

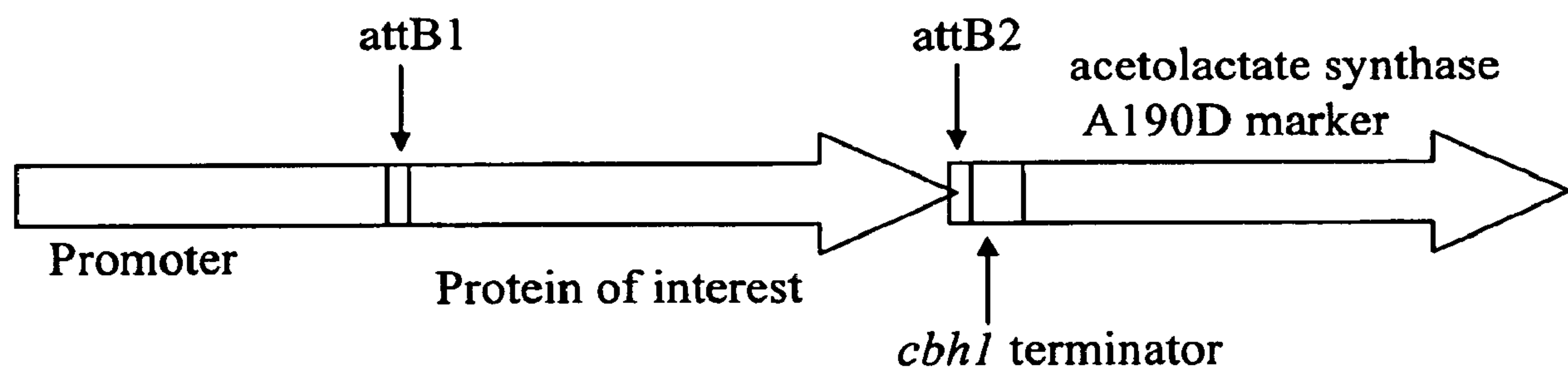


(86) Date de dépôt PCT/PCT Filing Date: 2007/09/21
(87) Date publication PCT/PCT Publication Date: 2008/04/03
(45) Date de délivrance/Issue Date: 2017/02/28
(85) Entrée phase nationale/National Entry: 2009/03/17
(86) N° demande PCT/PCT Application No.: US 2007/020484
(87) N° publication PCT/PCT Publication No.: 2008/039370
(30) Priorités/Priorities: 2006/09/22 (US60/846,804);
2006/09/22 (US60/846,656)

(51) Cl.Int./Int.Cl. *C12N 15/82* (2006.01),
C12N 9/88 (2006.01)
(72) Inventeurs/Inventors:
BOWER, BENJAMIN S., US;
DUNN-COLEMAN, NIGEL, US;
LEIVA, NICHOLAS, US
(73) Propriétaire/Owner:
DANISCO US INC., US
(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : MARQUEUR SELECTIONNABLE D'ACETOLACTATE SYNTHASE (ALS) A PARTIR DE TRICHODERMA REESEI

(54) Title: ACETOLACTATE SYNTHASE (ALS) SELECTABLE MARKER FROM TRICHODERMA REESEI



(57) Abrégé/Abstract:

A nucleic acid encoding an acetolactate synthase (ALS) protein that provides resistance to ALS inhibitors, e.g., sulphonylurea and imidazolinone compounds, is provided. The nucleic acid may be used as a selectable marker for expression of a protein of interest in host cells.

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 April 2008 (03.04.2008)

PCT

(10) International Publication Number
WO 2008/039370 A1

(51) International Patent Classification:
C12N 15/82 (2006.01) *C12N 9/88* (2006.01)

(74) Agent: KOLMAN, Michael F.; Danisco US, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304 (US).

(21) International Application Number:
PCT/US2007/020484

(22) International Filing Date:
21 September 2007 (21.09.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/846,656 22 September 2006 (22.09.2006) US
60/846,804 22 September 2006 (22.09.2006) US

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(71) Applicant (*for all designated States except US*):
DANISCO US, INC., GENENCOR DIVISION [US/US]; 925 Page Mill Road, Palo Alto, California 94304 (US).

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

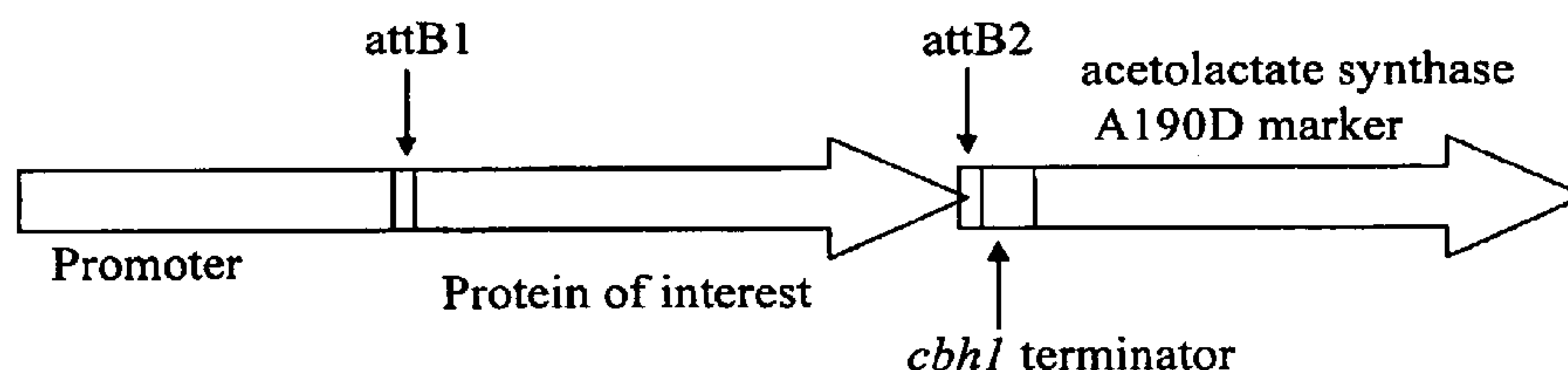
(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **BOWER, Benjamin S.** [US/US]; Danisco Us, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304 (US). **DUNN-COLEMAN, Nigel** [US/US]; Danisco Us, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304 (US). **LEIVA, Nicholas** [US/US]; Danisco Us, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304 (US).

