin (*aadA*, *spec/strep*), and tetracycline (*tetA*, *tetM*, *tetQ*);
- genes conferring tolerance to herbicides such as aminotriazole, amitrole, andrimid, aryloxyphenoxy propionates, atrazines, bipyridyliums, bromoxynil, cyclohexandione oximes dalapon, dicamba, diclfop, dichlorophenyl dimethyl urea (DCMU), difunone, diketonitriles, diuron, fluridone, glufosinate, glyphosate, halogenated hydrobenzonitriles, haloxyfop, 4-hydroxypyridines, imidazolinones, isoxasflutole, isoxazoles, isoxazolidinones, miroamide B, p-nitrodiphenylethers, norflurazon, oxadiazoles, m-phenoxybenzamides, N-phenyl imides, pinoxadin, protoporphyrionogen oxidase inhibitors, pyridazinones, pyrazolates, sulfonylureas, 1,2,4-triazol pyrimidine, triketones, or urea; acetyl

CoA carboxylase (ACCase); acetohydroxy acid synthase (*ahas*); acetolactate synthase (*als*, *csr1-1*, *csr1-2*, *imr1*, *imr2*), aminoglycoside phosphotransferase (*apt*), anthranilate synthase, bromoxynil nitrilase (*bxn*), cytochrome P450-NADH-cytochrome P450 oxidoreductase, dalapon dehalogenase (*dehal*), dihydropteroate synthase (*sul*), class I 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), class II EPSPS (*aroA*), non-class I/II EPSPS, glutathione reductase, glyphosate acetyltransferase (*gat*), glyphosate oxidoreductase (*gox*), hydroxyphenylpyruvate dehydrogenase, hydroxy-phenylpyruvate dioxygenase (*hppd*), isoprenyl pyrophosphate isomerase, lycopene cyclase, phosphinothricin acetyl transferase (*pat*, *bar*), phytoene desaturase (*crtI*), prenyl transferase, protoporphyrin oxidase, the *psbA* photosystem II polypeptide (*psbA*), and SMM esterase (*SulE*) superoxide dismutase (*sod*);

- genes that may be used in auxotrophic strains or to confer other metabolic effects, such as *arg7*, *his3*, *hisD*, *hisG*, *lysA*, *manA*, *metE*, *nit1*, *trpB*, *ura3*, *xylA*, a dihydrofolate reductase gene, a mannose-6-phosphate isomerase gene, a nitrate reductase gene, or an ornithine decarboxylase gene; a negative selection factor such as thymidine kinase; or toxin resistance factors such as a 2-deoxyglucose resistance gene.

[0070] A “reporter gene” is a gene encoding a protein that is detectable or has an activity that produces a detectable product. A reporter gene can encode a visual marker or enzyme that produces a detectable signal, such as *cat*, *lacZ*, *uidA*, *xylE*, an alkaline phosphatase gene, an α -amylase gene, an α -galactosidase gene, a β -glucuronidase gene, a β -lactamase gene, a horseradish peroxidase gene, a luciferin/luciferase gene, an R-locus gene, a tyrosinase gene, or a gene encoding a fluorescent protein, including but not limited to a blue, cyan, green, red, or yellow fluorescent protein, a photoconvertible, photoswitchable, or optical highlighter fluorescent protein, or any of variant thereof, including, without limitation, codon-optimized, rapidly folding, monomeric, increased stability, and enhanced fluorescence variants.

[0071] The term “terminator” or “terminator sequence” or “transcription terminator” as used herein refers to a regulatory section of genetic sequence that causes RNA polymerase to cease transcription.

[0072] The term “transformation” as used herein refers to the introduction of one or more exogenous nucleic acid sequences or polynucleotides into a host cell or organism by using one or more physical, chemical, or biological methods. Physical and chemical methods of transformation (*i.e.*, “transfection”) include, by way of non-limiting example, electroporation and liposome delivery. Biological methods of transformation (*i.e.*, “transduction”) include transfer of DNA using engineered viruses or microbes (*e.g.*, *Agrobacterium*).

[0073] The terms, “identical” or percent “identity”, in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window. The degree of amino acid or nucleic acid sequence identity can be determined by various computer programs for aligning the sequences to be compared based on designated program parameters. For example, sequences can be aligned and compared using the local homology algorithm of Smith & Waterman (1981) *Adv. Appl. Math.* 2:482–89, the homology alignment algorithm of Needleman & Wunsch (1970) *J. Mol. Biol.* 48:443–53, or the search for similarity method of Pearson & Lipman (1988) *Proc. Nat’l. Acad. Sci. USA* 85:2444–48, and can be aligned and compared based on visual inspection or can use computer programs for the analysis (for example, GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI).

[0074] The BLAST algorithm, described in Altschul *et al.* (1990) *J. Mol. Biol.* 215:403–10, is publicly available through software provided by the National Center for Biotechnology Information (*available at* <http://www.ncbi.nlm.nih.gov>). This algorithm identifies high scoring sequence pairs (HSPS) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as

the neighborhood word score threshold (Altschul *et al.*, 1990). Initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated for nucleotides sequences using the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. For determining the percent identity of an amino acid sequence or nucleic acid sequence, the default parameters of the BLAST programs can be used. For analysis of amino acid sequences, the BLASTP defaults are: word length (W), 3; expectation (E), 10; and the BLOSUM62 scoring matrix. For analysis of nucleic acid sequences, the BLASTN program defaults are word length (W), 11; expectation (E), 10; M=5; N=-4; and a comparison of both strands. The TBLASTN program (using a protein sequence to query nucleotide sequence databases) uses as defaults a word length (W) of 3, an expectation (E) of 10, and a BLOSUM 62 scoring matrix. *See*, Henikoff & Henikoff (1989) *Proc. Nat'l. Acad. Sci. USA* 89:10915–19.

[0075] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul (1993) *Proc. Nat'l. Acad. Sci. USA* 90:5873–87). The smallest sum probability (P(N)), provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, preferably less than about 0.01, and more preferably less than about 0.001.

[0076] Transcript abundance can be measured in terms of reads per kilobase of exon model per million mapped reads (“RPKM”) (Mortazavi *et al.* (2008) *Nature Methods* 5:621–28). RPKM for a gene is calculated by dividing the gene’s total number of exon

reads by the product of the number of mapped reads of that gene (in millions) multiplied by the exon length (in kilobases). A gene's total number of exon reads is the number of reads that have been mapped to a region in which an exon is annotated for the gene or across the boundaries of two exons or an intron and an exon for an annotated transcript of the gene. The mapped reads include all the reads uniquely mapped to the region of the gene as well as those of the reads which match in more places that have been allocated to the gene's region. Exon length is calculated as the sum of the lengths of all exons annotated for the gene. Each exon is included only once in this sum, even if it is present in more annotated transcripts for the gene. Partly overlapping exons count for their full length, even though they share the same region.

B. Nucleotide Sequences

[0077] Genes were identified and isolated from a *Tetraselmis* environmental isolate (designated "WT-105") as sources for promoter and terminator sequences that can find use in the expression of genes, such as but not limited to transgenes, in eukaryotic microorganisms. The method by which these new promoter and terminator sequences were discovered is described more fully in the examples herein. SEQ ID NOs:1–8 were identified as comprising promoters that were demonstrated to mediate expression of transgenes in WT-105, and SEQ ID NOs:9–10 were identified as comprising terminators functional in WT-105.

Table 1. Regulatory Sequence ID Numbers

Gene name	Plasmid	Promoter 3' UTR	Terminator 5' UTR
Actin	p05439	SEQ ID NO:1	SEQ ID NO:9
Glyceraldehyde 3-phosphate dehydrogenase	p05481	SEQ ID NO:2	SEQ ID NO:10
Oxygen evolving enhancer protein 1	p05482	SEQ ID NO:3	
Oxygen evolving enhancer protein 3	p05648	SEQ ID NO:5	
Photosystem II, reaction center W	p05605	SEQ ID NO:4	
40S Ribosomal protein S12	p05606	SEQ ID NO:6	

Photosystem I, light harvesting complex	p05607	SEQ ID NO:7
Reverse complement of PSI-LHC promoter	p05653	SEQ ID NO:8

[0078] Based on the demonstration that these sequences mediate expression of heterologous genes, isolated and recombinant DNA (nucleic acid) molecules are provided herein that correspond to SEQ ID NOs:1–10 and to nucleotide sequences having about 80% identity to at least 100 contiguous nucleotides to any one of SEQ ID NOs:1–10; also provided herein are isolated or recombinant nucleic acid molecules having at least 100 contiguous nucleotides from any one of SEQ ID NOs:1–10.

[0079] For example, an isolated DNA molecule as provided herein can include: a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:1; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:2; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:3; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:4; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:5; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:6; a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:7; or a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:8. In some examples, the isolated nucleic acid molecule can direct expression in a eukaryotic cell, such as but not limited to an algal, fungal, or heterokont cell, of a nucleic acid sequence to which it is operably linked, for example, the isolated nucleic acid molecule can direct expression of a nucleic acid sequence of a protein-encoding sequence or a sequence encoding a functional RNA.

[0080] In other examples, an isolated DNA molecule as provided herein can include a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:9 or a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NO:10. The isolated DNA molecule can mediate transcriptional termination of a gene to which it is operably linked.

[0081] Additionally or alternatively, an isolated DNA molecule as provided herein can include any of: a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:1; a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:2; a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:3; a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:4; a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:5; a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:6; a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:7; and a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:8

[0082] Additionally or alternatively, an isolated DNA molecule as provided herein can include any of: a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1; a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:2; a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:3; a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:4; a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:5; a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:6; a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:7; and a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:8 . The isolated DNA molecule can have promoter activity. For example the isolated DNA molecule can mediate or enhance expression of a heterologous nucleic acid sequence to which it is operably linked.

[0083] Additionally or alternatively, an isolated DNA molecule as provided herein can include a nucleotide sequence having at least 80% identity to SEQ ID NO:1; a nucleotide sequence having at least 80% identity to SEQ ID NO:2; a nucleotide sequence having at least 80% identity to SEQ ID NO:3; a nucleotide sequence having at least 80% identity to SEQ ID NO:4; a nucleotide sequence having at least 80% identity to SEQ ID NO:5; a nucleotide sequence having at least 80% identity to SEQ ID NO:6; a nucleotide sequence having at least 80% identity to SEQ ID NO:7; and a nucleotide sequence having at least 80% identity to SEQ ID NO:8. The isolated DNA molecule can have promoter activity. For example the isolated DNA molecule can mediate or enhance expression of a heterologous nucleic acid sequence to which it is operably linked.

[0084] Additionally or alternatively, an isolated DNA molecule as provided herein can have at least 85% or at least 90% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8,. In some examples, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1. Additionally or alternatively, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:2. Further additionally or alternatively, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:3. Further alternatively or additionally, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:4. Further additionally or alternatively, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:5. Additionally or alternatively, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:6. Further additionally or alternatively, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:7. Further additionally

or alternatively, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:8. The isolated DNA molecule can mediate or enhance expression of a heterologous nucleic acid sequence to which it is operably linked.

[0085] In another aspect, provided herein is an isolated DNA molecule comprising a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:9 and a nucleotide sequence having at least 80% identity to at least 200 contiguous nucleotides of SEQ ID NO:10. For example, an isolated DNA molecule can comprise a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:9, or a nucleotide sequence having at least 80% identity to at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:10, and can comprise a nucleotide sequence having at least 80% identity to SEQ ID NO:9 or a nucleotide sequence having at least 80% identity to SEQ ID NO:10. Further alternatively or additionally, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:9. Further alternatively or additionally, an isolated DNA molecule can have at least 95% sequence identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:10. The isolated DNA molecule can have terminator activity. For example, the isolated DNA molecule can mediate transcriptional termination of a gene to which is operably linked

[0086] Additionally or alternatively, an isolated DNA molecule as provided herein can include a nucleotide sequence that can have at least 95%, 96%, 97%, 98%, or 99% percent identity to at least 100 contiguous nucleotides to any one of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10.

[0087] Additionally or alternatively, an isolated DNA molecule can comprise at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1; at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:2; at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:3; at least 100, at least 200, at least 300,

at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:4; at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:5; at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:6; at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:7; at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:8; at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:9; or at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:10.

[0088] Additionally or alternatively, an isolated DNA molecule can comprise a nucleotide sequence having at least 95% identity to any of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10. For example, an isolated DNA molecule can comprise a nucleotide sequence having at least 96%, 97%, 98%, or 99% identity to any of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10. In further examples, an isolated DNA molecule can comprise any of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10.

[0089] The isolated DNA molecule can find use, for example, as a sequence that when operably linked to a nucleic acid sequence can affect expression of the nucleic acid sequence, which can comprise, for example, a sequence encoding a polypeptide or functional RNA. For example, the isolated DNA molecule can increase or decrease expression of a nucleic acid sequence (or a portion thereof) to which it is operably linked, or may mediate transcription of the operably-linked nucleic acid sequence (or a portion thereof) as a promoter. Methods for assessing the functionality of nucleotide sequences for promoter activity, as well as for enhancing or decreasing the activity of proximal promoters, are well-known in the art. For example, promoter function can be validated by confirming the ability of the putative promoter or promoter variant or fragment to drive expression of a selectable marker gene to which the putative promoter or promoter

fragment or variant is operably linked by detecting and, optionally, analyzing, resistant colonies after plating of cells transformed with the promoter construct on selective media.

[0090] Additionally or alternatively, promoter activity may be assessed by measuring the levels of RNA transcripts produced from a promoter construct, for example, using reverse transcription-polymerase chain reaction (RT-PCR, e.g., Watt *et al.* (2008) *PLoS ONE* 1:e1428), by detection of the expressed protein, or by *in vivo* assays that rely on an activity of the protein encoded by the transcribed sequence. For example, promoter activity can be assessed using chloramphenicol acetyltransferase (CAT) assays (where the heterologous sequence operably linked to the isolated nucleic acid molecule that comprises a putative promoter encodes chloramphenicol acetyltransferase, see, for example, Gerrish *et al.* (2000) *J. Biol. Chem.* 275:3485–92), luciferase assays, where the heterologous nucleic acid is a *lux* or *luc* gene, for example (*see, for example, Ferrante et al.* (2008) *PLoS ONE* 3:e3200), or *in vivo* assays using a fluorescent protein gene to determine the functionality of any of the sequences disclosed herein, including sequences of reduced size or having one or more nucleotide changes with respect to any of SEQ ID NOs 1–8. (*see, for example, Akamura et al.* (2011) *Anal. Biochem.* 412:159–64). Testing of sequence modifications, including deletions and base substitutions of the promoter-containing sequences using reporter constructs such as but not limited to those provided herein are well-known in the art (*see, for example, Quinn et al.* (2003) *Eukaryotic Cell* 2:995–1002; Ranjan *et al.* (2011) *J. Biotechnol.* 152:58–62; Gerrish *et al.*, 2000).

Promoters

[0091] Also provided herein are promoters comprising a nucleic acid sequence such as any described herein, for example, a nucleic acid sequence having at least 80%, at least 85%, at least 90%, or at least 95% identity to at least 100 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8. For example, a promoter as provided herein may include a nucleotide sequence that has at least 85% or at least 90% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:1, at least 85% or at least 90% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:2, at least 85% or at least 90% sequence identity to at least

100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:3, at least 85% or at least 90% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:4, at least 85% or at least 90% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:5, at least 85% or at least 90% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:6, at least 85% or at least 90% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:7, or at least 85% or at least 90% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:8.

[0092] Additionally or alternatively, a promoter can include a sequence that has at least 95% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:1, at least 95% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:2, at least 95% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:3, at least 95% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:4, at least 95% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:5, at least 95% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:6, at least 95% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:7, or at least 95% sequence identity to at least 100, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO:8.

[0093] Additionally or alternatively, a promoter as provided herein can be selected from the group consisting of an isolated DNA molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:1; an isolated DNA molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:2; an isolated DNA molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:3; an isolated DNA molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:4; an isolated DNA molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:5; an isolated DNA molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:6; an isolated DNA molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:7; and an isolated DNA molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:8.

[0094] A promoter provided herein can include a nucleotide sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 250, at least 300, at least 400, at least 500, at least 600, or at least 700 contiguous nucleotides of SEQ ID NO:1. For example, a promoter provided herein can have at least 95% sequence identity to SEQ ID NO:1.

[0095] Additionally or alternatively, a promoter provided herein can include a nucleotide sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 250, at least 300, at least 400, or at least 500, at least 750, at least 1000, at least 1200, at least 1300, at least 1400, or at least 1500 contiguous nucleotides of SEQ ID NO:2. For example, a promoter provided herein can have at least 95% sequence identity to SEQ ID NO:2.

[0096] Additionally or alternatively, a promoter provided herein can include a nucleotide sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 250, at least 300, at least 400, at least 500, at least 750, at least 1000, at least 1200, at least 1300, at least 1400, or at least 1500 contiguous nucleotides of SEQ ID NO:3. For example, a promoter provided herein can have at least 95% sequence identity to SEQ ID NO:3.

[0097] Additionally or alternatively, a promoter provided herein can include a nucleotide sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 250, at least 300, at least 400, at least 500, at least 600, at least 700, or at least 800 contiguous nucleotides of SEQ ID NO:4. For example, a promoter provided herein can have at least 95% sequence identity to SEQ ID NO:4.

[0098] Additionally or alternatively, a promoter provided herein can include a nucleotide sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 250, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 1000 contiguous nucleotides of SEQ ID NO:5. For example, a promoter provided herein can have at least 95% sequence identity to SEQ ID NO:5.

[0099] Additionally or alternatively, a promoter provided herein can include a nucleotide sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 250, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 950, or at least 1000 contiguous nucleotides of SEQ ID NO:6. For example, a promoter provided herein can have at least 95% sequence identity to SEQ ID NO:6.

[00100] Additionally or alternatively, a promoter provided herein can include a nucleotide sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 250, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 950 contiguous nucleotides of SEQ ID NO:7. For example, a promoter provided herein can have at least 95% sequence identity to SEQ ID NO:7.

[00101] Additionally or alternatively, a promoter provided herein can include a nucleotide sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 250, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 950 contiguous nucleotides of SEQ ID NO:8. For example, a promoter provided herein can have at least 95% sequence identity to SEQ ID NO:8.

[00102] In various examples of promoters of the present invention a promoter is selected from the group consisting of: SEQ ID NO:1; SEQ ID NO:2; SEQ ID NO:3; SEQ ID NO:4; SEQ ID NO:5; SEQ ID NO:6; SEQ ID NO:7; and SEQ ID NO:8.

[00103] A promoter as provided herein can be a constitutive promoter, for example the promoter may be active in culture conditions in which one or more nutrients are deficient as well as in culture conditions in which nutrients are sufficient for proliferation and/or growth of the culture. For example, a promoter as provided herein may direct expression of an operably linked nucleic acid sequence under conditions in which a host cell that includes the promoter construct is limited in nitrogen availability (nitrogen depletion/deficiency) as well as under conditions in which a host cell that includes the promoter construct is not limited in nitrogen availability (nitrogen replete conditions).

[00104] Without being bound by theory, promoters allow RNA polymerase to attach to DNA near a gene in order for transcription to take place. Promoters contain specific DNA sequences that provide transcription factors to an initial binding site from which they can recruit RNA polymerase binding. These transcription factors have specific protein motifs that enable them to interact with specific corresponding nucleotide sequences to regulate gene expressions. The minimal portion of the promoter required for proper transcription initiate include: (1) the Transcription Start Site (“TSS”) and elements directly upstream; (2) an RNA polymerase binding site; and (3) general transcription factor binding sites, *e.g.* a TATA box (sequence TATAAA).