Published:

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

(54) Title: ACETOLACTATE SYNTHASE (ALS) SELECTABLE MARKER FROM *TRICHODERMA REESEI*



(57) Abstract: A nucleic acid encoding an acetolactate synthase (ALS) protein that provides resistance to ALS inhibitors, e.g., sulphonylurea and imidazolinone compounds, is provided. The nucleic acid may be used as a selectable marker for expression of a protein of interest in host cells.

WO 2008/039370 A1

**ACETOLACTATE SYNTHASE (ALS) SELECTABLE MARKER FROM
*TRICHODERMA REESEI***

[01]

BACKGROUND

[02] Plants, fungi and bacteria synthesize the amino acids valine, leucine and isoleucine via a common pathway. One of the enzymes in this pathway is acetolactate synthase (ALS) (otherwise known as acetohydroxyacid synthase or AHAS), which converts pyruvate into 2-acetolactate as the first step in the synthesis of valine and leucine, and also converts pyruvate and 2-ketobutyrate into 2-aceto-2-hydroxybutyrate, the precursor of isoleucine. The activity of wild-type acetolactate synthase is sensitive to the action of several known classes of toxic compounds, including sulphonylurea and imidazolinone compounds. As such, such toxic compounds may be employed to kill cells containing acetolactate synthase proteins that are sensitive to those compounds.

[03] This disclosure relates to acetolactate synthase enzymes that provide resistance to toxic compounds, and hence are useful selectable markers for recombinant cells.

SUMMARY OF THE INVENTION

[04] An acetolactate synthase protein that provides resistance to ALS inhibitors, e.g., sulphonylurea and imidazolinone compounds, is provided, as well as a polynucleotide encoding the same. In certain embodiments, the amino acid sequence of the acetolactate protein is at least 80% identical to SEQ ID NO:1, the wild-type acetolactate protein of *Trichoderma reesei*. In particular embodiments, the ALS protein may contain an acidic amino acid at position 190, an acidic amino acid at position 241, or a histidine at position 372. In some embodiments, the ALS gene or polypeptide may be employed as a selectable marker in a wide variety of species. In certain cases, the protein may be non-naturally occurring.

[05] In certain embodiments, a polynucleotide encoding ALS may be operably linked to a promoter and a terminator to provide for expression of the ALS inhibitor resistance-conferring protein in a host cell. The promoter and terminator may be endogenous to the host cell in which

- 2 -

the polynucleotide is to be employed, and, in certain cases, the promoter and terminator may be the promoter and terminator of an ALS gene of the cell. The polynucleotide may, in certain embodiments, contain a single open reading frame encoding the acetolactate synthase protein, in which case the polynucleotide may be at least 70% identical to or may hybridize with SEQ ID NO:2. In other embodiments, the polynucleotide may comprise introns, in which case the polynucleotide may have a nucleotide sequence which is at least 70% identical to or may hybridize with SEQ ID NO:4. In other embodiments, the polynucleotide may be codon-optimized for expression of the acetolactate synthase protein in a particular host cell.

[06] A vector comprising the polynucleotide is also provided. In addition to the polynucleotide, the subject vector may contain an expression cassette for expression of a recombinant protein, e.g., an enzyme or therapeutic protein, in the cell.

[07] A host cell comprising the subject polynucleotide is also provided. In certain embodiments, the host cell is resistant to an ALS inhibitor, e.g., a toxic sulphonylurea or imidazolinone compound. The host cell may be any cell that is sensitive to the ALS inhibitor in the absence of the polynucleotide. In certain embodiments, the host cell may be a plant cell, e.g., a corn, soybean or *Arabidopsis* cell, a fungal cell, e.g., a filamentous fungal cell such as a *Trichoderma* sp. or *Aspergillus* sp. cell or a bacterial cell, e.g. a *Bacillus* sp. In particular embodiments, the host cell is a filamentous fungal. The cell may be present *in vitro*, or in a multicellular organism (e.g., a plant). The polynucleotide may be present in a genome of the host cell, or may be present in a vector that autonomously replicates in the host cell.

[08] Also provided is a method of selecting a cell. In certain embodiments, the method includes: introducing a subject polynucleotide which encodes an ALS into a plurality of cells, contacting the plurality of cells with an ALS inhibitor, and culturing the cells to provide for selection of the cell. In certain embodiments, the cell is a fungal cell. The cell may be cultured in a liquid medium containing the ALS inhibitor, or on a solid medium containing the ALS inhibitor. These methods may also include introducing a second polynucleotide into the host cell, where the second polynucleotide encodes a polypeptide that is to be produced by the host cell. The subject polynucleotide and the second polynucleotide may be present on the same or different nucleic acid, e.g., the same vector or different vectors. If different vectors are employed, they may be co-transformed into the same cells.

[09] Also provided are embodiments relating to relating to the 1818A and 1818B promoters

- 3 -

BRIEF DESCRIPTION OF THE FIGURES

[10] FIG. 1 shows the amino acid sequence of the wild type acetolactate synthase protein
5 from *T. reesei* (SEQ ID NO:1).

[11] FIG. 2 shows the nucleotide sequence of a *T. reesei* deduced cDNA (SEQ ID NO:2) encoding the wild type acetolactate protein of Fig. 1.