[00105] A proximal promoter sequence may be approximately 250 basepairs (bp) upstream of the translational start site of the open reading frame of the gene and may contain, in addition to sequences for binding RNA polymerase, specific transcription factor binding sites. Some promoters also include a distal sequence upstream of the gene that may contain additional regulatory elements, often with a weaker influence than the proximal promoter. Eukaryotic transcriptional complexes can bend the DNA back on itself, thus allowing for potential placement of additional regulatory sequences as far as several kilobases from the TSS. Many eukaryotic promoters contain a TATA box. The TATA box binds the TATA binding protein, which assists in the formation of the RNA polymerase transcriptional complex. TATA boxes usually lie within approximately 50 bp of the TSS. A promoter may be constitutive or expressed conditionally. Some promoters are inducible, and may activate or increase transcription in response to an inducing agent. In contrast, the rate of transcription of a gene under control of a constitutive promoter is not dependent on an inducing agent. A constitutive promoter can be made a conditional or inducible promoter by the addition of sequences that confer responsiveness to particular conditions or to an inducing agent. Thus, promoters provided herein may be constitutive or may be inducible or conditional. Further, promoters or portions of promoters may be combined in series to achieve a stronger level of expression or a more complex pattern of regulation.

[00106] In various examples, a promoter as provided herein, such as but not limited to a promoter that comprises a nucleotide sequence having at least 80%, at least 85%, at

least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8, can mediate transcription of an operably linked nucleic acid sequence in a eukaryotic algal cell, such as, for example, a *Tetraselmis* cell. In some instances, a promoter as provided herein can mediate transcription of an operably linked nucleic acid sequence in a eukaryotic cell, such as but not limited to a eukaryotic algal cell, during culturing of the cell under conditions of nitrogen depletion as well as during culturing of the cell under nitrogen replete conditions. For example, a promoter as described herein can preferably mediate transcription of an operably linked nucleic acid sequence in *Tetraselmis* cells cultured under conditions of nitrogen depletion or cultured under nitrogen replete conditions.

[00107] Additionally, as contemplated herein, a promoter or promoter region can include variants of the promoters disclosed herein derived by deleting sequences, duplicating sequences, or adding sequences from other promoters or as designed, for example, by bioinformatics, or by subjecting the promoter to random or site-directed mutagenesis, *etc.*

[00108] Any of the nucleic acid molecules described herein may comprise nucleic acid sequences comprising promoters. For example, promoters of the present invention can include nucleic acid sequences having at least 80%, at least 85%, at least 90%, at least 95%, or between 95% and 100% identity to the sequences between about 0 bp, 20 bp, 50 bp, 100 bp, 200 bp or 300 bp to about 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, or 1 kb upstream of the trinucleotide ATG sequence at the start site of a protein coding region of a native *Tetraselmis* gene, such as, for example, an actin gene, a glyceraldehyde 3-phosphate dehydrogenase gene, an oxygen evolving enhancer protein 1 (*psbO*) gene, a photosystem II reaction center W gene, an oxygen evolving enhancer protein 3 gene, a 40S ribosomal protein S12 gene, or a photosystem I light harvesting complex (PSI-LHC) gene.

[00109] Additionally or alternatively, promoters of the present invention can include nucleic acid sequences having at least 80%, at least 85%, at least 90%, at least 95%, or

between 95% and 100% identity to the reverse complement of sequences between about 0 bp, 20 bp, 50 bp, 100 bp, 200 bp or 300 bp to about 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, or 1 kb upstream of the trinucleotide ATG sequence at the start site of a protein coding region of a native *Tetraselmis* gene, such as, a photosystem I light harvesting complex gene.

[00110] The activity or strength of a promoter may be measured in terms of the amount of RNA it produces, or the amount of protein accumulation in a cell or tissue, which can optionally be measured by an activity of the expressed protein, *e.g.*, fluorescence, luminescence, acyltransferase activity, *etc.*, relative to a promoter whose transcriptional activity has been previously assessed, relative to a promoterless construct, or relative to non-transformed cells. For example, the activity or strength of a promoter may be measured in terms of the amount of mRNA accumulated that corresponds to a nucleic acid sequence to which it is operably linked in a cell, relative to the total amount of mRNA or protein produced by the cell. The promoter preferably expresses an operably linked nucleic acid sequence at a level greater than 0.01%; preferably in a range of about 0.5% to about 20% (w/w) of the total cellular RNA. The activity can also be measured by quantifying fluorescence, luminescence, or absorbance of the cells or a product made by the cells or an extract thereof, depending on the activity of a reporter protein that may be expressed from the promoter. The activity or strength of a promoter may be expressed relative to a well-characterized promoter (for which transcriptional activity was previously assessed). For example, a less-characterized promoter may be operably linked to a reporter sequence (*e.g.*, a fluorescent protein) and introduced into a specific cell type. A well-characterized promoter is similarly prepared and introduced into the same cellular context. Transcriptional activity of the unknown promoter is determined by comparing the amount of reporter expression, relative to the well characterized promoter.

[00111] A promoter described herein can have promoter activity in a eukaryotic cell, preferably in an algal cell or heterokont cell. In a particular examples, a promoter as provided herein is active in an algal or heterokont cell in nutrient replete and nutrient-depleted culture conditions. An algal promoter as provided herein can be used as a 5'

regulatory element for modulating expression of an operably linked gene or genes in algal species as well as other organisms, including fungi, heterokonts, and plants.

[00112] Using promoter assay methods, such as but not limited to the method described in **Examples 2 and 3**, the disclosed promoter sequences can be further modified, *e.g.* truncated or mutated, and screened to refine the active promoter regions.

Terminators

[00113] In another embodiment of the present invention terminators are provided in which the terminators comprise a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 100 or at least 150 contiguous nucleotides of SEQ ID NO:9 or SEQ ID NO:10.

[00114] For example, a terminator can have at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 950 contiguous nucleotides of SEQ ID NO:9.

[00115] Additionally or alternatively, a terminator can have at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 1000 contiguous nucleotides of SEQ ID NO:10.

[00116] Terminators are genetic sequences that mark the end of a gene for transcription. Without being bound by theory, the terminators of the present invention may improve expression of a nucleic acid sequence (amount of encoded RNA or protein produced), and may mediate polyadenylation or enhance RNA transcript stability. Most terminator sequences in eukaryotes consist of at least two DNA sequences: (1) a binding site for terminator proteins and (2) an upstream element located among the last twelve nucleotides of the transcript. The protein binding sites are usually orientation-sensitive and essential to termination. Termination usually occurs between twelve and twenty nucleotides upstream of the binding site. The upstream element's functionality usually

depends more on its overall base composition (T-rich) than on the specific sequence (Reeder & Lang (1997) *Trends Biochem. Sci.* 22:473–77).

C. Expression Cassettes

[00117] Expression cassettes are also provided in the present invention, in which the expression cassettes comprise one or more regulatory elements as described herein to drive the expression of transgenes. These cassettes comprise isolated nucleic acid molecules that include any one of the promoter sequences described herein or any combination thereof, operably linked to a gene of interest, with the gene of interest positioned downstream of the promoter sequence, and optionally with any one of the terminator sequences described herein or any combination thereof operably linked downstream of the transgene. For example, any of the promoters listed in **Table 1** can be used in combination with any terminator listed in **Table 1** in an expression cassette.

[00118] The basic techniques for operably linking two or more sequences of DNA together are familiar to the skilled worker, and such methods have been described in a number of texts for standard molecular biological manipulation (*see, e.g., "Molecular Cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; Gibson et al. (2009) Nature Methods 6:343–45*).

[00119] The promoters of the invention can be used with any heterologous or homologous gene(s). A heterologous or homologous gene according to the invention may encode a protein or polypeptide. Alternatively, the heterologous or homologous gene can encode a functional RNA, such as a tRNA, rRNA, small nucleolar RNA (snoRNA), ribozyme, antisense RNA (asRNA), micro RNA (miRNA), short hairpin RNA (shRNA), small interfering RNA (siRNA), or piwi-interacting RNA (piRNA). Any known or later-discovered heterologous or homologous gene which encodes a desired product can be operably linked to a promoter sequence of the invention using known methods. Non-limiting examples of known genes suitable for use with the promoters of the invention include genes encoding proteins associated with lipid biosynthesis; proteins having

lipolytic activity; proteins associated with carbohydrate metabolism; transporter polypeptides; proteins conferring resistance to an antibiotic, herbicide, or toxin; reporter proteins (*e.g.*, fluorescent proteins or enzymes that produce detectable products) polypeptides of the Calvin-Benson cycle; polypeptides that participate in photosynthesis (such as but not limited to, photosynthetic reaction center polypeptides, light-harvesting chlorophyll-binding proteins, oxygen-evolving complex polypeptides, cytochromes, ferredoxins, *etc.*); dehydrogenases, such as NADPH-forming dehydrogenases; transcription factors; proteins involved in cell signaling (*e.g.*, G proteins or kinases); or functional RNAs.

[00120] For example, an expression cassette can comprise a promoter as described herein (for example, a promoter comprising a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides of SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, or 8) operably linked to a gene encoding a polypeptide, where the polypeptide can be any polypeptide of interest, and in illustrative and nonlimiting examples, can be a protein associated with lipid biosynthesis, an acetyl-CoA carboxylase, a malonyl type 1 fatty acid synthase, a type 2 fatty acid synthase subunit, a beta ketoacyl-ACP synthase, a malonyl-CoA-malonyl-ACP acyltransferase, an acyl-ACP thioesterase, an acyl-CoA thioesterase, a 4-hydroxybenzoyl thioesterase, an alcohol forming acyl reductase, a wax synthase, an aldehyde decarbonylase, a fatty acid decarboxylase, a lipase, a glyceraldehyde 3 phosphate dehydrogenase, an acyl-CoA synthetase, a phospholipid diacylglycerol acyltransferase, a glycerol 3 phosphate acyltransferase, a lysophosphatidic acid acyltransferase, a phosphatidic acid phosphatase, a diacyl glycerol acyltransferase, a polypeptide that participates in photosynthesis, a chlorophyll-binding light harvesting polypeptide, a photosynthetic reaction center polypeptide, an oxygen-evolving complex polypeptide, a cytochrome, a ferredoxin, a protein associated with carbon fixation, a ribulose biphosphate carboxylase subunit, a carbonic anhydrase, a transporter protein, an ABC transporter, a FatB transporter, a dehydrogenase, an aldehyde dehydrogenase, a 2-hydroxyacid dehydrogenase, an isocitrate dehydrogenase, 6 phosphogluconate dehydrogenase, glucose 6 phosphate dehydrogenase, a transcription factor, a kinase, or a G protein.

[00121] In further examples, an expression cassette can comprise a promoter as described herein (for example, a promoter comprising a nucleotide sequence having at least 80% identity to at least 100 contiguous nucleotides SEQ ID NO:1, 2, 3, 4, 5, 6, 7, or 8) operably linked to a gene encoding a functional RNA, optionally wherein the functional RNA is an antisense RNA, a small hairpin RNA, a microRNA, an siRNA, an snoRNA, a piRNA, or a ribozyme.

D. Vectors

[00122] The present invention also provides vectors that can comprise the regulatory elements and/or expression cassettes described herein. The vectors can further optionally comprise at least one origin of replication (“ORI”) sequence for replication in a cell. The vectors may further optionally comprise one or more selectable markers under the control of one or more eukaryotic promoters, one or more selectable markers under the control of one or more prokaryotic promoters, and/or one or more sequences that mediate recombination of an exogenous nucleic acid sequence into the target cell’s genome.

[00123] An ORI is the sequence in a DNA molecule at which replication begins. The ORI serves as a base of assembly for the pre-replication complex. Depending on the ORI, such replication can proceed unidirectionally or bidirectionally. An expression vector as provided herein can include an ORI for replication of the expression vector in a cloning host, such as *E. coli* or *Saccharomyces*, and/or can include an ORI for replication of the expression vector in a target cell, which can be, for example, an algal or heterokont cell. The structural biology of ORIs is widely conserved among prokaryotes, eukaryotes, and viruses. Most ORIs possess simple tri-, tetra-, or higher nucleotide repetition patterns. Most are AT-rich and contain inverted repeats. Those skilled in the art will be familiar with the more common ORIs, such as p15A and the pUC ORI.

[00124] Additionally, a vector described herein may also carry a selectable marker. By way of example, a vector that includes an expression cassette may include, as a selectable marker, a gene conferring resistance to a poison, such as an antibiotic, a herbicide, or some other toxin, so that transformants can be selected by exposing the cells to the poison

and selecting those cells which survive the encounter. Non-limiting examples of selectable markers include:

- genes conferring resistance to antibiotics such as amikacin (*aphA6*), ampicillin (*amp^R*), blasticidin (*bls*, *bsr*, *bsd*), bleomicin or phleomycin (ZEOCIN™) (*ble*), chloramphenicol (*cat*), emetine (RBS14p or *cryI-1*), erythromycin (*ermE*), G418 (GENETICIN™) (*neo*), gentamycin (*aac3* or *aacC4*), hygromycin B (*aphIV*, *hph*, *hpt*), kanamycin (*nptII*), methotrexate (*DHFR mtx^R*), penicillin and other β -lactams (β -lactamases), streptomycin or spectinomycin (*aadA*, *spec/strep*), and tetracycline (*tetA*, *tetM*, *tetQ*);
- genes conferring resistance or tolerance to herbicides such as aminotriazole, amitrole, andrimid, aryloxyphenoxy propionates, atrazines (*psbA*), bipyridyliums, bromoxynil, cyclohexandione oximes dalapon, dicamba, diclufop, dichlorophenyl dimethyl urea (DCMU), difunone, diketonitriles, diuron, fluridone, glufosinate, glyphosate, halogenated hydrobenzonnitriles, haloxyfop, 4-hydroxypyridines, imidazolinones, isoxasflutole, isoxazoles, isoxazolidinones, miroamide B, p-nitrodiphenylethers, norflurazon, oxadiazoles, m-phenoxybenzamides, N-phenyl imides, pinoxadin, protoporphyrionogen oxidase inhibitors, pyridazinones, pyrazolates, sulfonylureas, 1,2,4-triazol pyrimidine, triketones, urea compounds; acetyl CoA carboxylase (ACCase), acetohydroxy acid synthase (*ahas*), acetolactate synthase (*als*, *csr1-1*, *csr1-2*, *imr1*, *imr2*), aminoglycoside phosphotransferase (*apt*), anthranilate synthase, bromoxynil nitrilase (*bxn*), cytochrome P450-NADH-cytochrome P450 oxidoreductase, dalapon dehalogenase (*dehal*), dihydropteroate synthase (*sul*), class I 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), class II EPSPS (*aroA*), non-class I/II EPSPS, glutathione reductase, glyphosate acetyltransferase (*gat*), glyphosate oxidoreductase (*gox*), hydroxyphenylpyruvate dehydrogenase, hydroxy-phenylpyruvate dioxygenase (*hppd*), isoprenyl pyrophosphate isomerase, lycopene cyclase, phosphinothricin acetyl transferase (*pat*, *bar*), phytoene desaturase (*crtI*), prenyl transferase, protoporphyrin oxidase, *psbA* of photosystem II (*psbA*), and SMM esterase (*SulE*) superoxide dismutase (*sod*);

- genes that may be used in auxotrophic strains or to confer autotrophic growth or other metabolic effects, such as *arg7*, *his3*, *hisD*, *hisG*, *lysA*, *manA*, *metE*, *nit1*, *trpB*, *ura3*, *xylA*, a dihydrofolate reductase gene, a mannose-6-phosphate isomerase gene, a nitrate reductase gene, or an ornithine decarboxylase gene; a negative selection factor such as thymidine kinase; or toxin resistance factors such as a 2-deoxyglucose resistance gene.

[00125] The selectable marker gene can be operably linked to and/or under the control of a promoter as provided herein. The promoter regulating expression of the selectable marker may be conditional or inducible but is preferably constitutive, and can be, for example, any promoter disclosed herein or another promoter. Alternatively, the selectable marker may be placed under the control of the expression cassette promoter. If a selectable marker is placed under the control of the expression cassette promoter, the selectable marker and the expression cassette may be operably linked with an internal ribosome entry site (“IRES”) element between the expression cassette and the selectable marker (Komar & Hatzoglou (2011) *Cell Cycle* 10:229–40 and Hellen & Sarnow (2001) *Genes & Dev.* 15:1593–612) or a “2A” sequence (Kim *et al.* (2011) *PLoS One* 6:e18556).

[00126] Further provided herein is a vector for transformation of a eukaryotic cell, such as but not limited to a eukaryotic microalgal cell or phytoplankter cell, in which the vector includes a selectable marker gene operably linked to a promoter as provided herein, for example, a promoter that includes a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, or at least 800 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8, or a promoter that comprises SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8. The transformation vector can further include one or more additional genes or constructs for transfer into the host cell, such as a gene encoding a polypeptide

such as but not limited to any disclosed hereinabove or a construct encoding a functional RNA, where the gene encoding a polypeptide or functional RNA can optionally be operably linked to a promoter as described herein, or can optionally be operably linked to another promoter.

[00127] Additionally or alternatively, the vectors as provided herein may comprise a terminator as provided herein. For example, a vector of the present invention may comprise a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% sequence identity to at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, or at least 800 contiguous nucleotides of SEQ ID NO:9 or SEQ ID NO:10. A gene of interest or a selectable marker gene on a vector of the present invention may be operably linked to a terminator sequence as provided herein. For example, a gene of interest or a selectable marker may be operably linked to SEQ ID NO:9. By way of further example, a gene of interest or a selectable marker gene may be operably linked to SEQ ID NO:10.

[00128] In an alternative transformation strategy, a selectable marker operably linked to a promoter such as a promoter described herein can be provided on a separate construct, where both the gene-of-interest construct and the selectable marker construct are used together in transformation protocols. Selected transformants are then analyzed for co-transformation of the construct that includes the gene-of-interest (*see, e.g.,* Kindle (1990) *Proc. Nat'l. Acad. Sci. USA* 87:1228–32).

[00129] If a vector as provided herein that includes an expression cassette lacks a selectable marker gene, transformants may be selected by routine methods familiar to those skilled in the art, such as, by way of a non-limiting example, extracting nucleic acid from the putative transformants and screening by PCR. Alternatively or in addition, transformants may be screened by detecting expression of a reporter gene, such as but not limited to a chloramphenicol acyltransferase gene (*cat*) *lacZ*, *uidA*, *xylE*, an alkaline phosphatase gene, an α -amylase gene, an α -galactosidase gene, a β -lactamase gene, a β -glucuronidase gene, a horseradish peroxidase gene, a luciferin/luciferase gene, an R-locus gene, a tyrosinase gene, or a gene encoding a fluorescent protein, such as any of the blue,

cyan, green, red, yellow, photoconvertible, or photoswitchable fluorescent proteins or any of their variants, including codon-optimized, rapidly folding, monomeric, increased stability, and enhanced fluorescence variants. A reporter gene used in a vector may optionally be regulated by a promoter as provided herein. A transformation vector may include a gene encoding a reporter, such as, for example, a fluorescent protein, operably linked to a promoter as provided herein.

[00130] In some examples, the vector is designed for integration of one or more genes (such as the expression cassette) into the host genome. For example, the expression vectors may include *Agrobacterium* flanking sequences designed for integrating transgenes into the genome of a target plant cell. In other embodiments, vectors can be targeted for integration into a plant or algal chromosome by including flanking sequences that enable homologous recombination into the chromosome or targeted for integration into endogenous host plasmids by including flanking sequences that enable homologous recombination into the endogenous plasmids. In some cases in which it may be advantageous to transform the chloroplast of a higher plant or alga, the expression vectors can be designed to have regions of sequences flanking the transgene that are homologous to chloroplast sequences to promote homologous recombination and integration of the sequence of interest. Further, a transformation vector can include sequences for site-specific recombination such as but not limited to *lox* sites on which the Cre recombinase acts.

[00131] In addition to the promoters provided herein, one skilled in the art would know various promoters, introns, enhancers, transit peptides, targeting signal sequences, 5' and 3' untranslated regions (UTRs), IRES, 2A sequences, and terminator sequences, as well as other molecules involved in the regulation of gene expression that are useful in the design of effective expression vectors. In some embodiments, the expression vector will contain one or more enhancer elements. Enhancers are short regions of DNA that can bind trans-acting factors to enhance transcription levels. Although enhancers usually act in *cis*, an enhancer need not be particularly close to its target gene, and may sometimes not be located on the same chromosome. Enhancers can sometimes be located in introns.

[00132] In some examples, a gene or genes encoding enzymes that participate in the synthesis of a fatty acid product (*e.g.*, a fatty acid, a fatty acid derivative, or a glycerolipid) is cloned into the vector as an expression cassette that includes a promoter as disclosed herein. The expression cassette may optionally include a transit peptide-encoding sequence for directing the expressed enzyme to the chloroplast or endoplasmic reticulum of transformed eukaryotic cells, an intron sequence, a sequence having a polyadenylation signal, *etc.* Additionally or alternatively, a vector is provided comprising an expression cassette as described herein, wherein the vector further comprises one or more of: a selectable marker gene, an origin of replication, and one or more sequences for promoting integration of the expression cassette into the host genome. Additionally or alternatively, a vector is provided comprising an isolated or recombinant nucleic acid molecule as described herein, wherein the isolated nucleic acid molecule is operably linked to a nucleic acid sequence encoding a selectable marker or a reporter protein, such as, for example, any described herein. In a particular embodiment, the vector further comprises one or more of: an origin of replication, one or more sequences for promoting integration of the expression cassette into the host genome, a sequence as reported herein that comprises a terminator, or an additional gene, wherein the additional gene encodes an antisense RNA, a microRNA, an shRNA, a ribozyme, structural protein, an enzyme, a transcription factor, or a transporter.

E. Chloroplast Transit Peptide

[00133] In a further aspect, provided herein is a nucleic acid molecule that comprises a nucleotide sequence encoding a chloroplast transit peptide, where the nucleic acid molecule that comprises a nucleotide sequence encoding a chloroplast transit peptide has at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:36 or a functional fragment thereof. The chloroplast transit peptide can direct a polypeptide to which it is operably linked to the stroma of a chloroplast, such as the chloroplast of a microalga. The nucleotide sequence encoding a chloroplast transit peptide as provided herein can be operably linked to a polypeptide-encoding nucleotide sequence such that when the chloroplast transit peptide-polypeptide construct is

expressed in a photosynthetic eukaryotic cell, the synthesized polypeptide is localized to the chloroplast of the cell.

[00134] The photosynthetic eukaryotic cell can be a plant or algal cell, and in some examples is a microalgal cell, such as a cell of any microalga disclosed herein. In certain examples, the microalgal cell is a nucleic acid molecule that comprises a nucleotide sequence encoding a chloroplast transit peptide having at least 80%, at least 85%, at least 90%, or at least 95% sequence identity to SEQ ID NO:36, operably linked to a nucleotide sequence encoding a polypeptide of interest. The polypeptide of interest can be any polypeptide, and can be, without limitation: a metabolic enzyme; a polypeptide that participates in light harvesting or electron transport; a transcriptional regulator; a translational regulator; a transporter; an NADH or NADPH-forming reductase; a chaperonin or protein that participates in protein or protein complex assembly or membrane assembly; a selectable marker; a reporter gene; etc. A construct that includes a nucleic acid molecule encoding a transit peptide as provided herein operably linked to a nucleotide sequence encoding a protein of interest can be introduced into a photosynthetic eukaryotic cell, such as a plant or alga cell, including an algal cell belonging to any of the taxonomic groups provided herein, and the photosynthetic eukaryotic cell can be subjected to conditions under which the construct is expressed, such that the polypeptide of interest is synthesized and localized to the chloroplast of the photosynthetic eukaryotic cell.

F. Transformation Methods

[00135] The present invention also provides transformation methods in which a eukaryotic cell is transformed with an expression vector as described herein. The methods comprise introducing an expression vector as provided herein that includes at least one promoter as provided herein and then selecting for a transformant. The expression vector may be introduced by many methods familiar to those skilled in the art including, as non-limiting examples: natural DNA uptake (Chung *et al.* (1998) *FEMS Microbiol. Lett.* 164:353–61); conjugation (Wolk *et al.* (1984) *Proc. Nat'l. Acad. Sci. USA* 81:1561–65); transduction; glass bead transformation (Kindle *et al.* (1989) *J. Cell*

Biol. 109:2589–601); silicon carbide whisker transformation (Dunahay *et al.* (1997) *Methods Mol. Biol.* (1997) 62:503–09); biolistics (Dawson *et al.* (1997) *Curr. Microbiol.* 35:356–62); electroporation (Kjaerulff *et al.* (1994) *Photosynth. Res.* 41:277–83); laser-mediated transformation; or incubation with DNA in the presence of or after pre-treatment with any of poly(amidoamine) dendrimers (Pasupathy *et al.* (2008) *Biotechnol. J.* 3:1078–82), polyethylene glycol (Ohnuma *et al.* (2008) *Plant Cell Physiol.* 49:117–20), cationic lipids (Muradawa *et al.* (2008) *J. Biosci. Bioeng.* 105:77–80), dextran, calcium phosphate, or calcium chloride (Mendez-Alvarez *et al.* (1994) *J. Bacteriol.* 176:7395–97), optionally after treatment of the cells with cell wall-degrading enzymes (Perrone *et al.* (1998) *Mol. Biol. Cell* 9:3351–65). *Agrobacterium*-mediated transformation can also be performed on algal cells, for example after removing or wounding the algal cell wall (*e.g.*, WO 2000/62601). Biolistic methods have been shown to be successful for transformation of the chloroplasts of plant and eukaryotic algal species (*see, e.g.*, WO 2007/133558y). When transforming chloroplasts, it can be useful to codon-optimize the gene of interest for expression in chloroplasts (*see, e.g.*, WO 2011/034863).

[00136] The eukaryotic cell transformed can be, for example, a fungal, heterokont, algal, or plant cell. For example, the eukaryotic cell transformed using an expression vector as provided herein can be an algal cell, such as a species of *Achnanthes*, *Amphiprora*, *Amphora*, *Ankistrodesmus*, *Asteromonas*, *Boekelovia*, *Bolidomonas*, *Borodinella*, *Botrydium*, *Botryococcus*, *Bracteococcus*, *Chaetoceros*, *Carteria*, *Chlamydomonas*, *Chlorococcum*, *Chlorogonium*, *Chlorella*, *Chroomonas*, *Chrysosphaera*, *Cricosphaera*, *Crypthecodinium*, *Cryptomonas*, *Cyclotella*, *Dunaliella*, *Ellipsoidon*, *Emiliana*, *Eremosphaera*, *Ernodesmus*, *Euglena*, *Eustigmatos*, *Franceia*, *Fragilaria*, *Gloeothamnion*, *Haematococcus*, *Halocafeteria*, *Heterosigma*, *Hymenomonas*, *Isochrysis*, *Lepocinclis*, *Micractinium*, *Monoraphidium*, *Nannochloris*, *Nannochloropsis*, *Navicula*, *Neochloris*, *Nephrochloris*, *Nephroselmis*, *Nitzschia*, *Ochromonas*, *Oedogonium*, *Oocystis*, *Ostreococcus*, *Pavlova*, *Parachlorella*, *Pascheria*, *Pelagomonas*, *Phaeodactylum*, *Phagus*, *Picochlorum*, *Platymonas*, *Pleurochrysis*, *Pleurococcus*, *Prototheca*, *Pseudochlorella*, *Pseudoneochloris*, *Pseudostaurastrum*,

Pyramimonas, *Pyrobotrys*, *Scenedesmus*, *Skeletonema*, *Spyrogyra*, *Stichococcus*, *Tetraselmis*, *Thalassiosira*, *Tribonema*, *Vaucheria*, *Viridiella*, *Vischeria*, and *Volvox*. For example, the eukaryotic cell transformed using the methods provided herein can optionally be a species of *Tetraselmis*, such as *Tetraselmis alacris*, *T. apiculata*, *T. ascus*, *T. astigmatica*, *T. chuii*, *T. convolutae*, *T. cordiformis*, *T. desikacharyi*, *T. gracilis*, *T. hazeni*, *T. impellucida*, *T. inconspicua*, *T. levis*, *T. maculate*, *T. marina*, *T. micropapillata*, *T. rubens*, *T. striata*, *T. suecica*, *T. tetrabrachia*, *T. tetrathele*, *T. verrucosa*, or *T. wettsteinii*.

[00137] In further examples, the eukaryotic cell can be an alga of the class *Chlorophyceae*, *Trebouxiophyceae*, or *Prasinophyceae*.

[00138] In some examples, a *Tetraselmis* cell is transformed by electroporation or particle bombardment. The expression vector used to transform the host cell may encode, for a selectable marker, a *ble* gene conferring bleomycin resistance, a polypeptide, or a functional RNA.

[00139] Additionally or alternatively, a *Tetraselmis* cell is transformed by a biolistic method. For example, the transformation may be achieved by use of a gene gun. By way of example, the gene gun may employ a rupture disc of 1100 rps, of 1350 rps, of 1550 rps, of 1800 rps, or of 2000 rps. A biolistic transformation of the present invention may involve any amount of DNA coated particles per bombardment; for example, a biolistic transformation method of the present invention may employ 500 µg, 1 mg, 2 mg, or 3 mg of DNA coated particles per bombardment.

[00140] The distance between a microcarrier and a plate in a biolistic transformation method of the present invention may vary according to the construct being transformed into a cell and the cells; for example the distance may be approximately 3 cm, approximately 4 cm, approximately 5 cm, approximately 6 cm, approximately 7 cm, approximately 8 cm, or even approximately 9 cm. A macrocarrier used in a biolistic method of the present invention may be in either the top position or bottom position, depending on the construct being transformed into a cell and the cells. DNA used in a biolistic method of the present invention may be linearized by restriction enzyme digest or may be circular supercoiled plasmid DNA, depending on the construct being

transformed into a cell and the cells. Cells grown for use in a biolistic method of the present invention may be grown in a variety of culture conditions; for example, cells destined for bombardment in a biolistic method of the present invention may be grown in PM024 liquid media or on PM024 agar plates for 2 days prior to bombardment. Cells used in a biolistic method of the present invention may be in different phases of a 14:10 diel cycle; for example, cells may be in either a light phase or dark phase, depending on the construct being transformed into a cell and the cells. A variety of antibiotic concentrations may be used to select transformants after transformation by a biolistic method of the present invention. One skilled in the art will recognize that these parameters are all subject to routine optimization involving only limited experimentation.

G. Culture

[00141] Eukaryotic host cells, such as any of the cells disclosed hereinabove, transformed with the expression vectors are also provided herein. Transformed algal cell cultures can be diluted, plated on agar, and allowed to grow until isolated colonies can be selected for further propagation as clonal strains.

[00142] Therefore, in one embodiment a eukaryotic cell is provided comprising an isolated or recombinant nucleic acid molecule as described herein or an expression cassette as described herein, or a vector as described herein.

[00143] Algae can be cultured phototrophically, in the absence of a fixed carbon source. Additionally or alternatively algae can be cultured mixotrophically, where the cultures are supplied with light for at least part of the day, and also supplied with a reduced carbon source, such as a sugar (*e.g.*, glucose, fructose, galactose, mannose, rhamnose, arabinose, xylose, lactose, sucrose, maltose), an organic acid (*e.g.*, acetate, citrate, succinate), or glycerol. The photosynthetic organism can be grown mixotrophically for a period of time, followed by a period of phototrophic growth, or vice versa.

[00144] Media for phototrophic or mixotrophic growth of algae are known in the art, and media can be optimized to enhance growth or production of fatty acid products for a

particular species. Artificial light sources can be used as the sole light source or to enhance or extend natural light.

[00145] Growth of algae can be in open areas, such as, for example, ponds, canals, channels, raceways, or tanks, or can be in bioreactors. Bioreactors are preferred for mixotrophic growth, and can also be used for phototrophic growth. The bioreactors can be of any sizes and form, and can include inlets for providing nutrients, additives, or gases, such as but not limited to air or CO₂. A bioreactor preferably also has an outlet for sampling of the culture. A bioreactor can be configured such that the algal culture is mixed during the growth period, for example, by stirring, rocking, shaking, inverting, bubbling of gases through the culture, *etc.* Outdoor ponds, raceways, tanks, canals, *etc.* can also be designed for mixing of cultures through, for example, paddles, pumps, hoses or jets for circulation of the culture media, or tubes, hoses or inlets for supplying air or CO₂ to the culture.

H. Further Embodiments

[00146] Embodiment 1. An isolated DNA molecule comprising a sequence selected from the group consisting of a nucleotide sequence having at least 80% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1; a nucleotide sequence having at least 80% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:2; a nucleotide sequence having at least 80% identity to at least 100, at least 200, at least 300, at least 400, or at least 500, contiguous nucleotides of SEQ ID NO:3; a nucleotide sequence having at least 80% identity to at least 100 at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:4; a nucleotide sequence having at least 80% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:5; a nucleotide sequence having at least 80% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:6; a nucleotide sequence having at least 80% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:7; and a nucleotide sequence having at least 80% identity to at

least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:8.

[00147] Embodiment 2. The isolated or recombinant nucleic acid molecule of Embodiment 1, wherein at least one of the following are satisfied:

the nucleic acid molecule has promoter activity in a eukaryotic cell, preferably an algal cell or heterokont cell;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 600, at least 700, or about 716 contiguous nucleotides of SEQ ID NO:1;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1200, at least 1400, or about 1504 contiguous nucleotides of SEQ ID NO:2;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1200, at least 1400, or about 1599 contiguous nucleotides of SEQ ID NO:3;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 600, at least 700, at least 800, or about 838 contiguous nucleotides of SEQ ID NO:4;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 600, at least 700, at least 800, at least 900, at least 1000, or about 1077 contiguous nucleotides of SEQ ID NO:5;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 600, at least 700, at least 800, at least 900, at least 1000, or about 1042 contiguous nucleotides of SEQ ID NO:6;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 600, at least 700, at least 800, at least 900, at least 950, or about 991 contiguous nucleotides of SEQ ID NO:7;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 600, at least 700, at least 800, at least 900, at least 950, or about 991 contiguous nucleotides of SEQ ID NO:8;

the isolated nucleic acid molecule comprises a constitutive promoter; and

the isolated nucleic acid molecule comprises a promoter active in a eukaryotic cell in nutrient replete and nutrient deficient culture conditions.

[00148] Embodiment 3. A promoter comprising a nucleic acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 100, at least 200, at least 300, at least 400, or at least 500 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8.

[00149] Embodiment 4. The promoter according to Embodiment 3 wherein the promoter is constitutive.

[00150] Embodiment 5. An isolated DNA molecule comprising a sequence selected from the group consisting of a nucleotide sequence having at least 80% at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 100 contiguous nucleotides of SEQ ID NO:9; and a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 100 contiguous nucleotides of SEQ ID NO:10.

[00151] Embodiment 6. An isolated or recombinant nucleic acid molecule comprising a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 100

or at least 150 contiguous nucleotides of SEQ ID NO:9 or SEQ ID NO:10, optionally wherein at least one of the following are satisfied:

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 950, or about 1011 contiguous nucleotides of SEQ ID NO:9;

the isolated nucleic acid molecule has at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or about 100% identity to at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1200, or about 1223 contiguous nucleotides of SEQ ID NO:10;

wherein the isolated or recombinant nucleic acid molecule preferably comprises a terminator sequence.

[00152] Embodiment 7. An expression cassette comprising:

a promoter according to Embodiment 3; and
a heterologous gene encoding a polypeptide or a functional RNA sequence operably linked to the promoter.

[00153] Embodiment 8. An expression cassette comprising the isolated nucleic acid molecule of any of Embodiments 1–4, wherein:

a) the isolated nucleic acid molecule is operably linked to a gene encoding a polypeptide, optionally wherein the polypeptide is:

a protein associated with lipid biosynthesis, an acetyl-CoA carboxylase, a malonyl type 1 fatty acid synthase, a Type 2 fatty acid synthase subunit, a beta ketoacyl-ACP synthase, a malonyl-CoA-malonyl-ACP acyltransferase, an acyl-ACP thioesterase, an acyl-CoA thioesterase, a 4-hydroxybenzoyl thioesterase, an alcohol forming acyl reductase, a wax synthase, an aldehyde decarbonylase, a fatty acid decarboxylase, a lipase, a glyceraldehyde 3 phosphate dehydrogenase, an acyl-CoA synthetase, a phospholipid diacylglycerol acyltransferase, a glycerol 3 phosphate acyltransferase, a lysophosphatidic acid acyltransferase, a phosphatidic acid phosphatase, a diacyl glycerol

acyltransferase, a polypeptide having lipolytic activity, a polypeptide that participates in photosynthesis, a chlorophyll binding light harvesting polypeptide, a photosynthetic reaction center polypeptide, an oxygen-evolving complex polypeptide, a cytochrome, a ferredoxin, a protein associated with carbon fixation, a ribulose biphosphate carboxylase subunit, a carbonic anhydrase, a transporter protein, an ABC transporter, a FatB transporter, a dehydrogenase, an aldehyde dehydrogenase, a 2-hydroxyacid dehydrogenase, an isocitrate dehydrogenase, 6 phosphogluconate dehydrogenase, glucose 6 phosphate dehydrogenase; a transcription factor, a protein involved in cell signaling, a kinase, or a G protein; or

(b) the isolated nucleic acid molecule is operably linked to a gene encoding a functional RNA, optionally wherein the functional RNA is an antisense RNA, a small hairpin RNA, a microRNA, an antisense RNA, a siRNA, a piRNA, or a ribozyme;

wherein the expression cassette optionally further comprises a nucleotide sequence according to Embodiment 6; wherein the nucleotide sequence of Embodiment 6 comprises a terminator, further wherein the terminator is operably linked to the heterologous gene coding sequence and/or the terminator is downstream of the heterologous gene coding sequence.

[00154] Embodiment 9. A vector comprising an expression cassette according to Embodiment 8, wherein the expression vector further comprises one or more of: an origin of replication; and one or more sequences for promoting integration of the expression cassette into the host genome, or a selectable marker or reporter gene; optionally wherein the selectable marker gene is selected from the group consisting of a gene conferring resistance to an antibiotic, a gene conferring resistance to an herbicide, a gene encoding acetyl CoA carboxylase (ACCase), a gene encoding acetohydroxy acid synthase (*ahas*), a gene encoding acetolactate synthase, a gene encoding aminoglycoside phosphotransferase, a gene encoding anthranilate synthase, a gene encoding bromoxynil nitrilase, a gene encoding cytochrome P450-NADH-cytochrome P450 oxidoreductase, a gene encoding dalapon dehalogenase, a gene encoding dihydropteroate synthase, a gene encoding a class I 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a gene encoding a class II EPSPS (*aroA*), a gene encoding a non-class I/II EPSPS, a gene

encoding glutathione reductase, a gene encoding glyphosate acetyltransferase, a gene encoding glyphosate oxidoreductase, a gene encoding hydroxyphenylpyruvate dehydrogenase, a gene encoding hydroxy-phenylpyruvate dioxygenase, a gene encoding isoprenyl pyrophosphate isomerase, a gene encoding lycopene cyclase, a gene encoding phosphinothricin acetyl transferase, a gene encoding phytoene desaturase, a gene encoding prenyl transferase, a gene encoding protoporphyrin oxidase, a gene encoding superoxide dismutase, *arg7*, *his3*, *hisD*, *hisG*, *manA*, *nit1*, *trpB*, *uidA*, *xylA*, a dihydrofolate reductase gene, a mannose-6-phosphate isomerase gene, a nitrate reductase gene, an ornithine decarboxylase gene, a thymidine kinase gene, a 2-deoxyglucose resistance gene; an R-locus gene, a tyrosinase gene, *lacZ*, an alkaline phosphatase gene, an α -amylase gene, a horseradish peroxidase gene, an α -galactosidase gene, a luciferin/luciferase gene, a beta-glucuronidase gene (*GUS*), and a gene encoding a fluorescent protein.