[12] FIG. 3 shows the nucleotide sequence of a *T. reesei* gene (SEQ ID NO:3) encoding the wild type acetolactate synthase protein of Fig. 1, wherein the promoter is in italics, the sequence
10 encoding ALS is underlined (SEQ ID NO:4), and the terminator sequence is in italics and underlined.

[13] FIG. 4 is a general diagram of a vector that may be used to test the ALS marker, wherein the vector includes two genes. The first gene corresponds to a polynucleotide encoding a protein of interest (e.g. a glucoamylase enzyme) and the second gene corresponds to a polynucleotide
15 coding for the ALS marker.

[14] FIGS. 5A and 5B show the nucleotide sequence of the promoters 1818A (SEQ ID NO:7) and 1818B (SEQ ID NO:8), respectively.

[15] FIG. 6 shows the nucleotide sequence (SEQ ID NO:9) of the pTrex-glucoamylase vector used to test the ALS marker which comprises the *Trichoderma reesei cbh1* promoter, attB1, a
20 polynucleotide encoding a *Trichoderma reesei* glucoamylase, attB2, the *Trichoderma reesei cbh1* terminator and an ALS marker (A190D) in the *E.coli* vector pSL1180.

[16] FIG. 7 illustrates an SDS-PAGE gel of supernatant samples from shake flask grown transformant. Lane 1 depicts the untransformed control. Expression of the *T. reesei* glucoamylase is shown in lanes 3 and 4 as highlighted by the arrow.

25 [17] FIG. 8 is a diagram of the vector used to test the 1818A and 1818B promoters.

[18] FIG. 9 shows the nucleotide sequence of a construct used to test the 1818A and 1818B promoters (SEQ ID NO:10).

DETAILED DESCRIPTION**Definitions**

30 [19] Unless defined otherwise herein, 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. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR

- 4 -

BIOLOGY, 2D ED., John Wiley and Sons, New York (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with general dictionaries of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described.

[20] All patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

[21] Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[22] The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

[23] As used herein, the term "selectable marker" refers to a gene or polynucleotide whose expression allows identification of cells that have been transformed with a DNA construct or vector containing the gene or polynucleotide. Selectable markers may provide resistance to toxic compounds such as antibiotics, herbicides, and the like.

[24] The term "acetolactate synthase (ALS)" refers to an enzyme that has an activity defined as EC 2.2.1.6, according to IUBMB Enzyme Nomenclature. The enzyme catalyses a reaction between two pyruvate molecules to produce 2-acetolactate and CO₂. The enzyme requires thiamine diphosphate, and may be referred to as acetohydroxyacid synthase (AHAS) elsewhere.

[25] The term "ALS inhibitor" refers to a compound that inhibits wild-type ALS protein and is toxic to cells that contain wild-type ALS. Such compounds include known herbicides, and include the sulfonylurea, imidazolinone, triazolopyrimidine, pyrimidinylthiobenzoate, or sulfonylamino-carbonyl-triazolinone compounds discussed below.

[26] The term "promoter" is defined herein as a nucleic acid that directs transcription of a downstream polynucleotide in a cell. In certain cases, the polynucleotide may contain a coding sequence and the promoter may direct the transcription of the coding sequence into translatable RNA.

[27] The term "isolated" as defined herein means a compound, a protein, cell, nucleic acid sequence or amino acid that is removed from at least one component with which it is naturally associated.

[28] The term "coding sequence" is defined herein as a nucleic acid that, when placed under

- 5 -

the control of appropriate control sequences including a promoter, is transcribed into mRNA which can be translated into a polypeptide. A coding sequence may contain a single open reading frame, or several open reading frames separated by introns, for example. A coding sequence may be cDNA, genomic DNA, synthetic DNA or recombinant DNA, for example. A coding sequence generally starts at a start codon (e.g., ATG) and ends at a stop codon (e.g., UAA, UAG and UGA).

[29] The term "recombinant" refers to a polynucleotide or polypeptide that does not naturally occur in a host cell. A recombinant molecule may contain two or more naturally occurring sequences that are linked together in a way that does not occur naturally.

[30] The term "heterologous" refers to elements that are not normally associated with each other. For example, a heterologous protein is a protein that is not produced in a wild-type host cell, a heterologous promoter is a promoter that is not present in nucleic acid that is endogenous to a wild type host cell, and a promoter operably linked to a heterologous coding sequence is a promoter that is operably linked to a coding sequence that it is not usually operably linked to in a wild-type host cell.

[31] The term "operably linked" refers to an arrangement of elements that allow them to be functionally related. For example, a promoter is operably linked to a coding sequence if it controls the transcription of the sequence, and a signal sequence is operably linked to a protein if the signal sequence directs the protein through the secretion system of a host cell.

[32] The term "nucleic acid" encompasses DNA, RNA, single or doubled stranded and modification thereof. The terms "nucleic acid" and "polynucleotide" may be used interchangeably herein.

[33] The term "DNA construct" as used herein means a nucleic acid sequence that comprises at least two DNA polynucleotide fragments.

[34] The term "signal sequence" or "signal peptide" refers to a sequence of amino acids at the N-terminal portion of a protein, which facilitates the secretion of the mature form of the protein outside the cell. The mature form of the extracellular protein lacks the signal sequence which is cleaved off during the secretion process.

[35] The term "vector" is defined herein as a polynucleotide designed to carry nucleic acid sequences to be introduced into one or more cell types. Vectors include cloning vectors, expression vectors, shuttle vectors, plasmids, phage or virus particles, DNA constructs, cassettes and the like. Expression vectors may include regulatory sequences such as promoters, signal sequences, a coding sequences and transcription terminators.