[00155] Embodiment 10. A method for transforming a eukaryotic cell comprising:
 introducing a vector according to Embodiment 9 into the eukaryotic cell; and
 selecting for a transformed eukaryotic cell, preferably wherein the eukaryotic host cell is a fungal, algal, heterokont, or plant host cell optionally wherein the eukaryotic host cell is a species of *Bacillariophyceae* (diatoms), *Bolidomonas*, *Chlorophyceae* (green algae), *Chrysophyceae* (golden algae), *Cyanophyceae* (cyanobacteria), *Eustigmatophyceae* (pico-plankton), *Glaucocystophytes*, *Pelagophyceae*, *Bolidophyceae*, *Prasinophyceae* (pico-plankton), *Raphidophyceae*, *Rhodophyceae* (red algae), *Synurophyceae* and *Xanthophyceae* (yellow-green algae), *Achnanthes*, *Amphiprora*, *Amphora*, *Ankistrodesmus*, *Asteromonas*, *Boekelovia*, *Bolidomonas*, *Borodinella*, *Botrydium*, *Botryococcus*, *Bracteococcus*, *Chaetoceros*, *Carteria*, *Chlamydomonas*, *Chlorococcum*, *Chlorogonium*, *Chlorella*, *Chroomonas*, *Chrysosphaera*, *Cricosphaera*, *Crypthecodinium*, *Cryptomonas*, *Cyclotella*, *Dunaliella*, *Ellipsoidon*, *Emiliania*, *Eremosphaera*, *Ernodesmus*, *Euglena*, *Eustigmatos*, *Franceia*, *Fragilaria*, *Gloeothamnion*, *Haematococcus*, *Halocafeteria*, *Heterosigma*, *Hymenomonas*, *Isochrysis*, *Lepocinclis*, *Micractinium*, *Monoraphidium*, *Nannochloris*, *Nannochloropsis*, *Navicula*, *Neochloris*, *Nephrochloris*, *Nephroselmis*, *Nitzschia*, *Ochromonas*, *Oedogonium*, *Oocystis*, *Ostreococcus*, *Pavlova*, *Parachlorella*, *Pascheria*, *Pelagomonas*,

Phaeodactylum, *Phagus*, *Picochlorum*, *Platymonas*, *Pleurochrysis*, *Pleurococcus*, *Prototheca*, *Pseudochlorella*, *Pseudoneochloris*, *Pseudostaurastrum*, *Pyramimonas*, *Pyrobotrys*, *Scenedesmus*, *Skeletonema*, *Spyrogyra*, *Stichococcus*, *Tetraselmis*, *Thalassiosira*, *Tribonema*, *Vaucheria*, *Viridiella*, *Vischeria*, or *Volvox*.

[00156] Embodiment 11. The method according to Embodiment 10, wherein the vector is introduced by a biolistic procedure, optionally wherein at least about 300 psi of pressure is used to impel biolistic microcarriers coated with vector DNA into the eukaryotic cell.

[00157] Embodiment 12. A method for co-transforming a eukaryotic cell comprising:
introducing an expression cassette according to Embodiment 8 and a nucleic acid sequence encoding a selectable marker into the eukaryotic cell; and

selecting for the presence of the selectable marker in a transformed eukaryotic cell to provide a eukaryotic cell transformed with the expression cassette, optionally wherein the eukaryotic cell is an algal cell, optionally wherein the selectable marker gene is operably linked to a promoter according to Embodiment 3, optionally wherein the selectable marker gene is selected from the group consisting of a gene conferring resistance to an antibiotic, a gene conferring resistance to an herbicide, a gene encoding acetyl CoA carboxylase (ACCase), a gene encoding acetohydroxy acid synthase (*ahas*), a gene encoding acetolactate synthase, a gene encoding aminoglycoside phosphotransferase, a gene encoding anthranilate synthase, a gene encoding bromoxynil nitrilase, a gene encoding cytochrome P450-NADH-cytochrome P450 oxidoreductase, a gene encoding dalapon dehalogenase, a gene encoding dihydropteroate synthase, a gene encoding a class I 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a gene encoding a class II EPSPS (*aroA*), a gene encoding a non-class I/II EPSPS, a gene encoding glutathione reductase, a gene encoding glyphosate acetyltransferase, a gene encoding glyphosate oxidoreductase, a gene encoding hydroxyphenylpyruvate dehydrogenase, a gene encoding hydroxy-phenylpyruvate dioxygenase, a gene encoding isoprenyl pyrophosphate isomerase, a gene encoding lycopene cyclase, a gene encoding phosphinothricin acetyl transferase, a gene encoding phytoene desaturase, a gene encoding prenyl transferase, a gene encoding protoporphyrin oxidase, a gene encoding

superoxide dismutase, *arg7*, *his3*, *hisD*, *hisG*, *manA*, *nit1*, *trpB*, *uidA*, *xylA*, a dihydrofolate reductase gene, a mannose-6-phosphate isomerase gene, a nitrate reductase gene, an ornithine decarboxylase gene, a thymidine kinase gene, a 2-deoxyglucose resistance gene; and an R-locus gene.

[00158] Embodiment 13. A method according to Embodiment 12, wherein the algal cell is from a species selected from the group consisting of *Achnanthes*, *Amphiprora*, *Amphora*, *Ankistrodesmus*, *Asteromonas*, *Boekelovia*, *Bolidomonas*, *Borodinella*, *Botrydium*, *Botryococcus*, *Bracteococcus*, *Chaetoceros*, *Carteria*, *Chlamydomonas*, *Chlorococcum*, *Chlorogonium*, *Chlorella*, *Chroomonas*, *Chrysosphaera*, *Cricosphaera*, *Crypthecodinium*, *Cryptomonas*, *Cyclotella*, *Dunaliella*, *Ellipsoidon*, *Emiliana*, *Eremosphaera*, *Ernodesmus*, *Euglena*, *Eustigmatos*, *Franceia*, *Fragilaria*, *Gloeothamnion*, *Haematococcus*, *Halocafeteria*, *Heterosigma*, *Hymenomonas*, *Isochrysis*, *Lepocinclis*, *Micractinium*, *Monoraphidium*, *Nannochloris*, *Nannochloropsis*, *Navicula*, *Neochloris*, *Nephrochloris*, *Nephroselmis*, *Nitzschia*, *Ochromonas*, *Oedogonium*, *Oocystis*, *Ostreococcus*, *Pavlova*, *Parachlorella*, *Pascheria*, *Pelagomonas*, *Phaeodactylum*, *Phagus*, *Picochlorum*, *Platymonas*, *Pleurochrysis*, *Pleurococcus*, *Prototheca*, *Pseudochlorella*, *Pseudoneochloris*, *Pseudostaurastrum*, *Pyramimonas*, *Pyrobotrys*, *Scenedesmus*, *Schizochlamydeella*, *Skeletonema*, *Spyrogyra*, *Stichococcus*, *Tetrachlorella*, *Tetraselmis*, *Thalassiosira*, *Tribonema*, *Vaucheria*, *Viridiella*, *Vischeria*, and *Volvox*.

[00159] Embodiment 14. A eukaryotic host cell comprising an isolated or recombinant nucleic acid molecule, expression cassette, or vector of any of Embodiments 7–9, wherein the eukaryotic host cell is optionally a microalgal cell, optionally a species of wherein the eukaryotic cell is an algal cell, optionally, a species of *Bacillariophyceae*, *Bolidomonas*, *Chlorophyceae*, *Chrysophyceae*, *Eustigmatophyceae*, *Glaucocystophytes*, *Pelagophyceae*, *Bolidophyceae*, *Prasinophyceae*, *Raphidophyceae*, *Rhodophyceae* (red algae), *Synurophyceae* or *Xanthophyceae* (yellow-green algae), or of a genus selected from the group consisting of: *Achnanthes*, *Amphiprora*, *Amphora*, *Ankistrodesmus*, *Asteromonas*, *Boekelovia*, *Bolidomonas*, *Borodinella*, *Botrydium*, *Botryococcus*, *Bracteococcus*, *Chaetoceros*, *Carteria*, *Chlamydomonas*, *Chlorococcum*, *Chlorogonium*,

Chlorella, Chroomonas, Chrysosphaera, Cricosphaera, Crypthecodinium, Cryptomonas, Cyclotella, Dunaliella, Ellipsoidon, Emiliana, Eremosphaera, Ernodesmius, Euglena, Eustigmatos, Franceia, Fragilaria, Gloeothamnion, Haematococcus, Halocafeteria, Heterosigma, Hymenomonas, Isochrysis, Lepocinclis, Micractinium, Monoraphidium, Nannochloris, Nannochloropsis, Navicula, Neochloris, Nephrochloris, Nephroselmis, Nitzschia, Ochromonas, Oedogonium, Oocystis, Ostreococcus, Pavlova, Parachlorella, Pascheria, Pelagomonas, Phaeodactylum, Phagus, Picochlorum, Platymonas, Pleurochrysis, Pleurococcus, Prototheca, Pseudochlorella, Pseudoneochloris, Pseudostaurastrum, Pyramimonas, Pyrobotrys, Scenedesmus, Schizochlamydeella, Skeletonema, Spyrogyra, Stichococcus, Tetrachlorella, Tetraselmis, Thalassiosira, Tribonema, Vaucheria, Viridiella, Vischeria, and Volvox.

[00160] Embodiment 15. A nucleic acid molecule that comprises a nucleotide sequence encoding a chloroplast transit peptide, where the nucleic acid molecule that comprises a nucleotide sequence encoding a chloroplast transit peptide has at least 80%, at least 85%, at least 90%, or at least 95% identity to SEQ ID NO:36 or a functional fragment thereof.

[00161] Embodiment 16. The nucleic acid molecule of Embodiment 15, wherein the sequence encoding a chloroplast transit peptide is operably linked to a nucleic acid sequence encoding a polypeptide that is localized to the chloroplast of the cell when the nucleic acid molecule is expressed in a eukaryotic photosynthetic cell.

[00162] Embodiment 17. A eukaryotic photosynthetic cell comprising the nucleic acid molecule of Embodiment 15 or 16, wherein the eukaryotic photosynthetic cell is preferably an algal cell, for example an algal cell of the class *Chlorophyceae*, *Prasinophyceae*, or *Trebouxiophyceae*.

EXAMPLES

[00163] The following examples are merely illustrative, and do not limit this disclosure in any way.

Example 1. Strains and Media

[00164] *Tetraselmis* is a green alga in the *Prasinophyceae* class (*Chlorophyte* phylum). A strain of *Tetraselmis*, designated WT-105, was isolated from an environmental sample from South Laguna Madre, Texas. Strain P-925 is a classically improved derivative of WT-105 with increased lipid productivity traits.

[00165] Media used for the growth of *Tetraselmis* included the following:

[00166] PM024: 35 g/L Instant Ocean Salts (Aquatic Eco Systems, Apopka, FL), 10x Guillard's F/2 marine water enrichment solution (from 50x stock from Sigma-Aldrich, St. Louis, MO, cat. No. G0154; final concentrations of components in media: 8.825 mM Sodium nitrate; 320 μ M Sodium phosphate monobasic; 205 nM biotin; 420 nM Cobalt chloride•6H₂O; 400 nM Cupric sulfate•5H₂O; 116.5 μ M Ferric chloride•6H₂O; 117.13 μ M Disodium EDTA•2H₂O; 9.095 μ M Manganese chloride•4H₂O; 248 nM Sodium molybdate•2H₂O; 2.965 μ M Thiamine•HCl; 37 nM Vitamin B₁₂; 765 nM Zinc sulfate•7H₂O).

[00167] PM032 is filtered seawater to which 1.3 mL per liter (of final media volume) PROLINE® F/2 Algae Feed Part A (Aquatic Eco-Systems) and 1.3 ml/L PROLINE® F/2 Algae Feed Part B (Aquatic Eco-Systems) have been added. The final media composition includes 8.8 mM NaNO₃, 0.361 mM NaH₂PO₄•H₂O, 10x the Guillard's F/2 medium trace metal concentration, and 10X the Guillard's F/2 medium trace vitamins (Guillard (1975) Culture of phytoplankton for feeding marine invertebrates. in "Culture of Marine Invertebrate Animals." (eds: Smith W.L. and Chanley M.H.) Plenum Press, New York, USA. pp 26–60).

[00168] Nutrient Replete Media: 35 g/L Instant Ocean Salts, 0.5 g/L NaHCO₃, 10 mM MOPS pH 8, 7 mM NH₄Cl, 140 μ M KH₂PO₄; final concentration of trace elements and vitamins: (400 μ M MgSO₄•7H₂O; 48 μ M CaCl₂•2H₂O; 162 μ M H₃BO₃; 2 μ M NaVO₃; 47 μ M FeCl₃•6H₂O; 47 μ M Na₂EDTA•2H₂O; 3.6 μ M MnCl₂•4H₂O; 304 nM ZnSO₄•7H₂O; 168 nM CoCl₂•6H₂O; 156 nM CuSO₄•5H₂O; 104 nM Na₂MoO₄•2H₂O; 1.2 μ M Thiamine•HCl; 8 nM Biotin; 1.5 nM Cyanocobalamine).

[00169] Nitrogen Deficient (a.k.a. N minus) Media: 35 g/L Instant Ocean Salts, 0.5 g/L NaHCO₃, 10 mM MOPS pH 8, 140 μ M KH₂PO₄; final concentration of trace elements and vitamins: (400 μ M MgSO₄•7H₂O; 48 μ M CaCl₂•2H₂O; 162 μ M H₃BO₃; 2 μ M NaVO₃; 47 μ M FeCl₃•6H₂O; 47 μ M Na₂EDTA•2H₂O; 3.6 μ M MnCl₂•4H₂O; 304 nM ZnSO₄•7H₂O; 168 nM CoCl₂•6H₂O; 156 nM CuSO₄•5H₂O; 104 nM Na₂MoO₄•2H₂O; 1.2 μ M Thiamine•HCl; 8 nM Biotin; 1.5 nM Cyanocobalamine).

[00170] Phosphate Deficient Media 1: 35 g/L Instant Ocean Salts, 500 mg/L NaHCO₃, 10 mM MOPS pH 8, 7 mM NH₄Cl, 35 μ M KH₂PO₄; final concentration of trace elements and vitamins: (400 μ M MgSO₄•7H₂O; 48 μ M CaCl₂•2H₂O; 162 μ M H₃BO₃; 2 μ M NaVO₃; 47 μ M FeCl₃•6H₂O; 47 μ M Na₂EDTA•2H₂O; 3.6 μ M MnCl₂•4H₂O; 304 nM ZnSO₄•7H₂O; 168 nM CoCl₂•6H₂O; 156 nM CuSO₄•5H₂O; 104 nM Na₂MoO₄•2H₂O; 1.2 μ M Thiamine•HCl; 8 nM Biotin; 1.5 nM Cyanocobalamine).

[00171] Phosphate Deficient Media 2: 35 g/L Instant Ocean Salts, 500 mg/L NaHCO₃, 10 mM MOPS pH 8, 7 mM NH₄Cl, 18 μ M KH₂PO₄; final concentration of trace elements and vitamins: (400 μ M MgSO₄•7H₂O; 48 μ M CaCl₂•2H₂O; 162 μ M H₃BO₃; 2 μ M NaVO₃; 47 μ M FeCl₃•6H₂O; 47 μ M Na₂EDTA•2H₂O; 3.6 μ M MnCl₂•4H₂O; 304 nM ZnSO₄•7H₂O; 168 nM CoCl₂•6H₂O; 156 nM CuSO₄•5H₂O; 104 nM Na₂MoO₄•2H₂O; 1.2 μ M Thiamine•HCl; 8 nM Biotin; 1.5 nM Cyanocobalamine).

[00172] Phosphate Deficient Media 3: 35 g/L Instant Ocean Salts, 500 mg/L NaHCO₃, 10 mM MOPS pH 8, 7 mM NH₄Cl, 10 μ M KH₂PO₄; final concentration of trace elements and vitamins: (400 μ M MgSO₄•7H₂O; 48 μ M CaCl₂•2H₂O; 162 μ M H₃BO₃; 2 μ M NaVO₃; 47 μ M FeCl₃•6H₂O; 47 μ M Na₂EDTA•2H₂O; 3.6 μ M MnCl₂•4H₂O; 304 nM ZnSO₄•7H₂O; 168 nM CoCl₂•6H₂O; 156 nM CuSO₄•5H₂O; 104 nM Na₂MoO₄•2H₂O; 1.2 μ M Thiamine•HCl; 8 nM Biotin; 1.5 nM Cyanocobalamine).

[00173] Stock cultures were cultured in shake flasks with 50 mL PM024 containing 250 μ g/mL ZEOCIN™ phleomycin D1 antibiotic (“Zeocin”) when appropriate. Shake flasks were cultivated at 25°C, 100 rpm, 60 μ E constant light, 1% CO₂.

Example 2. *Tetraselmis* transformation vector design and synthesis

[00174] Vector p05320 (**Figure 1**), containing the *Streptoalloteichus hindustanus* (*Sh*) *ble* gene conferring Zeocin-resistance and codon-optimized for expression in *Chlamydomonas reinhardtii* (SEQ ID NO:30) under the control of the WT-105 actin promoter fragment (SEQ ID NO:1) and WT-105 actin terminator fragment (SEQ ID NO:9), was assembled using the Gibson assembly method (Gibson *et al.* (2008) *Science* 319:1215–20).

[00175] The WT-105 actin gene (**Figure 2**) was identified based on genome sequencing and homology to actin genes of other organisms. The sequence from 1–713 bp upstream of the actin gene start codon (SEQ ID NO:1) was amplified as the actin gene promoter fragment and the sequence from 1–1000 bp downstream of the actin gene termination codon (SEQ ID NO: 9) was amplified as the actin gene terminator fragment to direct expression of the *ble* gene of plasmid p05320 (**Figure 1**). After PCR amplification using the primers in **Table 2** and WT-105 genomic DNA as the template, a 713 bp upstream region (actin promoter fragment; SEQ ID NO:1) and 1 kbp downstream region (actin terminator fragment; SEQ ID NO:9) were assembled with the PCR-amplified *Sh ble* gene (SEQ ID NO:30) and cloned into a plasmid containing the pUC19 backbone between HindIII and EcoRI restriction sites, forming the plasmid p05320 (**Figure 1**; SEQ ID NO:29). The *Sh ble* coding sequence, codon-optimized for *Chlamydomonas*, (SEQ ID NO:30), was PCR amplified with primers BLE-ActinWT105-5 (SEQ ID NO:15) and BLE-ActinWT105-3 (SEQ ID NO:16) (**Table 2**).

Table 2. Primers for p05320 construction

Primer name	Primer sequence	SEQ ID NO:
WT105ActinPro-EcoRI	5'-aaacgacggccagtgaattccgtttctgacacctgggcctt	11
WT105ActinPro-BLE-3	5'-gacgtcagcttgcccaacatggtgcctttctcgttaagag	12
WT105ActinTER-BLE-5	5'-tcgccgaagagcaggattaagcaagcgctgtgcaatcaa	13
WT105ActinTER-HindIII	5'-accatgattacccaagcttcattggtggtggagccct	14
BLE-ActinWT105-5	5'-ctcttacgaagaaggcaccatgttgccaagctgacgtc	15
BLE-ActinWT105-3	5'-ttgattgcacagcgcttgcttaactctgctcttcggcga	16

[00176] The p05320 construct also contained a gene coding for Turbo-GFP (Evrogen, Moscow, Russia), codon-optimized for expression in WT-105 based on a codon table

derived from an analysis of over 100000 WT-105 genes. This Turbo-GFP coding sequence was flanked by two *Chlamydomonas reinhardtii* regulatory sequences: the promoter fragment from the *C. reinhardtii* actin gene (GenBank accession #D50838, SEQ ID NO:31) and the terminator fragment from the *C. reinhardtii* *psaD* gene (GenBank accession #XM_001697670, SEQ ID NO:32). After PCR amplification using the primers below (Table 3), the *C. reinhardtii* actin gene 713 bp upstream region (promoter fragment) and *C. reinhardtii* *psaD* gene 1 kbp downstream region (terminator fragment) were assembled with the PCR-amplified GFP by the Gibson assembly method and cloned into a PciI restriction site. The *Pontellina plumata* GFP coding sequence was PCR amplified with primers GFP-ActinWT105-5 (SEQ ID NO:21) and GFP-ActinWT105-3 (SEQ ID NO:22) (Table 3).

Table 3. Primers for p05322 construction

Primer name	Primer sequence	SEQ ID NO:
WT105ActinPro-PciI	5'-gctggccttttctcacatgtcgtttctgacctgggcct	17
WT105ActinPro-GFP-3	5'-cccggatcgtccgactccatggtgcctttctcgtaaga	18
WT105ActinTER-GFP-5	5'-ccgacgctggagaagagtaaagcaagcgtgtgcaatcaa	19
WT105ActinTER-PciI	5'-taacgcaggaaagaacatgtccaaatggctcgtgtttggtg	20
GFP-ActinWT105-5	5'-tcttacgaagaaaggcaccatggagtcggacgaatccggg	21
GFP-ActinWT105-3	5'-ttgattgcacagcgttgccttactcttccagcgtcgg	22

Example 3. Transformation of WT-105

[00177] To introduce the p05320 plasmid into cells, WT-105 cells were harvested in late log phase and bombarded with ScaI-linearized p05320 coated on gold beads. Gold microcarriers used for bombardment of algal cells using the PDS-1000/He Particle Delivery System were S550d DNADEL™ Gold Carrier Particles (Seashell Technology). The gold particles were sonicated briefly to dissociate any aggregates prior to formulation and diluted into binding buffer to yield a final concentration of 30 mg/mL. The linearized DNA (overnight restriction enzyme digested and phenol/chloroform extracted) was added at a ratio of 2–5 µg DNA per 1 mg gold and vortexed briefly. An equal volume of precipitation buffer was added, vortexed briefly, and incubated for 3 minutes at room temperature. The DNA coated particles were centrifuged at 10 krpm in Eppendorf microfuge tubes for 10 seconds. Pelleted gold particles were washed with 500 µL of cold

100% ethanol once and resuspended in cold 100% ethanol to reach 1 mg particles per 10 μ L. Prior to particle bombardment, the DNA-coated particles were sonicated for 1 minute to disperse any aggregates.

[00178] Equipment set up and detailed operation procedures were according to the BIOLISTIC® PDS-1000/He Particle Delivery System User manual (*available at* http://biorad.com/webroot/web/pdf/lsr/literature/Bulletin_9075.pdf). Late-log phase cells were collected by centrifugation at $2000 \times g$ for 10 minutes at 25°C and 1×10^7 cells were plated on PM024 agar covering the center 5 cm of the plate. Bombardment was performed with a BioRad PDS-1000 unit at 1800 psi with a chamber vacuum of 10 Hg, with the disk positioned at 6 cm distance from the cells.

[00179] Cells were spread on PM024 agar plates containing 250 μ g/mL Zeocin and grown under constant light for three weeks at room temperature, after which ten colonies were patched on a second PM024 (containing 250 μ g/mL Zeocin) agar plate. No colonies were detected on the control plates (bombardment without plasmid). Five of the ten isolates patched from the initial selection plate grew on the second selection plate (isolates C2, C4, C6, C8, and C10) and were transferred again to new selection plates. All five grew, and were screened by colony PCR for the presence of the *ble* gene and the endogenous Na^+/H^+ ATPase gene. The five isolates that did not grow, *i.e.* that were only transiently resistant to the antibiotic, may have expressed the Ble protein from an un-integrated DNA which was subsequently lost.

[00180] *Colony PCR for p05320 Tetraselmis transformants:* Cells were picked off of patch plates and placed into 100 μ L of TE buffer. Samples were incubated at 98°C for 10 minutes to disrupt the cell wall and membrane, and then cell debris was removed by centrifugation. A PCR reaction was set up using the supernatant as the DNA template to amplify the *Sh ble* gene. The p05320 vector was used as a positive control. 10 μ M of each of the primers LF-SGE5320F1 (SEQ ID NO:27) and BLEcodon3 (SEQ ID NO:28) were used to amplify a 645 bp fragment.

[00181] To verify the sensitivity of the PCR assay, an endogenous single copy gene was used as a positive control. To determine an appropriate single copy gene, protein sequences deduced from five publically available algal genomes (*Chlamydomonas*

reinhardtii, *Ostreococcus lucimarinus*, *Ostreococcus taurii*, *Phaeodactylum tricornutum*, and *Thalassiosira pseudonana*) were clustered into families by homology. Genes were identified which had single protein per cluster and the Na⁺/H⁺ translocating ATPase gene was chosen from this group for use as an endogenous single-copy control gene. Primers NaHtransB5 (SEQ ID NO:25) and NaHtransC3 (SEQ ID NO:26) were designed to amplify a 670 bp fragment from the endogenous Na⁺/H⁺ ATPase gene. For the PCR reactions, 2x PHUSION® master PCR mix reagent mixture (New England Biolabs) was used with 200 nM of each of the primers and about 20 ng of template in a 25 µL reaction. The following PCR program was used: Step 1, 98°C for 3 min.; Step 2, 98°C for 30 sec.; Step 3, 58°C for 30 sec.; Step 4, 72°C for 1 min.; Cycle to Step 2, 35 more times; Step 5, 72°C for 5 min.; and Step 6, hold at 4°C. 15 µL of each reaction product was loaded on a 1% agarose gel and separated at 100 V for 30 minutes.

[00182] Colony PCR of the putative transformants showed, as expected, that the endogenous Na⁺/H⁺ ATPase gene was detected in all five Zeocin-resistant isolates and in the wild type cells (**Figure 3**, upper panel), indicating that sufficient template was present. No bands were detected in the negative controls (water or TE buffer alone). The *ble* gene was detected only in isolate C8, which produced a band that co-migrated with the band amplified from the p05320 vector (**Figure 3**, lower panel). The other four Zeocin-resistant isolates and wild type did not produce a band of the expected size. The C8 isolate was chosen for further analysis to determine whether the *ble* gene was integrated into the genome, and if the Ble protein was expressed.

[00183] To determine whether the *ble* gene in isolate C8 was integrated into the WT-105 genome, a Southern blot was performed using a digoxigenin-labeled PCR product amplified from the *ble* gene and restriction digested genomic DNA. Genomic DNA (5 µg) was restriction digested with either HindIII and BglII or KpnI and MfeI in a 50 µL reaction at 37°C overnight. A positive control (Ble coding sequences) was prepared by PCR using the p05320 plasmid as a template. A 378 bp digoxigenin-labeled probe was synthesized as according to the user's manual (PCR DIG Probe Synthesis Kit; Cat 11 636 090 910; Roche), by PCR with the primers of **Table 4** using the p05320 plasmid as a template.

Table 4. Primers for *ble* probe construction

Primer name	Primer sequence	SEQ ID NO:
BLE-5	5'-atgttgccaagctgacgtc	23
BLE-3	5'-ttaatcctgctcttcgga	24

[00184] Electrophoresis was performed using a 0.7% agarose gel with 50 ng/mL ethidium bromide in 1x Tris-acetate-EDTA (TAE) at 80 volts for 5 hours. Restriction digested genomic DNA (50 μ L) was loaded into each lane along with a size standard, DIG-labeled DNA Molecular Weight Marker II (Roche cat. no. 11 218 590 910). Once the gel electrophoresis was completed, the DNA was depurinated prior to transfer by submerging the gel in 250 mM HCl, with shaking at room temperature, until the bromophenol blue marker changed from blue to yellow (~10 minutes). The DNA in the gel was denatured by submerging the gel twice in denaturing solution (500 mM NaOH, 1.5 M NaCl) for 15 minutes each time, at room temperature, with gentle shaking. After rinsing in distilled water, the gel was submerged twice in neutralization buffer (500 mM Tris-HCl, pH 7.5; 1.5 M NaCl) for 15 minutes each time at room temperature. The gel was equilibrated for at least 10 minutes in 20x SSC. The DNA transfer was conducted by capillary transfer overnight onto a nylon membrane using 20x SSC. The damp nylon membrane was placed DNA side up on Whatman paper that had been soaked in 2x SSC and the membrane was placed into a crosslinker and exposed to 1.2 mJ for 5 minutes. The blot was prehybridized in pre-warmed (42°C) Easy Hyb buffer (cat# 11 603 558 001) and incubated at 42°C for 30 minutes. 80 μ L of probe was denatured at 100°C for 5 minutes and cooled quickly on ice, then added to 10 mL of pre-warmed hybridization buffer. The hybridization was conducted at 45°C overnight. The blot was washed twice in low stringency buffer (2x SSC, 0.1% SDS, 5 minutes each wash) at room temperature and washed twice in high stringency buffer (0.1x SSC, 0.1% SDS, 15 minutes each wash) at 68°C. The hybridization of the probe was detected by chemiluminescence using the DIG Wash and Block Buffer Set (Roche cat# 11 585 762 001), and an anti-digoxigenin antibody with CDP-Star chemiluminescent substrate (Roche cat# 11 685 627 001)

according to the user's manual. The chemiluminescence signal was detected by using the MULTIIMAGE® II detection and imaging system (Alpha Innotech).

[00185] As expected, no signal was detected in genomic DNA isolated from the wild type parent, whereas the anti-*ble* probe did detect bands in both digest of the C8 genomic DNA. **Figure 4** provides an image of the Southern blot of the HindIII and BglII digested genomic DNAs and **Figure 5** provides an image of the Southern blot of KpnI and MfeI digested genomic DNAs. Using DNA prepared from isolate C8, two strong bands were detected by the anti-*ble* probe, indicting the presence of the *ble* gene in the genomic DNA. Both of these bands were larger than the 1.5 kbp fragment expected from restriction digestion of p05320, although a faint band appears at this size. The larger size fragments may result from partial digestion of the genomic DNA, or from truncation of the integrated plasmid resulting in the loss of a restriction site. The presence of multiple bands indicates that the plasmid may be integrated in more than one chromosomal location.

[00186] To determine whether the *ble* gene was expressed and translated into protein, isolate C8 was also screened for the presence of Ble (bleomycin resistance protein). Shake flasks of 50 mL PM024 (250 µg/mL Zeocin) were inoculated with wild-type WT-105 or the p05320-8 WT-105 transformation isolate. Cultures were cultivated at 25°C, 100 rpm, 60 µE constant light, 1% CO₂. On day 7 post-inoculation, cell pellets were collected from 45 mL of each culture by centrifugation at 4000 × *g* for 10 minutes and re-suspended in 1 mL PBS. The resuspended cultures were transferred into 2 mL Eppendorf tubes, each containing ~500 µL of 200 µm Zirconium beads. Samples were chilled on ice for 10 minutes before lysis by bead beating for 5 minutes. The supernatants of the lysed cultures were collected after centrifugation at 4000 × *g* for 10 minutes, and pigments were solubilized by adding 4 volumes of methanol, 1 volume of chloroform, and 3 volumes of H₂O. After vortexing briefly, protein precipitates were pelleted by centrifugation at 15000 × *g* for 2 minutes. The aqueous top layers were removed and pellets were washed twice with 4 volumes of methanol. The protein precipitates were re-dissolved in 200 µL 4x LOS sample loading buffer (Invitrogen, cat. # NP0007) containing 10% β-mercaptoethanol. 40 µL of each protein extract was loaded onto a 18%

tris-glycine NOVEX® gel (Invitrogen, Carlsbad). Protein transfer was performed with iBlot (program #7, 7 minutes) (Invitrogen, Carlsbad). The membrane was incubated in blocking buffer (5% nonfat milk, TRIS-buffered saline with 1% Tween-20 (TBST)) for 1 hour before incubation in 25 mL of primary antibody (anti-BLE, InvivoGen, #ANTI-0010, 1:500) for 1 hour. The blot was washed five times with TBST for 5 minutes each wash and then incubated in secondary antibody (goat anti-rabbit IgG (H+L) HRP-conjugated, Bio-Rad cat. # 170-6515) for 45 minutes. The blot was then washed 5 times with TBST for 5 minutes each wash before chemiluminescent signal was detected with chemiluminescent peroxidase substrate (CPS-1, Sigma-Aldrich, Saint Louis) using the MULTIIMAGE® II detection and imaging system (Alpha Innotech). The size standard was MAGICMARK™ XP Western Protein Standard (Invitrogen cat. # LC5602). The *ble* gene was cloned into an *E. coli* expression vector (p05406) and crude lysates were used as a positive control.

[00187] A 13 kDa band (the size of Ble) was detected using the anti-Ble antibody in isolate C8 (**Figure 6**). The slight difference in migration between the control proteins and the C8 preparations may reflect differences in levels of total protein loaded. No band was detected in the wild type preparations using the same amount of protein.

[00188] The assays indicate that the *Tetraselmis* nuclear genome was transformed by the *ble* selectable marker, which conferred Zeocin resistance on the transformant through expression of the Ble (bleomycin resistance) protein.

Example 4. Identification of *Tetraselmis* regulatory sequences

[00189] *Tetraselmis* genes that were expressed under nutrient replete and nutrient limited culture conditions were identified and used for the identification of regulatory sequences. To identify constitutively expressed genes, WT-105 cells were grown in nutrient-replete medium, and then shifted into either nitrogen-deficient or phosphate-deficient medium (*see, Example 1, supra*), while control cultures were maintained in nutrient-replete medium. For nitrogen-deprivation, cells were grown in Nutrient Replete medium, and then cultures were centrifuged and resuspended in either Nutrient Replete

medium (+N), or Nitrogen Deficient medium (-N). The process for phosphate-deprivation was essentially the same, except that after centrifugation cells were shifted to one of three varieties of phosphate deficient media, containing 35, 18, or 10 μM KH_2PO_4 (*see, Example 1, supra*). A concentration of 140 μM of KH_2PO_4 is considered phosphate replete. Total RNA was collected before medium shift and two days post medium shift and transcripts were sequenced.

[00190] Thirty-three cDNAs were identified as exhibiting high (>1500 RPKM) transcript abundance under nutrient replete conditions as well as under both nitrogen-deprivation and phosphate-deprivation. These were analyzed for whether or not the cDNA assembly represented the entire ORF, as judged by BLAST (basic local alignment search tool; *available at* <http://blast.ncbi.nlm.nih.gov/>) results against orthologous proteins. Those cDNAs that were determined to constitute complete open reading frames (ORFs) were homology searched against a proprietary *Tetraselmis* genomic DNA (gDNA) database. BLAST hits having high confidence and representing coverage of the entire cDNA contig were selected. Six cDNAs of the thirty-three cDNAs screened provided satisfactory results. These cDNAs corresponded to genes encoding a glyceraldehyde-3-phosphate dehydrogenase (GAPDH), photosystem II reaction center W protein (PSIIw), oxygen evolving enhancer protein 3 (OEEp3), 40S ribosomal protein S12 (RB40S12), a photosystem I light harvesting complex protein (PSILHC), and the photosystem II PsbO protein (PSBo).

[00191] Promoter fragment sequences from these six genes and a terminator fragment sequence from the GAPDH gene were identified using GENESCAN® software (Applied Biosystems). Primers were designed for cloning the promoter and terminator fragment sequences. As the genomic sequence upstream of the GAPDH gene as delineated in the WT-105 genome assembly included a gap, the complete upstream sequence was cloned from genomic DNA by PCR using the GAPDH upstream primers.

Example 5. Functional verification of cloned regulatory sequences in P-925

[00192] The putative promoter fragments isolated from the 5' upstream region of the genes as identified by bioinformatic analysis of a draft genome sequence of WT-105, a wild-type strain of *Tetraselmis*, were cloned into constructs for promoter function testing. Promoter fragments were cloned as 800 bp to 1.5 kbp of sequence upstream of an identified *Tetraselmis* gene, such that the 3' end of the putative promoter fragment corresponded to the 3'-most base pair of the 5' untranslated region of the *Tetraselmis* gene (that is, the 3' end of each promoter fragment was the nucleotide immediately upstream of the initiating ATG codon of the identified gene). In addition, a 1223 bp fragment 3' of the stop codon of the WT-105 GAPDH gene (SEQ ID NO:10) was cloned as a terminator fragment.

[00193] Upstream regions of genes that were cloned for promoter fragment testing included: the glyceraldehyde 3-phosphate dehydrogenase gene upstream fragment GAPDH_P (SEQ ID NO:2); the oxygen evolving enhancer protein 1 gene upstream fragment PSBo_P (SEQ ID NO:3); photosystem II reaction center W gene upstream fragment PsIIW_P (SEQ ID NO:4); oxygen evolving enhancer protein 3 OEEp3_P (SEQ ID NO:5); 40S ribosomal protein S12 RB40S12_P (SEQ ID NO:6); and photosystem I light harvesting complex protein gene upstream fragment PSILHCdirect_P (SEQ ID NO:7). In addition, the proximity of the putative promoter fragment PSILHCdirect_P (SEQ ID NO:7) to an oppositely-oriented gene suggested that the fragment might include a bidirectional promoter. Therefore, the putative promoter-containing fragment PSILHCdirect_P (SEQ ID NO:7) was cloned in the reverse orientation (designated as PSILHCinverted_P and provided as SEQ ID NO:8) and was also tested for promoter function. The actin gene upstream fragment (SEQ ID NO:1) that demonstrated promoter activity in **Example 2** was also further assessed in these transformation experiments.

[00194] Putative promoter fragments were placed upstream from the *Sh ble* gene (SEQ ID NO:30), which was positioned immediately upstream of the WT-105 actin terminator fragment (SEQ ID NO:9) (see **Figure 7**). Vectors were sequence confirmed and were linearized with restriction enzymes and transformed into *Tetraselmis* strain P-925, a

classically improved derivative of WT-105 with increased lipid productivity relative to WT-105.

Example 6. Transformation into *Tetraselmis* strain P-925

[00195] *PDS-1000 biolistics transformation:* Although a large set of parameters were tested with the Biorad PDS-1000 gene gun for transformation of *Tetraselmis* (see, for example, **Example 2**), low numbers (*e.g.*, 0–4) of Zeocin resistant colonies per bombardment resulted regardless of the conditions used.

[00196] *HELIOS® biolistics transformation:* Plasmid DNAs were isolated from overnight *E. coli* cultures using the Invitrogen CHARGESWITCH® Pro Filter Plasmid Maxiprep Kit and linearized with an appropriate endonuclease from New England Biolabs. Linearized plasmids were purified using a phenol/chloroform (pH 8) extraction and precipitation in 70% cold ethanol and quantitated with the Invitrogen Qubit 2.0 fluorometer.

[00197] Cartridges for the HELIOS® Gene Gun (Bio-Rad) were prepared according to the manufacturer's instructions. In brief, CaCl_2 and spermidine were used to precipitate 1 μg linearized plasmid onto 500 μg of 600 nm BioRad gold microcarriers per cartridge. DNA-coated microcarriers were suspended in 100% ethanol with 10 $\mu\text{g}/\text{mL}$ PVP and immobilized to the inside of GOLD-COAT™ (Bio-Rad) tubing. For every 40 bullets, 20 mg gold was measured into a 1.5 mL tube to achieve 500 μg gold/shot. 100 μL of 50 mM spermidine was added, vortexed, and sonicated for 5 seconds. Plasmid DNA was added and vortexed. 100 μL of 1 M CaCl_2 was then added dropwise while vortexing. This mixture was incubated at room temperature for 10 minutes. Meanwhile, 30" Tefzel tubing was simultaneously dried by running nitrogen through it for 15 minutes at 0.3–0.4 LPM. The gold was then centrifuged for ~10–15 seconds. The supernatant was removed and the pellet was washed three times with 1 mL ethanol, with vortexing and centrifugation between each wash step. The pellet was resuspended in 1 mL ethanol/PVP solution (10 $\mu\text{g}/\text{mL}$ of PVP in ethanol). This 1 mL slurry was then moved to a 15 mL conical tube and an additional 1.5 mL of ethanol/PVP was added. The gold suspension

was vortexed and immediately pulled into the Tefzel tubing with a syringe, leaving several inches of space on both sides. Having checked to verify that the nitrogen was off, the Tefzel tubing was placed into a Prep Station, leaving the free end of the Tefzel tubing sealed with the syringe connection. The tube of gold was incubated for 5 minutes to allow the gold to settle and adhere to the Tefzel tubing. Suction was then applied with the syringe to remove the ethanol from the tubing at a rate of approximately 100 μ L/second. Once the ethanol was all removed, the tubing was immediately rotated approximately 120° every five seconds for a minute. After that first minute, nitrogen was introduced at 0.1 LPM and rotation continued for another minute. After the second minute, nitrogen was increased to approximately 0.3–0.4 LPM and rotation continued for another five minutes. The tubing was then cut into half-inch cartridges and place in a sealed storage bottle with desiccant for storage at 4°C.