[36] An "expression vector" as used herein means a DNA construct comprising a coding

- 6 -

sequence that is operably linked to suitable control sequences capable of effecting expression of a protein in a suitable host. Such control sequences may include a promoter to effect transcription, an optional operator sequence to control transcription, a sequence encoding suitable ribosome binding sites on the mRNA, enhancers and sequences which control termination of transcription and translation.

[37] As used herein, the terms "polypeptide" and "protein" are used interchangeably and include reference to a polymer of any number of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analog of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The terms also apply to polymers containing conservative amino acid substitutions such that the polypeptide remains functional. "Peptides" are polypeptides having less than 50 amino acid residues.

[38] A "host cell" is a cell which that contains a subject recombinant nucleic acid, either in the genome of the host cell or in an extrachromosomal vector that replicates autonomously from the genome of the host cell. A host cell may be any cell type.

[39] The term "filamentous fungi" refers to all filamentous forms of the subdivision Eumycotina (See, Alexopoulos, C. J. (1962), INTRODUCTORY MYCOLOGY, Wiley, New York). These fungi are characterized by a vegetative mycelium with a cell wall composed of chitin, glucans, and other complex polysaccharides. The filamentous fungi of the present invention are morphologically, physiologically, and genetically distinct from yeasts. Vegetative growth by filamentous fungi is by hyphal elongation and carbon catabolism is obligatory aerobic.

[40] A "non-pathogenic" cell is a strain that is not pathogenic to humans.

[41] "Transformation" means introducing DNA into a cell so that the DNA is maintained in the cell either as an extrachromosomal element or chromosomal integrant.

[42] Unless otherwise indicated, all amino acid positions in an acetolactate synthase protein are relative to SEQ ID NO:1, after alignment of that protein with SEQ ID NO:1 using the BLASTP program (Altschul, Nucl. Acids Res. 1997 25:3389-3402; Schäffer, Bioinformatics 1999 15:1000-1011) under default conditions, as available from the world wide website of the National Center of Biotechnology Information (NCBI).

Polynucleotides

[43] Provided herein is a polynucleotide that encodes an acetolactate synthase protein that

- 7 -

provides resistance to ALS inhibitors. In certain embodiments, the acetolactate synthase protein is a non-naturally occurring protein. In general terms, the polynucleotide encodes a protein that: a) possesses acetolactate synthase activity (i.e., can catalyze a reaction between two pyruvate molecules to produce 2-acetolactate), b) confers resistance to ALS inhibitors, e.g.,

5 sulphonylurea and imidazolinone compounds, and c) has an amino acid sequence that is at least 85% identical (e.g., at least 90% identical, at least 93% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 98.5% identical, at least 99% identical or at least 99.5% identical) to SEQ ID NO:1, where SEQ ID NO:1 sets forth the amino acid sequence of a wild-type acetolactate synthase from *Trichoderma reesei*.

10 [44] In certain embodiments, the polynucleotide may encode a protein having one or more of: a) an acidic amino acid (e.g., Asp or Glu) at position 190, b) an acidic amino acid (e.g., Asp or Glu) at position 241, or a His at position 372. Positions 190, 241 and 372 in the wild-type *T. reesei* ALS protein are underlined in Fig. 1. As noted above, positions 190, 241 and 372 in a different ALS protein (e.g., an ALS protein that is, for example, shorter, longer or contains
15 deletions and/or insertions relative to the wild-type *T. reesei* ALS protein), are defined herein as being the positions in that protein that correspond to (i.e., align with or lie across from) positions 190, 241 and 372 of the wild-type *T. reesei* ALS protein when the wild-type *T. reesei* ALS protein and the other protein are aligned using standard sequence alignment methods, e.g., BLASTP (Altschul, Nucl. Acids Res. 1997 25:3389-3402; Schäffer, Bioinformatics 1999
20 15:1000-1011) using default parameters. ALS proteins, in general, are well characterized enzymes and have been investigated in great detail functionally and structurally. ALS proteins have been reviewed in several publications (see, e.g., Chipman, Biochim. Biophys. Acta 1998 1385:401-19; Chipman, Curr. Opin. Chem. Biol. 2005 9:475-81), and have been crystallized (see, e.g., Pang, J. Biol. Chem. 2004 279:2242-53; Pang, J. Mol. Biol. 2002 317:249-62), as
25 well as subjected to mutagenesis to identify essential and non-essential residues (see, e.g., Ibdah, Biochemistry 1996 35:16282-91; Mendel J. Mol. Biol. 2001 307:465-771; Hill, Biochem. J. 1998 335:653-61). Other herbicide resistance-conferring mutations are also known (see, e.g., Jung, Biochem. J. 2004 383:53-61; Duggelby, Eur. J. Biochem. 2003 270:2895-904). Further, the amino acid sequences of several hundred ALS proteins are known, and publicly
30 available via NCBI's Genbank database. As such, a wide variety of amino acid changes that could be made to a subject ALS protein, some of which may confer resistance to ALS inhibitors, without abolishing its activity would be readily apparent.

[45] In certain cases, the amino acid changes described herein may be transferred into any other ALS protein, from any species, to render that protein herbicide resistant. In other words,

- 8 -

an amino acid at position 190, 241 or 372 (relative to SEQ ID NO:1) of any other ALS protein may be substituted by an acidic amino acid (e.g., Asp or Glu), an acidic amino acid (e.g., Asp or Glu), or a His, respectively, to provide an ALS protein that confers resistant to an ALS inhibitor.

5 [46] For example, the amino acid sequence of the ALS proteins of several fungal species are known and are deposited into NCBI's Genbank database. In certain embodiments, the above-described amino acid alterations can be transferred to the ALS proteins of those fungi in order to provide other ALS inhibitor-resistant proteins. In other embodiments, the amino acid sequence other fungal ALS proteins may be employed to make further changes in the subject *T.*
 10 *reesei*-based ALS proteins that do not abolish the ALS activity of those proteins. For example, a fusion between two ALS proteins from different species, or a protein containing amino acid substitutions, deletions or insertions could be made. Exemplary ALS amino acid sequences from other fungal species, including other filamentous fungal species, are deposited at NCBI's Genbank database as GIDS: 39977967 and 2547090 (*Magnaporthe grisea*), GID:85108881
 15 (*Neurospora crassa*), GID:46108408 (*Gibberella zeae*), GID:90302929 (*Coccidioides immitis*), GID: 67537572 (*Aspergillus nidulans*); GID: 70999742 (*Aspergillus fumigatus*); GIDs: 83767597 and 83771596 (*Aspergillus oryzae*); GID: 111063308 (*Phaeosphaeria nodorum*), GID: 50547615 (*Yarrowia lipolytica*), GID: 49657303 (*Debaryomyces hansenii*), GID: 68468265 (*Candida albicans*), GID: 21615550 (*Saccharomycopsis fibuligera*), GID: 49641223
 20 (*Kluyveromyces lactis*); GID: 49527687 (*Candida glabrata*) and GID: 817866 (*Saccharomyces cerevisiae*).

[47] Because of the redundancy of the genetic code, a subject polynucleotide may comprise
 25 any one of a number of nucleotide sequences. In particular embodiments, the subject polynucleotide may have a nucleotide sequence that is: a) at least 70% identical to (e.g., at least 80%, at least 90%, at least 95%, at least 98% or at least 99% identical to), or b) hybridizes with under stringent hybridization conditions to SEQ ID NO:2 or SEQ ID NO:4, which sequences set forth the nucleotide sequence of a wild-type cDNA and gene of *T. reesei*, respectively. As
 30 such, the encoding polypeptide may contain introns, or may contain a single open reading frame encoding the protein.

[48] As would be apparent, in certain embodiments, the polynucleotide may have a nucleotide sequence that encodes a protein having one or more of: a) an acidic amino acid (e.g., Asp or

- 9 -

Glu) at position 190, b) an acidic amino acid (e.g., Asp or Glu) at position 241, or a His at position 372.

[49] The term "identity" in the context of two nucleic acid sequences refers to nucleotides residues in the two sequences that are the same when aligned for maximum correspondence, as measured using any of the following sequence comparison algorithms. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection.

[50] An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul, *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information available on the world wide web ([www](http://www.ncbi.nlm.nih.gov)) [ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). The BLAST algorithm performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). A nucleotide sequence of at least 100, at least 500, at least 1000, up to entire length of the SEQ ID NOS:2 or 4 may be employed in sequence comparisons.

[51] As noted above, the polynucleotide may include a nucleotide sequence that hybridizes under stringent hybridization conditions to a polynucleotide having the nucleotide sequence of SEQ ID NO: 2 or 4, where stringent hybridization conditions encompass low, medium, high and very high stringency hybridization conditions.

[52] "Low-stringency" conditions refer to washing with a solution of 1X SSC/0.1% SDS at 20° C for 15 minutes. "Medium-stringency" conditions refer to washing with a solution of 1X SSC/0.1% SDS at 65° C for 60 minutes. "High-stringency" conditions refer to washing with a solution of 0.2X SSC/0.1% SDS at 65° C for 10 minutes. "Very high-stringency" conditions refer to washing with a solution of 0.2X SSC/0.1% SDS at 65° C for 60 minutes.

[53] Hybridization methods are described in great detail in Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL* (2nd Ed., 1989 Cold Spring Harbor, NY). In one exemplary hybridization assay, a DNA sample is electrophoresed through an agarose gel (for example, 0.8% agarose) so that of the DNA fragment can be visualized by ethidium bromide staining. The gel is then briefly rinsed in distilled H₂O and subsequently depurinated in an appropriate

- 10 -

solution (such as, for example, 0.25M HCl) with gentle shaking followed by denaturation for 30 minutes (in, for example, 0.4 M NaOH) with gentle shaking. A renaturation step may be included, in which the gel is placed in 1.5 M NaCl, 1MTris, pH 7.0 with gentle shaking for 30 minutes. The DNA is then transferred onto an appropriate positively charged membrane, for example, *Maximum Strength Nytran Plus* membrane (Schleicher & Schuell, Keene, N.H.), using a transfer solution (such as, for example, 6XSSC, i.e., 900 mM NaCl, 90 mM trisodium citrate). Once the transfer is complete, generally after about 2 hours, the membrane is rinsed in e.g., 2X SSC (300 mM NaCl, 30 mM trisodium citrate) and air dried at room temperature. The membrane may be prehybridized (for approximately 2 hours or more) in a suitable prehybridization solution (such as, for example, an aqueous solution containing per 100 mL: 20-50 mL formamide, 25 mL of 20X SSPE (1X SSPE = 0.18 M NaCl, 1 mM EDTA, 10 mM NaH₂PO₄, pH 7.7), 2.5 mL of 20% SDS, and 1 mL of 10 mg/mL sheared herring sperm DNA). As would be known to one of skill in the art, the amount of formamide in the prehybridization solution may be varied depending on the nature of the reaction obtained according to routine methods. Thus, a lower amount of formamide may result in more complete hybridization in terms of identifying hybridizing molecules than the same procedure using a larger amount of formamide. On the other hand, a strong hybridization band may be more easily visually identified by using more formamide.