[00198] Two days prior to transformation, cultures of P-925 were grown to mid-log phase (between 5×10^5 and 1×10^6 cells/mL) in PM032 media. Cell counts were determined using an Accuri cytometer. Cells were grown at 25°C, 1% CO₂ with shaking at 125 rpm on a 16:8 light:dark cycle. On the day of transformation, cells were harvested by centrifugation at $3000 \times g$ for 5 minutes at 25°C and re-suspended in PM032 to a concentration of 5×10^7 cells/mL. Fifteen 4 cm diameter circles with 1×10^7 cells per circle were spread on a 22 x 22 cm PM032 1.5% agar plate and allowed to dry. Fifteen replicate shots were fired with the HELIOS® Gene Gun (Bio-Rad) at 400 psi and 500 psi at a distance of 5 cm. The shooting plates were left at 25°C in ambient light for 24 hours to allow cells to recover. Cells from the fifteen replicates were washed off the shooting plates with 40 mL PM032 and pooled together in a 50 mL BD Falcon tube. The samples were then pelleted at $3000 \times g$ for 5 minutes at 25°C and re-suspended with PM032 to a final volume of 4 mL. The suspension was spread onto two 22 x 22 cm PM032 plates (2 mL cell suspension per plate) with 250 μ g/mL Zeocin. After air-drying in a biological safety cabinet, plates were incubated under 100 μ E light at 25°C. After two weeks, colonies present on the primary selection plates were streaked for secondary selection on PM032 agar with Zeocin at 800 μ g/mL. Using this method the following promoter fragments were validated: SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, and SEQ ID NO:8. The number of Zeocin 250 μ g/mL colonies were

recorded, and a random selection were re-streaked on Zeocin 800 µg/mL containing plates. Of those that grew on Zeocin 800 µg/mL, colony PCR was performed to confirm presence of *Sh ble* gene. A summary of these results are presented in **Table 5** below. All promoter fragments tested demonstrated the ability to direct expression of the *Sh ble* gene, as was evident by the high percentages of colonies that were resistant to a high concentration (800 µg/ml) of Zeocin.

Table 5. Screening of HELIOS® Gene Gun Transformants

Promoter fragment	# Resistant to 250 µg/mL Zeocin	% Resistant to 800 µg/mL Zeocin	% Positive by PCR screen
SEQ ID NO:1	629	25% (N = 629)	95% (N = 145)
SEQ ID NO:3	1150	97% (N = 32)	70% (N = 31)
SEQ ID NO:4	415	100% (N = 32)	100% (N = 32)
SEQ ID NO:6	4150	97% (N = 32)	79% (N = 31)
SEQ ID NO:7	1250	97% (N = 32)	92% (N = 31)
SEQ ID NO:8	161	100% (N = 32)	100% (N = 32)

Example 7. Cloned regulatory sequences in P-925 driving expression of a heterologous gene encoding a metabolic enzyme

[00199] The GAPDH promoter fragment (SEQ ID NO:2) and *psbO* promoter fragment (SEQ ID NO:3) were tested in *Tetraselmis* for their ability to drive expression of a heterologous gene encoding a metabolic enzyme. In these experiments, *Tetraselmis* strain P-925 was transformed with a plasmid that included a gene encoding a cyanobacterial sedoheptulose biphosphatase (SBPase) engineered to include an N-terminal chloroplast stroma targeting sequence from the WT-105 SBPase (SEQ ID NO:34) and a C-terminal FLAG tag, in which the gene was operably linked to either the GAPDH promoter fragment (SEQ ID NO:1) or the *psbO* promoter fragment (SEQ ID NO:3). A general diagram of the p05841 and p05842 constructs that included the *Synechococcus* PCC7942 SBPase gene tagged at the C-terminus with a FLAG epitope tag and fused at the N-terminus with a stromal targeting sequence from the endogenous WT-105 SBPase gene, followed by the GAPDH terminator fragment (SEQ ID NO:10) is provided as **Figure 8**.

In the experiments detailed herein, the “promoter of interest” in **Figure 8** is the GAPDH promoter fragment (SEQ ID NO:2) for construct p05841 and the *psbO* promoter fragment (SEQ ID NO:3) for construct p05842.

[00200] *Construction of p05481 and p05482 vectors for over-expression of a cyanobacterial SBPase:* Two vectors, p05464 and p05465, were synthesized for over-expressing heterologous genes in *Tetraselmis*. In addition to either the GAPDH promoter fragment (SEQ ID NO:2) or the *psbO* promoter fragment (SEQ ID NO:3), paired with the GAPDH terminator fragment (SEQ ID NO:10) for target gene expression, these plasmids both included the Zeocin resistance cassette derived from p05320 (the *Sh ble* gene operably linked to the WT-105 actin promoter fragment (SEQ ID NO:1) and the WT-105 actin terminator fragment (SEQ ID NO:9)) (see **Figure 8**). p05464 was assembled by PCR amplifying and gel purifying DNA fragments, which were eventually assembled into the vector. Each fragment was designed to have 15–40 bases of overlap to neighboring fragments to enable homology based cloning. The following DNA fragments were used to assemble p05464: A) a 2.1 kbp cloning backbone from pUC19 vector containing the beta-lactamase selectable marker and origin of replication for *E.coli*; B) an algal selectable marker cassette containing the bleomycin resistance gene (SEQ ID NO:30) under the control of WT-105 actin promoter (SEQ ID NO:1) and WT-105 actin terminator (SEQ ID NO:9); as well as C) a WT-105 GAPDH promoter fragment (SEQ ID NO:2); and D) a WT-105 GAPDH terminator fragment (SEQ ID NO:10) positioned in tandem such that a gene of interest could be inserted between the GAPDH promoter (C) and the GAPDH terminator (D). Fragments A–D were assembled by Gibson Assembly to produce expression vector p05464. Vector p05465 was assembled in the same manner, but with a WT-105 *psbO* promoter fragment for element C (SEQ ID NO:3). The *Synechococcus elongatus* PCC7942 sedoheptulose-1,7-bisphosphatase encoding gene (SBPase, Genbank accession: BAA08536, SEQ ID NO:33) was inserted into expression vector p05464 downstream of the GAPDH promoter fragment to create p05481 and inserted into expression vector p05465 downstream of the *psbO* promoter fragment to create p05482. In both constructs, a nucleotide sequence encoding an N-terminal 61 amino acid chloroplast targeting sequence (SEQ ID NO:34) from the WT-105 SBPase gene (SEQ ID NO:33) and a nucleotide sequence encoding a C-terminal FLAG epitope

tag were added upstream and downstream of the SBPase encoding sequence, respectively, using the Gibson Assembly method.

[00201] *Tetraselmis* P-925 cells were transformed with a using a HELIOS® gene gun according to the method provided in Example 6. Zeocin resistant colonies were selected as described above, and lines were further assessed for SBPase protein expression by western blotting.

[00202] Accumulation of the recombinant protein in transgenic lines was detected using by immunoblotting with an anti-FLAG epitope tag antibody (see **Figure 9**). Total cellular proteins were extracted in EZ buffer (Martinez-Garcia *et al.* (1999) *Plant J.* 20:251–57) by treatment in a water bath sonicator for approximately 30 seconds, followed by denaturation at 85°C for 10 minutes with occasional vortexing. Total protein concentration was determined using the BioRad RCDC protein assay kit. SDS-PAGE gels were loaded on an equal total protein basis and proteins were separated by electrophoresis. Proteins were transferred to PVDF membranes using the Invitrogen iBLOT® dry blot system (Life Technologies, Carlsbad, CA) using program #3, for 7 minutes. Membranes were blocked for at least 90 minutes in tris-buffered saline (TBS) containing 5% powdered milk and 0.15% Tween 20 (TBST), followed by incubation with primary antibodies and HRP-conjugated secondary antibody (Biorad goat anti-rabbit IgG, diluted 1:40000 in TBST-milk). Antibodies used in this report included anti-AGPase (AS11 1739, Agrisera), rabbit polyclonal peptide antibody recognizing FLAG epitope tags, and Bip2 (AS09 481, Agrisera). Immuno-reacting bands were detected with Chemiluminescent peroxidase substrate (Sigma-Aldrich) and imaged with the MULTIIMAGE® II (Alpha Innotech) gel imaging station.

[00203] A ~40 kDa band (the expected mass if the chloroplast targeting sequence is properly cleaved) was observed in 19 of 49 p05481 derived lines, and 5 of 23 p05482 derived lines. These results indicate that transgenic protein was detected in ~20–40% of transgenic lines, as summarized in **Table 6** below. As such, the promoter function of both the GAPDH promoter fragment (SEQ ID NO:2) and the *psbO* promoter fragment (SEQ ID NO:3) has been confirmed.

Table 6. Summary of results from driving transgene expression

Promoter fragment	# Lines tested	% Zeocin ^R & FLAG ⁺
GAPDH	50	42
<i>psbO</i>	20	20

[00204] Regulatory elements have been identified from *Tetraselmis* to use in recombinant protein expression in eukaryotic cells, particularly microalgae. Further, the inventors have discovered that transformation of *Tetraselmis* using a biolistic method results in greater than 100 transformants per day even though *Tetraselmis* has been known to be difficult to stably transform. Transgenic protein over-expression was shown to be readily achievable in *Tetraselmis* using the transformation protocol established in this report.

[00205] Example embodiments are provided so that this disclosure will be thorough, and will fully convey the scope to those who are skilled in the art. Numerous specific details are set forth such as examples of specific components and methods, to provide a thorough understanding of embodiments of the present disclosure. It will be apparent to those skilled in the art that specific details need not be employed, that example embodiments may be embodied in many different forms and that neither should be construed to limit the scope of the disclosure. In some example embodiments, well-known processes, well-known structures, and well-known technologies are not described in detail.

[00206] An index of all of the sequences detailed in this application is set forth below in **Table 7**.

Table 7. Index of Sequence Listings

SEQ ID NO:	Description	SEQ ID NO:	Description
1	<i>Tetraselmis</i> Actin promoter fragment	19	WT105ActinTER-GFP-5 primer
2	<i>Tetraselmis</i> GAPDH promoter fragment	20	WT105ActinTER-PciI primer
3	<i>Tetraselmis psbO</i> promoter fragment	21	GFP-ActinWT105-5 primer
4	<i>Tetraselmis</i> PS-II W promoter fragment	22	GFP-ActinWT105-3 primer
5	<i>Tetraselmis</i> OEE p3 promoter fragment	23	BLE-5 primer

6	<i>Tetraselmis</i> RB40 S12 promoter fragment	24	BLE-3 primer
7	<i>Tetraselmis</i> PS-I HLC direct promoter fragment	25	NaHtransB5 primer
8	<i>Tetraselmis</i> PS-I HLC inverted promoter fragment	26	NaHtransC3 primer
9	<i>Tetraselmis</i> Actin terminator fragment	27	LF-SGE5320F1 primer
10	<i>Tetraselmis</i> GAPDH terminator fragment	28	BLEcodon3 primer
11	WT105ActinPro-EcoRI primer	29	p05320
12	WT105ActinPro-BLE-3 primer	30	<i>Sh ble</i> gene
13	WT105ActinTER-BLE-5 primer	31	<i>C. reinhardtii</i> actin gene
14	WT105ActinTER-HindIII primer	32	<i>C. reinhardtii</i> <i>psaD</i>
15	BLE-ActinWT105-5 primer	33	WT-105 SBPase gene
16	BLE-ActinWT105-3 primer	34	WT-105 SBPase
17	WT105ActinPro-PciI primer	35	WT-105 SBPase chloroplast targeting sequence gene
18	WT105ActinPro-GFP-3 primer	36	WT-105 SBPase chloroplast targeting sequence

CLAIMS:

1. An expression cassette comprising a heterologous nucleic acid sequence encoding a protein or functional RNA, operably linked to a promoter comprising a nucleotide sequence having at least 95% identity to at least 300 contiguous nucleotides at the 3' end of SEQ ID NO:6, wherein the at least 300 contiguous nucleotides have promoter activity.
2. An expression cassette comprising a heterologous nucleic acid sequence encoding a protein or functional RNA, operably linked to a promoter comprising a nucleotide sequence having 100% identity to at least 100 contiguous nucleotides at the 3' end of SEQ ID NO:6, wherein the at least 100 contiguous nucleotides have promoter activity.
3. The expression cassette according to claim 1 or 2, wherein the promoter is a constitutive promoter.
4. The expression cassette according to any one of claims 1 to 3, wherein the nucleic acid sequence encoding a protein or functional RNA is further operably linked to a terminator comprising a nucleotide sequence selected from the group consisting of a nucleotide sequence having at least 95% identity to SEQ ID NO:9 and a nucleotide sequence having at least 95% identity to SEQ ID NO:10.
5. The expression cassette according to any one of claims 1 to 4, wherein the nucleic acid sequence encoding a protein or functional RNA encodes a protein.
6. The expression cassette according to claim 5, wherein the nucleic acid sequence encoding a protein or functional RNA encodes (a) a protein associated with lipid biosynthesis, (b) a protein having lipolytic activity, (c) a photosynthetic protein, (d) a carbon fixation protein, (e) a transporter protein, (f) a dehydrogenase, (g) a transcription factor, or (h) a cell signaling protein.
7. The expression cassette according to any one of claims 1 to 4, wherein the nucleic acid sequence encoding a protein or functional RNA encodes a functional RNA selected from the group consisting of an antisense sequence, a microRNA, a shRNA, an siRNA, and a ribozyme.
8. A vector comprising the expression cassette according to any one of claims 1 to 3.

9. The vector according to claim 8, further comprising a selectable marker or reporter gene.
10. The vector according to claim 8 or 9, further comprising a terminator comprising a nucleotide sequence selected from the group consisting of a nucleic acid sequence having at least 95% identity to SEQ ID NO:9 and a nucleic acid sequence having at least 95% identity to SEQ ID NO:10; wherein the terminator is operably linked to the nucleic acid sequence encoding a protein or a functional RNA and/or the terminator is downstream of the nucleic acid sequence encoding a protein or a functional RNA.
11. The vector according to claim 10, wherein the terminator comprises the nucleotide sequence having at least 95% identity to SEQ ID NO:9.
12. The vector according to claim 10, wherein the terminator comprises the nucleotide sequence having at least 95% identity to SEQ ID NO:10.
13. The vector according to any one of claims 8 to 12, wherein the vector comprises at least one origin of replication.
14. The vector according to claim 9, further comprising a second promoter operably linked to the selectable marker or reporter gene.
15. The vector according to claim 9, wherein the selectable marker or reporter gene is selected from the group consisting of a gene conferring resistance to an antibiotic, a gene conferring resistance to an herbicide, a gene encoding acetyl CoA carboxylase (ACCase), a gene encoding acetohydroxy acid synthase (ahas), a gene encoding acetolactate synthase, a gene encoding aminoglycoside phosphotransferase, a gene encoding anthranilate synthase, a gene encoding bromoxynil nitrilase, a gene encoding cytochrome P450-NADH-cytochrome P450 oxidoreductase, a gene encoding dalapon dehalogenase, a gene encoding dihydropteroate synthase, a gene encoding a class I 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a gene encoding a class II EPSPS, a gene encoding a non-class I/II EPSPS, a gene encoding glutathione reductase, a gene encoding glyphosate acetyltransferase, a gene encoding glyphosate oxidoreductase, a gene encoding hydroxyphenylpyruvate dehydrogenase, a gene encoding hydroxy-phenylpyruvate dioxygenase, a gene encoding isoprenyl pyrophosphate isomerase, a gene encoding lycopene cyclase, a gene encoding phosphinothricin acetyl transferase, a gene encoding phytoene desaturase, a gene encoding prenyl transferase, a gene encoding

protoporphyrin oxidase, a gene encoding superoxide dismutase, *arg7*, *his3*, *hisD*, *hisG*, *manA*, *nit1*, *trpB*, *uidA*, *xylA*, a dihydrofolate reductase gene, a mannose-6-phosphate isomerase gene, a nitrate reductase gene, an ornithine decarboxylase gene, a thymidine kinase gene, a 2-deoxyglucose resistance gene, an R-locus gene, a tyrosinase gene, *lacZ*, an alkaline phosphatase gene, an α -amylase gene, a horseradish peroxidase gene, an α -galactosidase gene, a luciferin/luciferase gene, a beta-glucuronidase gene (GUS), and a gene encoding a fluorescent protein.

16. A method for transforming a eukaryotic cell comprising: introducing the vector according to any one of claims 8 to 15 into the eukaryotic cell; and selecting for a transformed eukaryotic cell.

17. The method according to claim 16, wherein the vector is introduced by a biolistic procedure.

18. The method according to claim 17, wherein about 300 psi or more of pressure is used to impel biolistic microcarriers coated with vector DNA into the eukaryotic cell.

19. The method according to claim 16, wherein the transformed eukaryotic cell is stably transformed.

20. The method according to claim 16, wherein the eukaryotic cell is selected from the group consisting of a fungal cell, a heterokont cell, an algal cell, and a plant cell.

21. The method according to claim 16, wherein the eukaryotic cell is an algal cell.

22. The method according to claim 21, wherein the algal cell is from a species belonging to a genus selected from the group consisting of *Achnanthes*, *Amphiprora*, *Amphora*, *Ankistrodesmus*, *Asteromonas*, *Boekelovia*, *Bolidomonas*, *Borodinella*, *Botrydium*, *Botryococcus*, *Bracteococcus*, *Chaetoceros*, *Carteria*, *Chlamydomonas*, *Chlorococcum*, *Chlorogonium*, *Chlorella*, *Chroomonas*, *Chrysosphaera*, *Cricosphaera*, *Crypthecodinium*, *Cryptomonas*, *Cyclotella*, *Dunaliella*, *Ellipsoidon*, *Emiliania*, *Eremosphaera*, *Emodesmius*, *Euglena*, *Eustigmatos*, *Franceia*, *Fragilaria*, *Gloeothamnion*, *Haematococcus*, *Halocafeteria*, *Heterosigma*, *Hymenomonas*, *Isochrysis*, *Lepocinclis*, *Micractinium*, *Monoraphidium*, *Nannochloris*, *Nannochloropsis*, *Navicula*, *Neochloris*, *Nephrochloris*, *Nephroselmis*, *Nitzschia*, *Ochromonas*, *Oedogonium*, *Oocystis*, *Ostreococcus*, *Pavlova*, *Parachlorella*, *Pascheria*, *Pelagomonas*, *Phaeodactylum*, *Phagus*, *Picochlorum*, *Platymonas*, *Pleurochrysis*, *Pleurococcus*,

Prototheca, *Pseudochlorella*, *Pseudoneochloris*, *Pseudostaurastrum*, *Pyramimonas*, *Pyrobotrys*, *Scenedesmus*, *Schizochlamydeella*, *Skeletonema*, *Spyrogyra*, *Stichococcus*, *Tetraselmis*, *Thalassiosira*, *Tribonema*, *Vaucheria*, *Viridiella*, *Vischeria*, and *Volvox*.