[54] A DNA probe of at least 100, at least 500, at least 1000, up to entire length of the SEQ ID NOS:2 or 4 may be employed in hybridization assays or in sequence comparisons. The DNA probe may be isolated by electrophoresis in an agarose gel, the fragment excised from the gel, and recovered from the excised agarose. This purified fragment of DNA may be labeled (using, for example, the *Megaprime* labeling system according to the instructions of the manufacturer) to incorporate P³² in the DNA. The labeled probe is denatured by heating to 95°C for 5 minutes and immediately added to the membrane and prehybridization solution. The hybridization reaction should proceed for an appropriate time and under appropriate conditions, for example, for 18 hours at 37°C with gentle shaking or rotating. The membrane is rinsed (for example, in 2X SSC/0.3% SDS) and then washed in an appropriate wash solution, as described above, with gentle agitation. Hybridization can be detected by autoradiography.

[55] In particular embodiments, the polynucleotide may be codon optimized for expression in particular host cell. In other embodiments, the polynucleotide may have a nucleotide sequence that contains less than 10 (e.g., 9, 8, 7, 6, 5, 4, 3, 2, or 1) differences compared to the wild type ALS sequence (e.g., either the genomic or cDNA sequence) of the particular host cell in which it the polynucleotide is to be employed.

- 11 -

[56] In certain cases, in addition to a coding sequence, the polynucleotide may further contain other elements that are necessary for expression of the encoded protein in a host cell. For example, in one embodiment, the polynucleotide may be flanked by sequences to form an expression cassette that provides expression of the encoded protein in a host cell. In certain
5 embodiments, the expression cassette may contain a promoter for transcription of the coding sequence, and a transcriptional terminator, and a sequence encoding a 5' untranslated region (UTR) that allows translational initiation, each in operable linkage to the coding sequence. Promoters, enhancers, terminators, UTRs, polyadenylation signals and other regulatory sequences for a wide variety of host cells, particularly those from plants, bacteria and fungi, are
10 well known in the art (see, e.g., Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995; Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, 1989 Cold Spring Harbor, N.Y.).

[57] In particular embodiments, the subject expression cassette may contain a promoter and terminator for expression of the subject protein in a filamentous fungal cell. Examples of
15 suitable promoters and terminators for directing the transcription of a subject nucleic acid in a filamentous fungal host cell are promoters and terminators obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae*
20 alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, *Fusarium venenatum* amyloglucosidase, *Fusarium oxysporum* trypsin-like protease (WO 96/00787), *Trichoderma reesei* cellobiohydrolase I, *Trichoderma reesei* cellobiohydrolase II, *Trichoderma reesei* endoglucanase I, *Trichoderma reesei* endoglucanase II, *Trichoderma reesei* endoglucanase III, *Trichoderma reesei* endoglucanase IV, *Trichoderma*
25 *reesei* endoglucanase V, *Trichoderma reesei* xylanase I, *Trichoderma reesei* xylanase II, *Trichoderma reesei* beta-xylosidase, as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase); and mutant, truncated, and hybrid promoters thereof.

[58] The promoter and/or terminator may be native or non-endogenous to the host cell in
30 which the subject ALS protein is to be expressed and, in certain embodiments, the promoter and terminator may be the ALS gene promoter and terminator from the host cell. For example, in one embodiment, an expression cassette for use in *Trichoderma reesei* cells may comprise the *Trichoderma reesei* ALS gene promoter, the *Trichoderma reesei* ALS coding sequence (containing one or more of the nucleotide changes outlined above) and *Trichoderma reesei*

- 12 -

ALS terminator, in operable linkage (Fig. 3). The promoter and/or terminator may be from an ALS gene of the host cell in which the polynucleotide is to be employed.

[59] The polynucleotide may be integrated into a genome of the host cell, or may be present on a vector that autonomously replicates in the host cell.

5

Recombinant nucleic acid

[60] A recombinant nucleic acid comprising a subject polynucleotide is also provided. A subject recombinant nucleic acid may comprise a subject polynucleotide, e.g., an expression cassette for production of a resistance-conferring ALS protein in a host cell, as well as a second expression cassette for expression of a protein of interest in the host cell. In a particular embodiment, the subject polynucleotide is employed as a marker for selection of host cells that contain the recombinant nucleic acid over other cells that do not contain the recombinant nucleic acid (i.e., the subject polynucleotide is employed as a "selectable marker" for cells that contain the subject polynucleotide).

15

[61] The protein of interest encoded by the second expression cassette may be for example an enzyme, a therapeutic protein, a reporter protein, a food additive or a foodstuff or the like.

[62] In one embodiment, the protein of interest encoded by the second expression cassette may be an enzyme such as a carbohydrase, such as an α -amylase, an alkaline α -amylase, a β -amylase, a cellulase; a dextranase, an α -glucosidase, an α -galactosidase, a glucoamylase, a hemicellulase, a pentosanase, a xylanase, an invertase, a lactase, a naringanase, a pectinase or a pullulanase; a protease such as an acid protease, an alkali protease, bromelain, ficin, a neutral protease, papain, pepsin, a peptidase, rennet, rennin, chymosin, subtilisin, thermolysin, an aspartic proteinase, or trypsin; a lipase or esterase, such as a triglyceridase, a phospholipase, a pregastric esterase, a phosphatase, a phytase, an amidase, an iminoacylase, a glutaminase, a lysozyme, or a penicillin acylase; an isomerase such as glucose isomerase; an oxidoreductases, e.g., an amino acid oxidase, a catalase, a chloroperoxidase, a glucose oxidase, a hydroxysteroid dehydrogenase or a peroxidase; a lyase such as a acetolactate decarboxylase, an aspartic β -decarboxylase, a fumarese or a histadase; a transferase such as cyclodextrin glycosyltransferase; or a ligase, for example. In particular embodiments, the protein may be an aminopeptidase, a carboxypeptidase, a chitinase, a cutinase, a deoxyribonuclease, an α -galactosidase, a β -galactosidase, a β -glucosidase, a laccase, a mannosidase, a mutanase, a pectinolytic enzyme, a polyphenoloxidase, ribonuclease or transglutaminase, for example. The enzyme may be a wild-type enzyme or a variant of a wild-type enzyme. In addition the enzyme may be a hybrid

30

- 13 -

enzyme that includes fragments of different enzymes.