23. The method according to claim 22, wherein the algal cell is a *Tetraselmis* algal cell.
24. A method for co-transforming a eukaryotic cell comprising:
 - introducing the expression cassette according to claim 1 and a nucleic acid sequence encoding a selectable marker into the eukaryotic cell;
 - selecting for the presence of the selectable marker in the eukaryotic cell; and
 - identifying a eukaryotic cell transformed with the expression cassette.
25. The method according to claim 24, wherein the eukaryotic cell is an algal cell.
26. The method according to claim 24 or 25, wherein the selectable marker is encoded by a gene selected from the group consisting of a gene conferring resistance to an antibiotic, a gene conferring resistance to an herbicide, a gene encoding acetyl CoA carboxylase (ACCase), a gene encoding acetohydroxy acid synthase (ahas), a gene encoding acetolactate synthase, a gene encoding aminoglycoside phosphotransferase, a gene encoding anthranilate synthase, a gene encoding bromoxynil nitrilase, a gene encoding cytochrome P450-NADH-cytochrome P450 oxidoreductase, a gene encoding dalapon dehalogenase, a gene encoding dihydropteroate synthase, a gene encoding a class I 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a gene encoding a class II EPSPS, a gene encoding a non-class I/II EPSPS, a gene encoding glutathione reductase, a gene encoding glyphosate acetyltransferase, a gene encoding glyphosate oxidoreductase, a gene encoding hydroxyphenylpyruvate dehydrogenase, a gene encoding hydroxy-phenylpyruvate dioxygenase, a gene encoding isoprenyl pyrophosphate isomerase, a gene encoding lycopene cyclase, a gene encoding phosphinothricin acetyl transferase, a gene encoding phytoene desaturase, a gene encoding prenyl transferase, a gene encoding protoporphyrin oxidase, a gene encoding superoxide dismutase, *arg7*, *his3*, *hisD*, *hisG*, *manA*, *nit1*, *trpB*, *uidA*, *xylA*, a dihydrofolate reductase gene, a mannose-6-phosphate isomerase gene, a nitrate reductase gene, an ornithine decarboxylase gene, a thymidine kinase gene, a 2-deoxyglucose resistance gene, and an R-locus gene.

27. The method according to any one of claims 24 to 26, wherein the selectable marker gene is operably linked to a second promoter active in the eukaryotic cell.
28. The method according to any one of claims 24 to 27, wherein the expression cassette and the nucleic acid sequence encoding a selectable marker are present in the same vector.
29. The method according to claim 25, wherein the algal cell is from a species belonging to a genus selected from the group consisting of *Achnanthes*, *Amphiprora*, *Amphora*, *Ankistrodesmus*, *Asteromonas*, *Boekelovia*, *Bolidomonas*, *Borodinella*, *Botrydium*, *Botryococcus*, *Bracteococcus*, *Chaetoceros*, *Carteria*, *Chlamydomonas*, *Chlorococcum*, *Chlorogonium*, *Chlorella*, *Chroomonas*, *Chrysosphaera*, *Cricosphaera*, *Crypthecodinium*, *Cryptomonas*, *Cyclotella*, *Dunaliella*, *Ellipsoidon*, *Emiliana*, *Eremosphaera*, *Ernodesmius*, *Euglena*, *Eustigmatos*, *Franceia*, *Fragilaria*, *Gloeothamnion*, *Haematococcus*, *Halocafeteria*, *Heterosigma*, *Hymenomonas*, *Isochrysis*, *Lepocinclis*, *Micractinium*, *Monoraphidium*, *Nannochloris*, *Nannochloropsis*, *Navicula*, *Neochloris*, *Nephrochloris*, *Nephroselmis*, *Nitzschia*, *Ochromonas*, *Oedogonium*, *Oocystis*, *Ostreococcus*, *Pavlova*, *Parachlorella*, *Pascheria*, *Pelagomonas*, *Phaeodactylum*, *Phagus*, *Picochlorum*, *Platymonas*, *Pleurochrysis*, *Pleurococcus*, *Prototheca*, *Pseudochlorella*, *Pseudoneochloris*, *Pseudostaurastrum*, *Pyramimonas*, *Pyrobotrys*, *Scenedesmus*, *Schizochlamydeella*, *Skeletonema*, *Spyrogyra*, *Stichococcus*, *Tetrachlorella*, *Tetraselmis*, *Thalassiosira*, *Tribonema*, *Vaucheria*, *Viridiella*, *Vischeria*, and *Volvox*.
30. The method according to claim 29, wherein the algal cell is a *Tetraselmis* algal cell.
31. A eukaryotic host cell comprising the expression cassette according to any one of claims 1 to 7.
32. The eukaryotic host cell according to claim 31, wherein the eukaryotic host cell is an algal cell.
33. A vector comprising the expression cassette according to any one of claims 4 to 7.
34. The vector according to claim 33, further comprising a selectable marker or reporter gene.

1/9

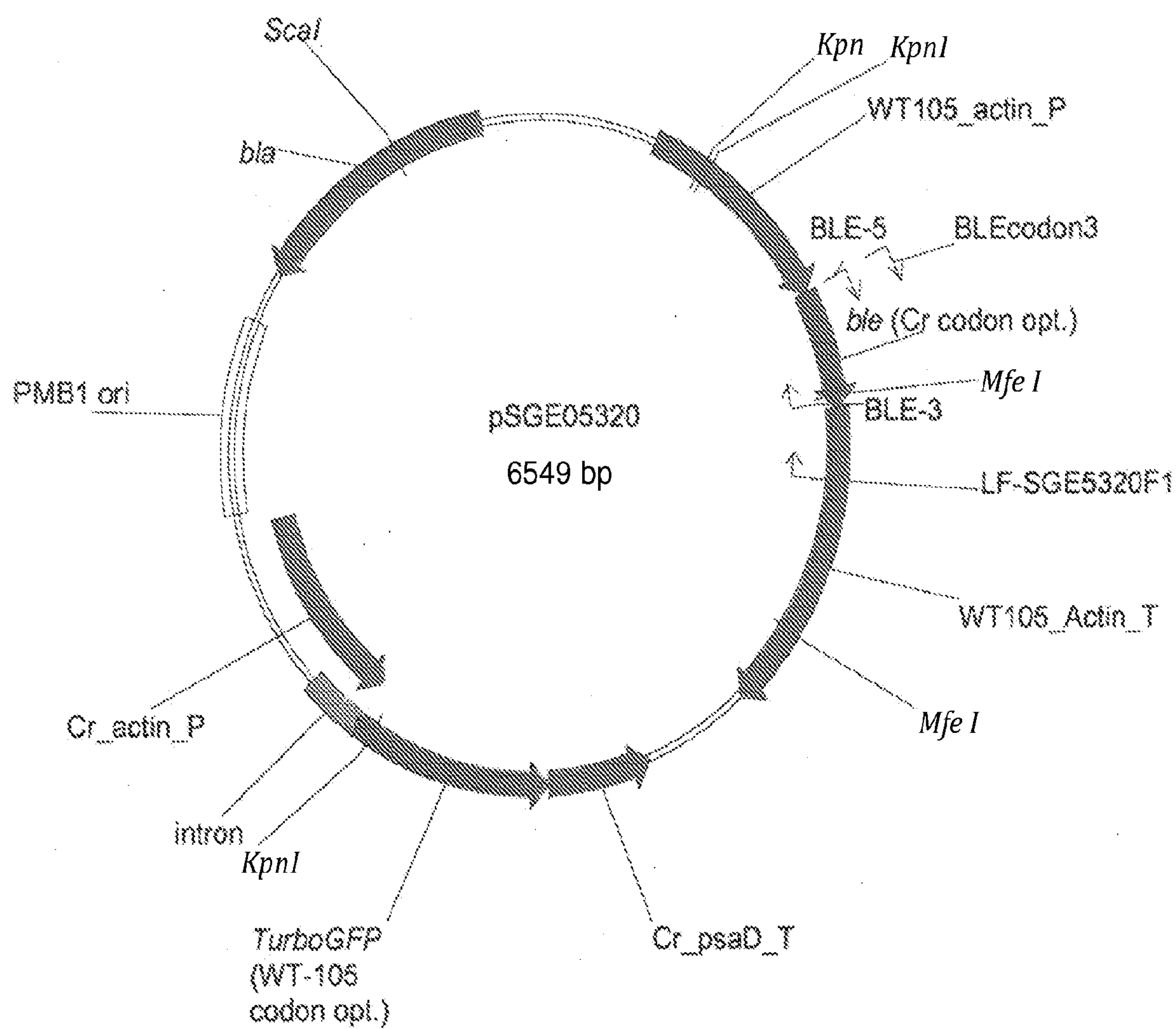


Figure 1

2/9

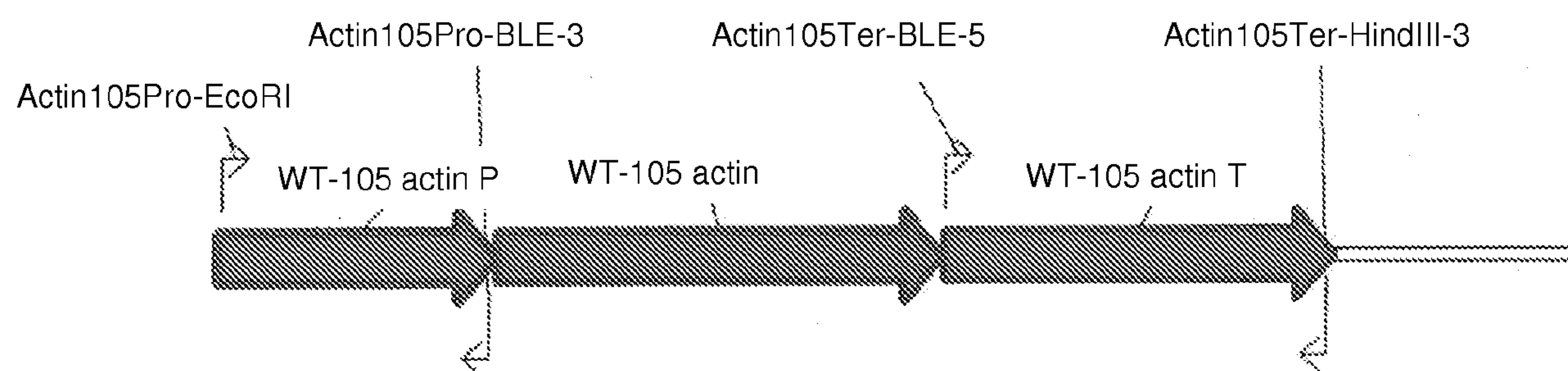
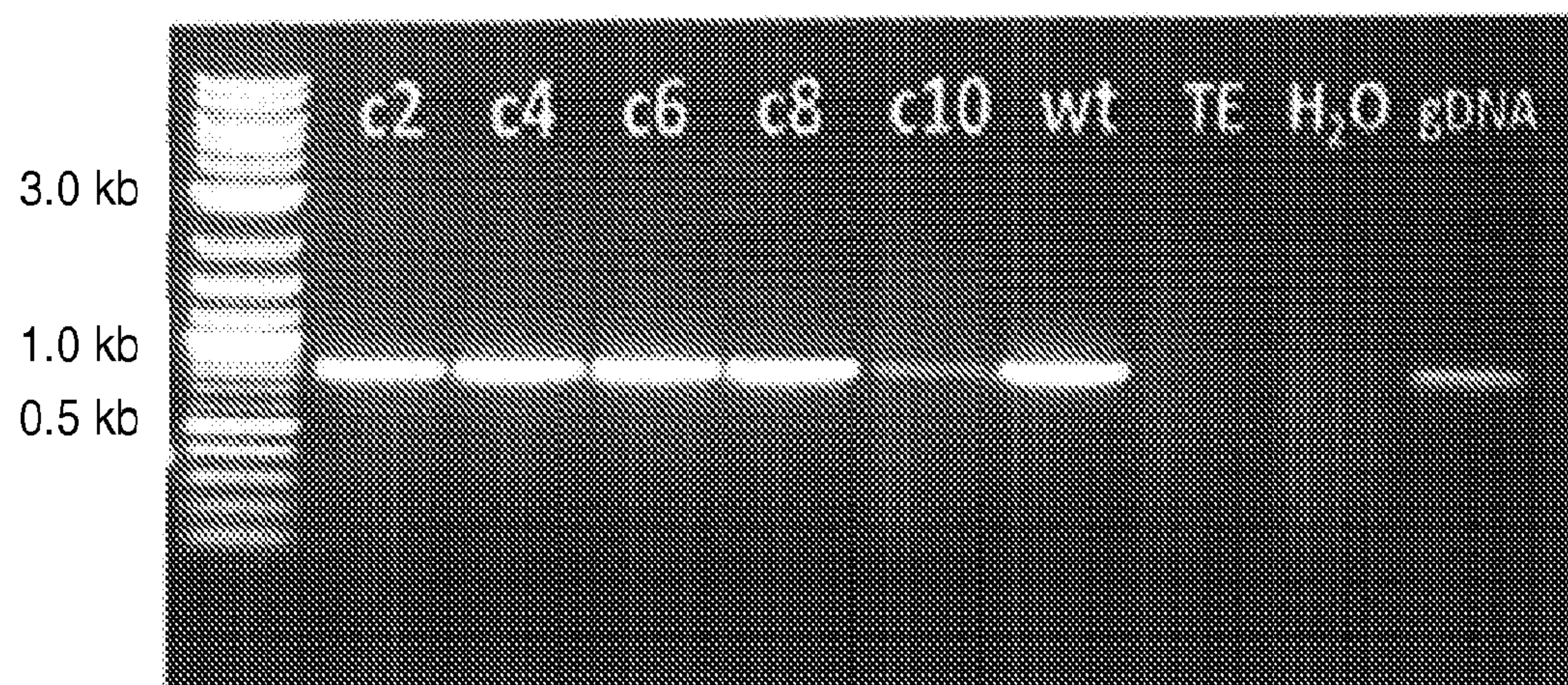