[63] In other embodiment, the protein of interest encoded by the second expression cassette may be a therapeutic protein (i.e., a protein having a therapeutic biological activity). Examples of suitable therapeutic proteins include: erythropoietin, cytokines such as interferon- α ,
5 interferon- β , interferon- γ , interferon- δ , and granulocyte-CSF, GM-CSF, coagulation factors such as factor VIII, factor IX, and human protein C, antithrombin III, thrombin, soluble IgE receptor α -chain, IgG, IgG fragments, IgG fusions, IgM, IgA, interleukins, urokinase, chymase, and urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, annexin V fusion protein, angiostatin, vascular endothelial growth factor-2,
10 myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1-antitrypsin, α -feto proteins, DNase II, kringle 3 of human plasminogen, glucocerebrosidase, TNF binding protein 1, follicle stimulating hormone, cytotoxic T lymphocyte associated antigen 4-Ig, transmembrane activator and calcium modulator and cyclophilin ligand, soluble TNF receptor Fc fusion, glucagon like protein 1 and IL-2 receptor agonist. Antibody proteins, e.g., monoclonal antibodies that may be
15 humanized, are of particular interest.

[64] In a further embodiment, the protein encoded by the second expression cassette may be a reporter protein. Such reporter proteins may be optically detectable or colorigenic, for example. In this embodiment, the protein may be a β -galactosidase (*lacZ*), β -glucuronidase (GUS), luciferase, alkaline phosphatase, nopaline synthase (NOS), chloramphenicol acetyltransferase
20 (CAT), horseradish peroxidase (HRP) or a fluorescent protein green, e.g., green fluorescent protein (GFP), or a derivative thereof.

[65] In certain embodiments, particularly those in which the host cell is a filamentous fungal host cell, the coding sequence of the second expression cassette may encode a fusion protein. In some of these embodiments, the fusion protein may provide for secretion of the protein from
25 the host cell in which it is expressed and, as such, may contain a signal sequence operably linked to the N-terminus of the protein, where the signal sequence contains a sequence of amino acids that directs the protein to the secretory system of the host cell, resulting in secretion of the protein from the host cell into the medium in which the host cell is growing. The signal sequence is cleaved from the fusion protein prior to secretion of the protein. The signal
30 sequence employed may be endogenous or non-endogenous to the host cell and, in certain embodiments, may be signal sequence of a protein that is known to be highly secreted from a host cell. In particular embodiments, the signal sequence protein may be any signal sequence that facilitates protein secretion from a filamentous fungal (e.g., *Trichoderma* or *Aspergillus*) host cell. Such signal sequence include, but are not limited to: the signal sequence of

- 14 -

cellobiohydrolase I, cellobiohydrolase II, endoglucanases I, endoglucanases II, endoglucanases III, α -amylase, aspartyl proteases, glucoamylase, mannanase, glycosidase and barley endopeptidase B (see Saarelainen, Appl. Environ. Microbiol. 1997 63: 4938-4940), for example. Other of signal sequences are those originating from the fungal amyloglucosidase (AG) gene (*glaA*), the α factor gene (yeasts e.g. *Saccharomyces*, *Kluyveromyces* and *Hansenula*) or the α amylase gene (*Bacillus*). In certain embodiments, therefore, the subject recombinant nucleic acid may comprise: a signal sequence-encoding nucleic acid operably linked to a protein-encoding nucleic acid, where translation of the nucleic acid in a host cell produces a fusion protein comprising a protein having an N-terminal signal sequence for secretion of the protein from the host cell.

[66] In particular embodiments, the fusion protein may further contain a "carrier protein", which is a portion of a protein that is endogenous to and highly secreted by the host cell. Suitable carrier proteins include those of *T. reesei* mannanase I (Man5A, or MANI), *T. reesei* cellobiohydrolase II (Cel6A, or CBHII) (see, e.g., Paloheimo et al Appl. Environ. Microbiol. 2003 December; 69(12): 7073-7082) or *T. reesei* cellobiohydrolase I (CBHI). In one embodiment, the carrier protein is a truncated *T. reesei* CBH1 protein that includes the CBH1 core region and part of the CBH1 linker region. A fusion protein containing, from amino-terminus to carboxy-terminus, a signal sequence, a carrier protein and a subject protein in operable linkage is therefore provided, as well as a nucleic acid encoding the same.

[67] In certain embodiments, the polynucleotide may be codon optimized for expression of the protein in a particular host cell. Since codon usage tables listing the usage of each codon in many cells are known in the art (see, e.g., Nakamura et al, Nucl. Acids Res. 2000 28: 292) or readily derivable, such nucleic acids can be readily designed giving the amino acid sequence of a protein to be expressed.

[68] A subject recombinant nucleic acid may be present, e.g., integrated, into a genome (i.e., the nuclear genome) of a host cell, or may be present in a vector, e.g., a phage, plasmid, viral, or retroviral vector, that autonomously replicates in the host cell. In certain embodiments, the vector may be an expression vector for expressing a protein in a host cell and, as such, may further contain the second expression cassette discussed above. In certain embodiments, the vector may be an expression vector for expressing a recombinant polypeptide in a filamentous fungal cell.

[69] Vectors for expression of recombinant proteins are well known in the art (Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995; Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

- 15 -

Selection methods

[70] A method for selecting host cells that contain a subject polynucleotide is also provided. In certain embodiments, the method includes introducing a subject polynucleotide into a plurality of cells, contacting the plurality of cells with an ALS inhibitor, e.g., a toxic sulphonylurea or imidazolinone compound, and culturing the cells to provide for selection of the cell. The cell may be selected on solid medium, e.g., by plating the plurality of cells onto an agar plate containing the ALS inhibitor, or in liquid medium, e.g., by culturing the plurality of in a liquid medium containing the ALS inhibitor.