Figure 2

3/9

Control Na/H⁺ transporter ~700bp

Ble gene ~630bp

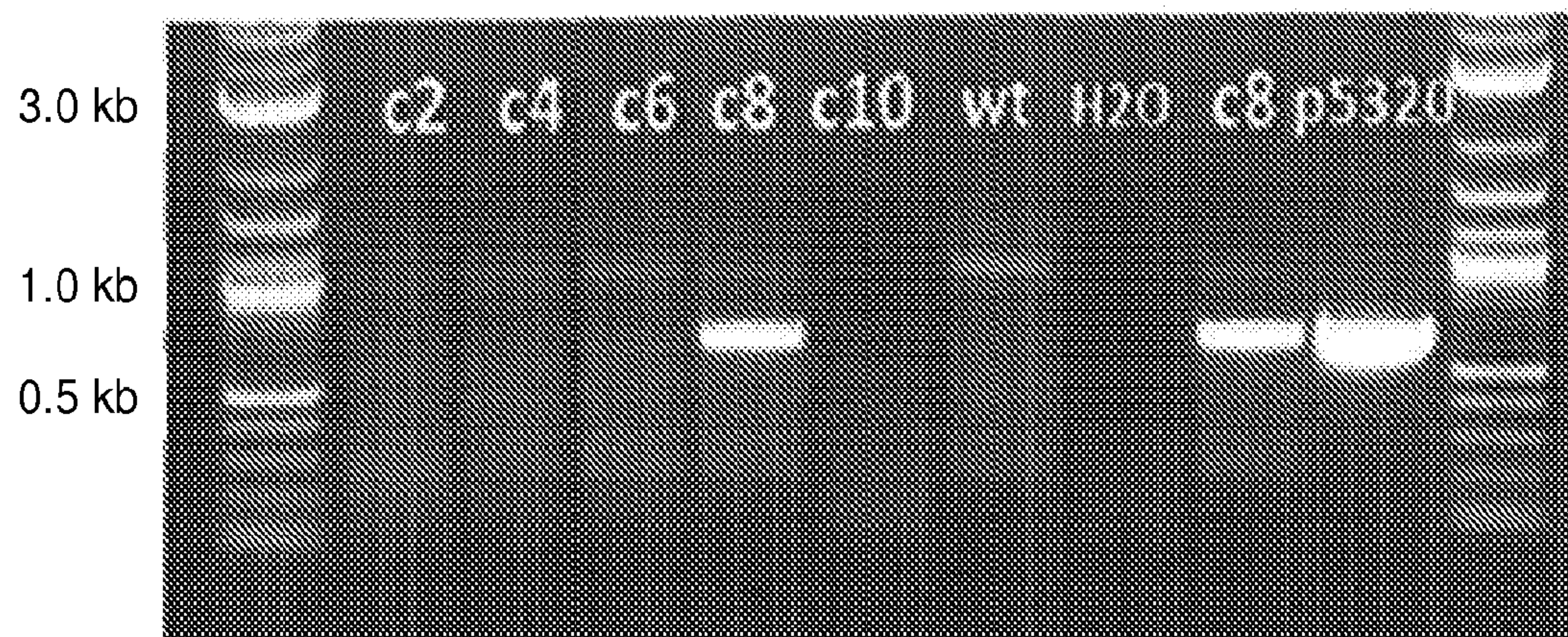


Figure 3

4/9

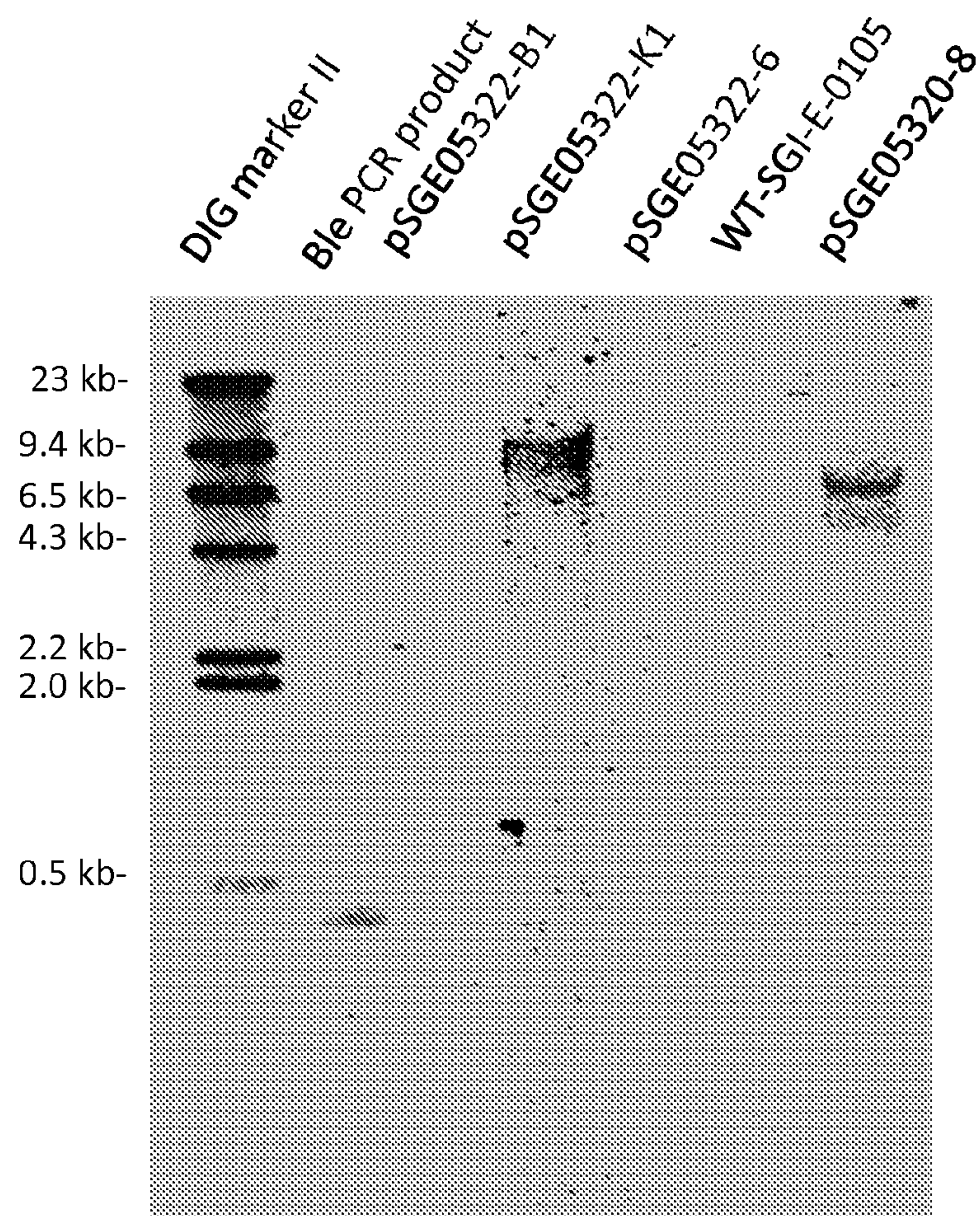


Figure 4

5/9

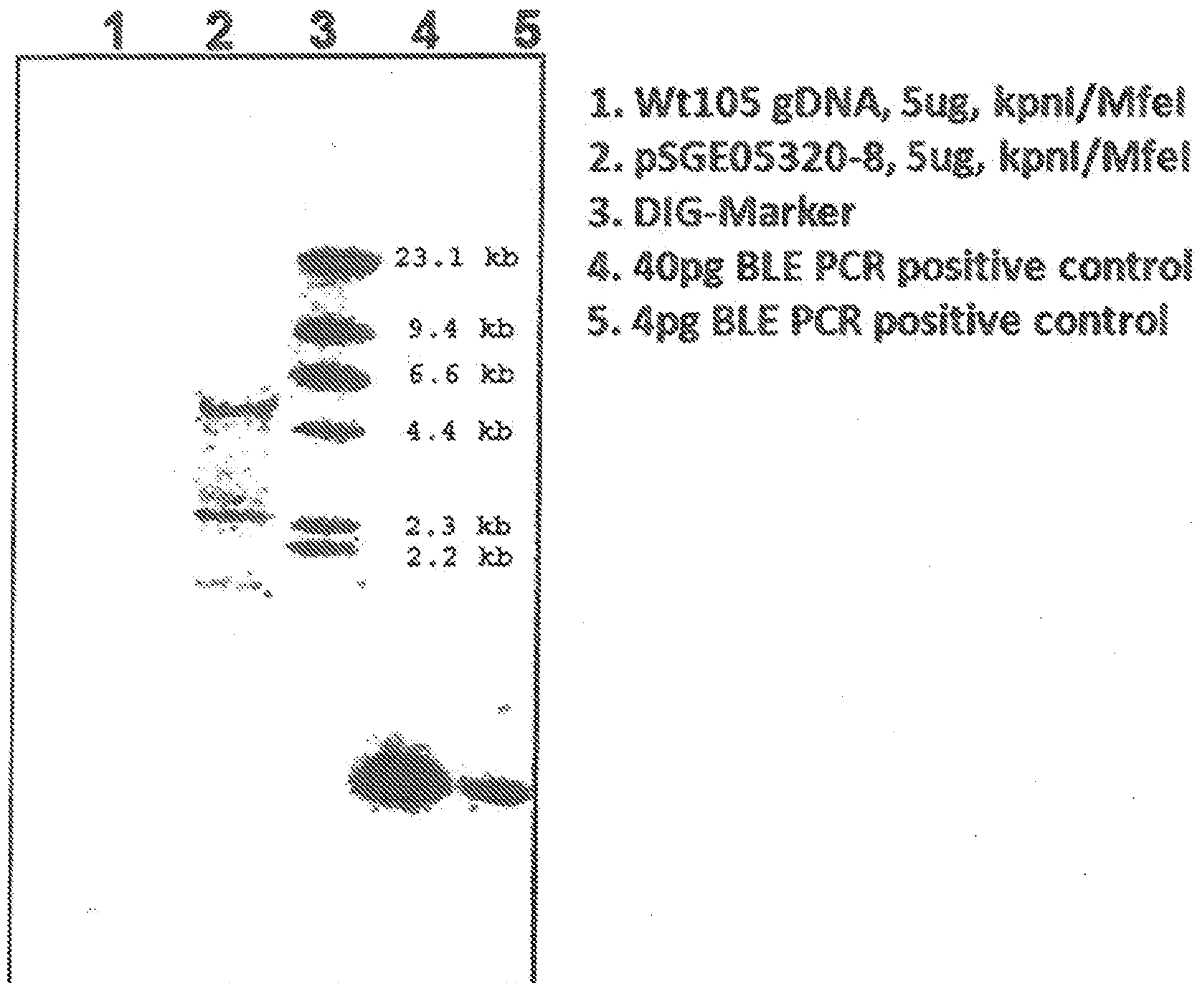


Figure 5

6/9

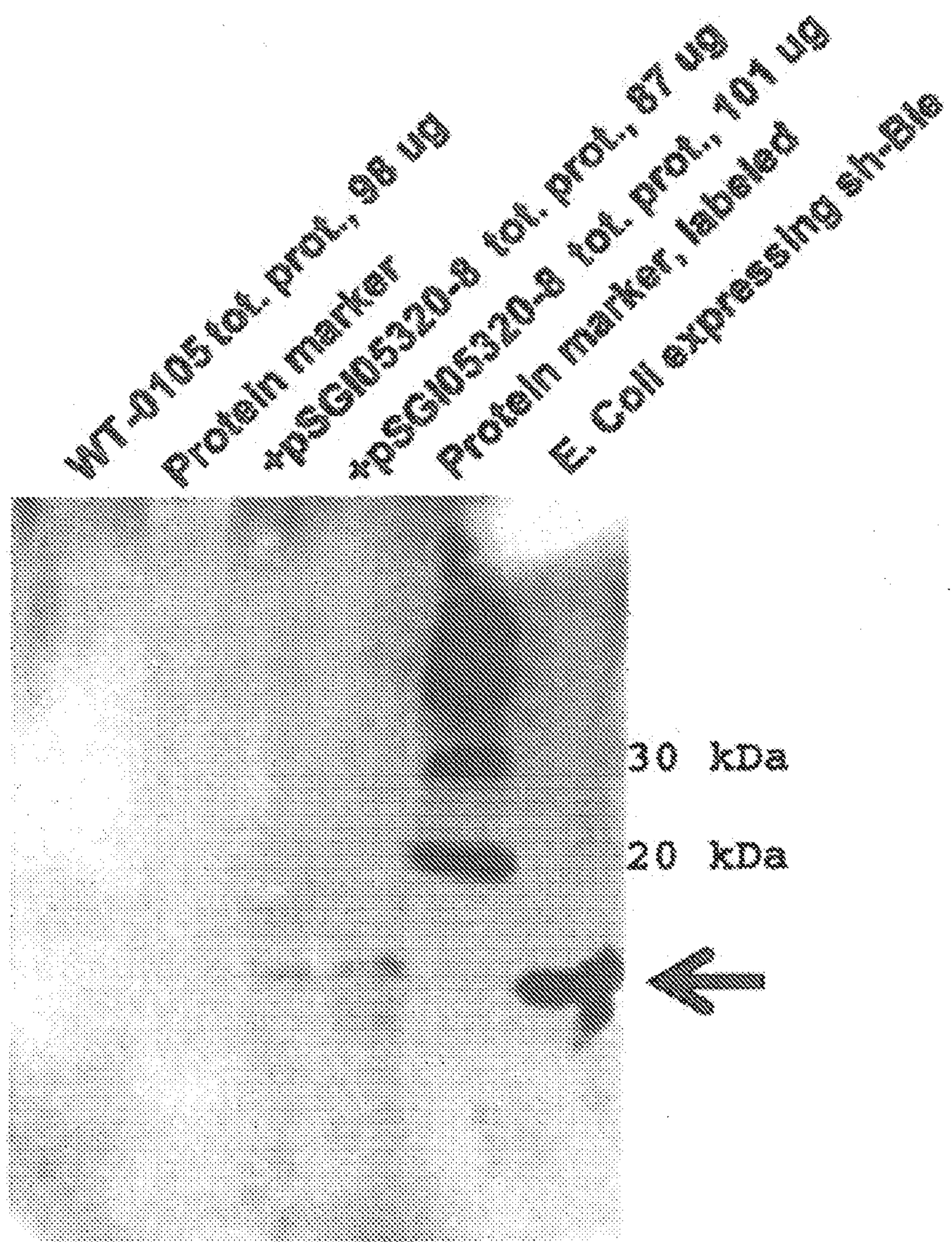


Figure 6

7/9

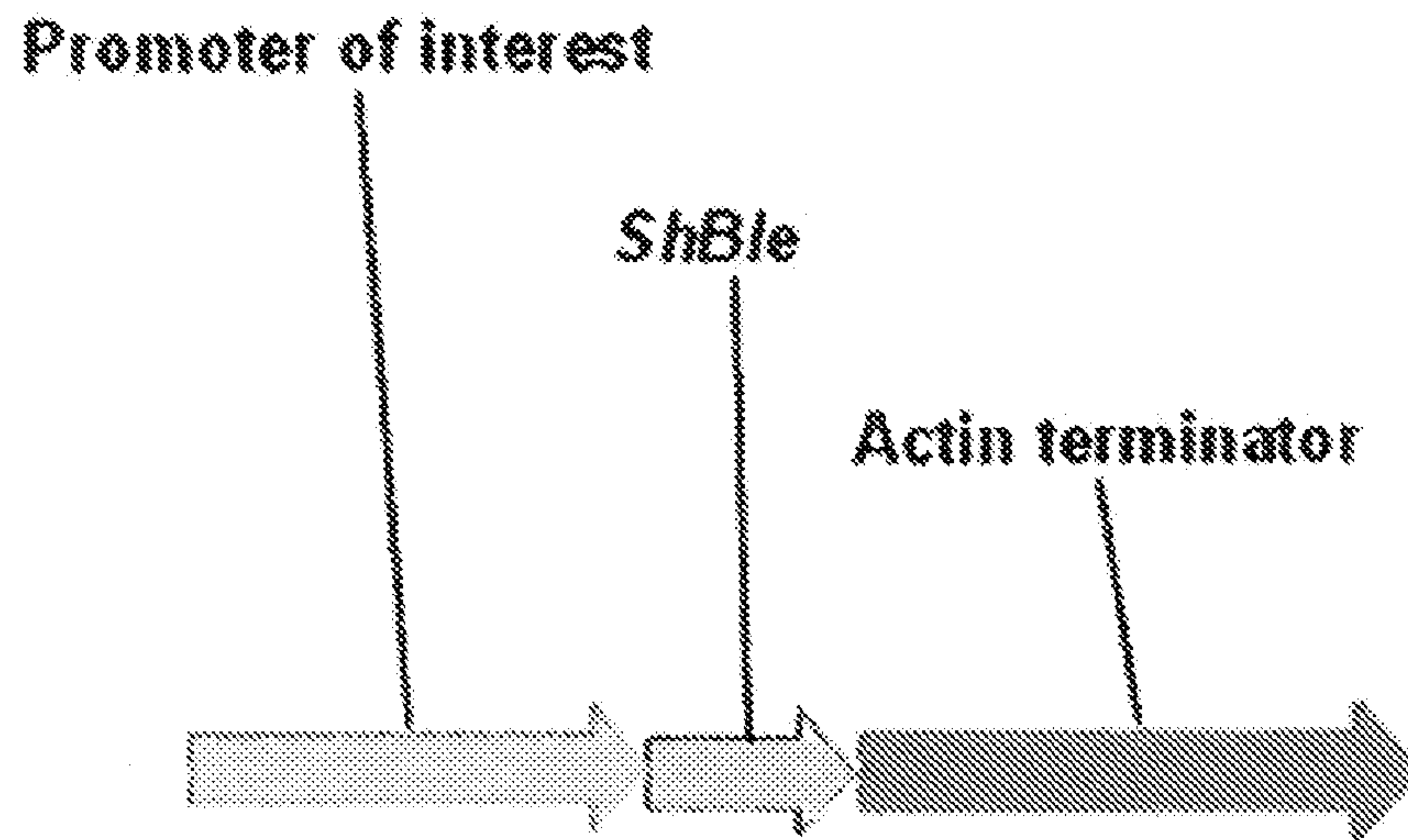


Figure 7

8/9

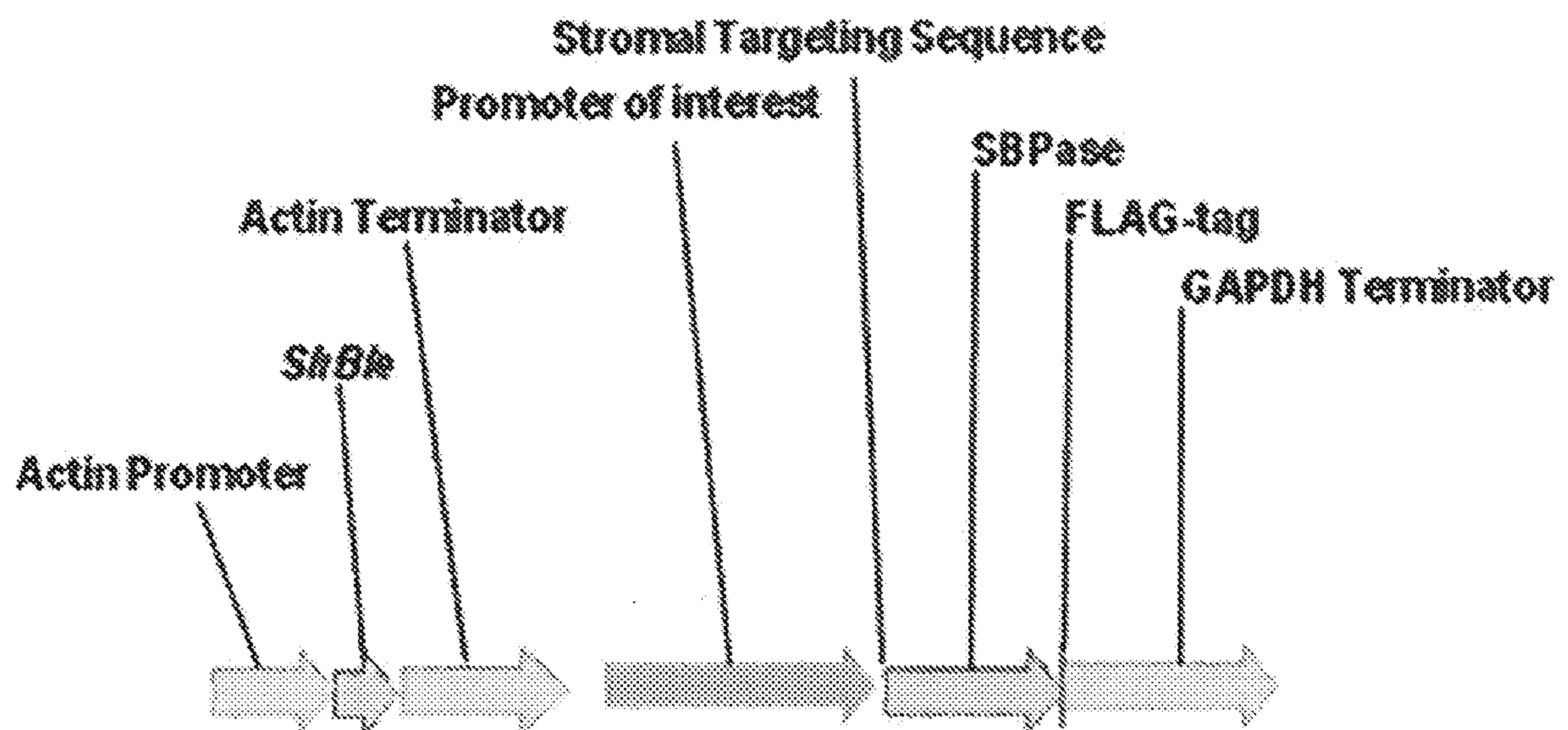


Figure 8

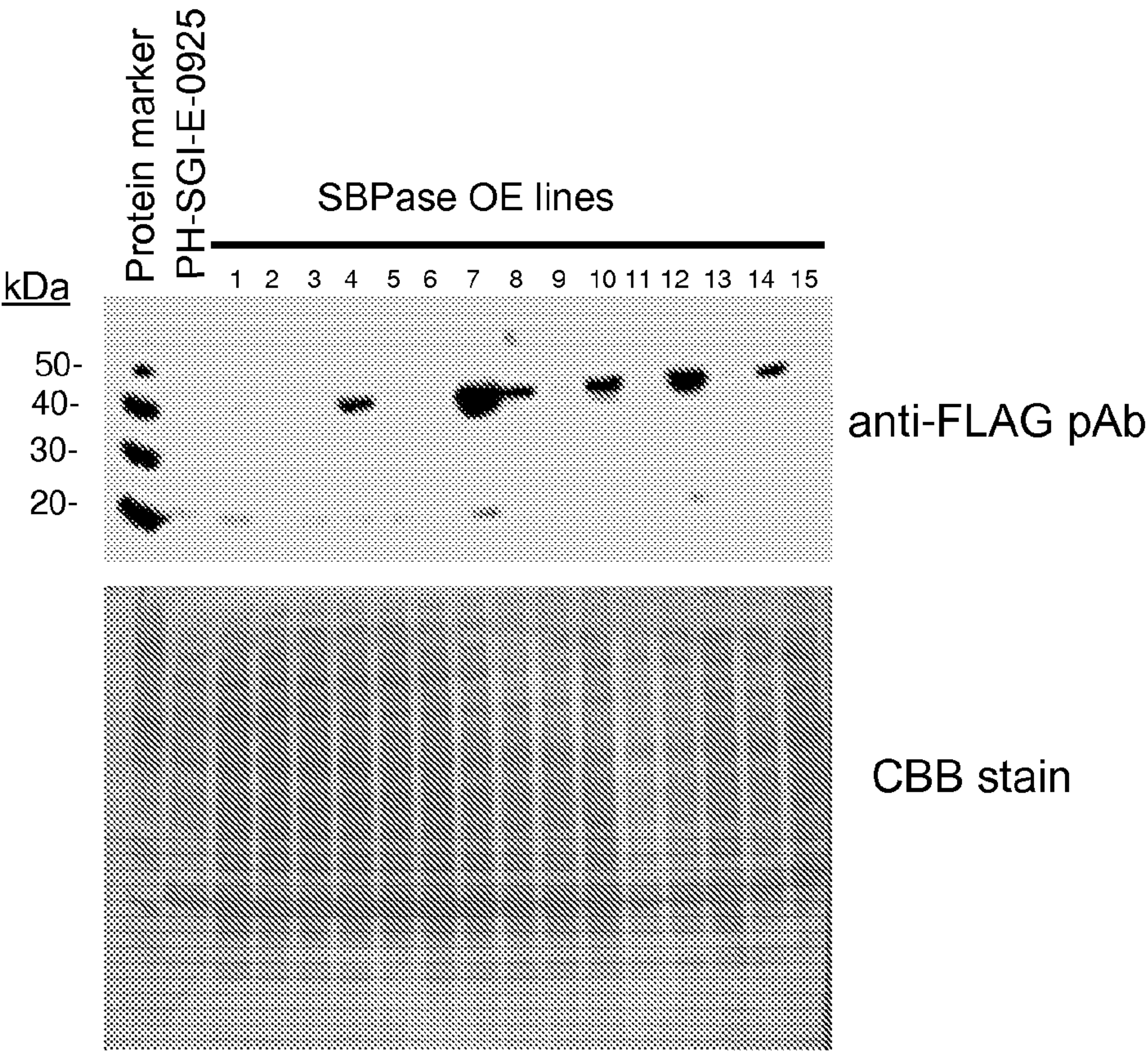


Figure 9