[71] In certain embodiments, the subject selection methods may be employed to select for cells containing a second expression cassette that encodes a protein of interest. The second expression cassette may be present in a recombinant nucleic acid that also contains the instant ALS protein-encoding polynucleotide (i.e., in a single recombinant nucleic acid molecule). In these embodiments, the instant selection methods may be employed to select for cells that contain the recombinant nucleic acid. Since the recombinant nucleic acid also contains the second expression cassette, cells that contain the second expression cassette are selected using the ALS inhibitor. In alternative embodiments, the second expression cassette may be present on a recombinant nucleic acid that does not contain the instant ALS protein-encoding polynucleotide (e.g., a different vector). In these embodiments, a subject polynucleotide may be co-transformed with (i.e., transformed at the same time as) a separate and distinct polynucleotide molecule (e.g., a different nucleic acid molecule or vector) that contains the expression cassette. As such, the instant selection methods may be employed to select for cells that contain the second expression cassette, even though the second expression cassette is on a different nucleic acid molecule to the polynucleotide.

[72] As such, the subject selection methods may be employed to select for host cells that express the protein of interest. The protein of interest may be native or non-native to the host cells used.

[73] The exact concentration of ALS inhibitor employed may vary according to the particular ALS inhibitor used and the type of host cell to be selected. In general terms, the ALS inhibitor is used at a concentration that provides selection of the host cell containing the polynucleotide. In certain embodiments, the ALS inhibitor may be employed at a concentration of 0.5 ppm to 10,000 ppm, for example 1 ppm to 10,000 ppm or 10 ppm to 1,000 ppm. For example, in certain cases, the ALS inhibitor may be employed at a concentration in the range of 25 ppm to

- 16 -

100 ppm, e.g., 50 ppm or 100 ppm, or at a concentration in the range of 100 ppm to 500 ppm, e.g., 200 ppm or 500 ppm. For example, in one embodiment, the ALS inhibitor may be employed at a concentration in the range of 1 µg/ml to 1 mg/ml, e.g., 10 µg/ml to 500 µg/ml.

[74] ALS inhibitors include any compounds that: a) kill cells that do not have resistance to the compounds via inhibiting an ALS enzyme and b) do not kill cells having a subject polynucleotide. ALS inhibitors of particular interest include sulfonylurea (SU), imidazolinone (IMI), triazolopyrimidine (TP), pyrimidinylthiobenzoate (PTB), and sulfonylamino-carbonyl-triazolinone (SCT) compounds that are known ALS inhibitors and, in certain cases, may be commonly employed has herbicides. Examples of sulfonylurea compounds that may be employed in the subject methods include: I) phenylsulfonylureas, including a) chlorimuron ethyl (see Agricultural Chemicals Book II "Herbicides" by W. T. Thompson, Thompson Publications, Fresno Calif., U.S.A. 1990, page 152); b) primisulfuron (CGA 136,872, see Brighton Crop Prot. Conf. "Weeds" 1989, p. 41-48), c) 3-(4-ethyl-6-methoxy-1,3,5-triazin-2-yl)-1-(2,3-dihydro-1,1-dioxo-2-methyl benzo[b]thiophen-7-sulfonyl)-urea (see, e.g., EP-A-79,683), d) 3-(4-ethoxy-6-ethyl-1,3,5-triazin-2-yl)-1-(2,3-dihydro-1,1-dioxo-2-methylbenzo[b]thiophen-7-sulfonyl)urea (see, e.g., EP-A-79,683), e) tribenuron-methyl (see "The Pesticide Manual", British Crop Protection Council 9th Edition (1990/91), page 840), f) metsulfuron-methyl (see Proc. Int. Congr. Plant Prot., 10th, 1983, Vol. 1, 324), g) chlorsulfuron (see U.S. Pat. No. 4,127,405; Weeds Weed Control, 1980, 21st, 24), h) triasulfuron (see "The Pesticide Manual" 9th Ed., p. 837) and i) sulfometuron-methyl (see "The Pesticide Manual" 9th Ed., p. 774); II) thienylsulfonylureas, for example thifensulfuron-methyl (see Agricultural Chemicals Book II "Herbicides" by W. T. Thompson, Thompson Publications, Fresno Calif., U.S.A. 1990, page 155); III) pyrazolylsulfonylureas, for example: a) pyrazosulfuron-ethyl (NC 311, see "The Pesticide Manual" 9th Ed., p. 735) and b) methyl 3-chloro-5-(4,6-dimethoxypyrimidin-2-ylcarbamoylsulfamoyl)-1-methyl-pyrazole-4-carboxylate (see EP 282,613); IV) Sulfonediarnide derivatives, for example amidosulfuron and structural analogs (see EP-A-0,131,258 and Z. Pfl.Krankh. Pfl.Schutz, Special Issue XII, 489-497 (1990); V) pyridylsulfonylureas, for example: a) nicosulfuron (SL-950, see Kimura et al., Brighton Crop Protection Conference "Weeds" 1989, p. 29-34); b) DPX-E 9636 (see Brighton Crop Prot. Conf.--Weeds--1989, p. 23 et seq.) and c) pyridylsulfonylureas as are described in German Patent Applications P 4000503.8 (WO-91/10660) and P 4030577.5 and VI) Phenoxy sulfonylureas such as those described in, for example, EP-A-0,342,569, EP-A-4,163, EP-A-113,956, U.S. Pat. No. 4,678,500 and U.S. Pat. No. 4,581,059. Examples of imidazolinone compounds that may be employed in the subject methods include: a)

- 17 -

imazethapyr (see Ch. R. Worthing's "The Pesticide Manual" 8th Edition 1987, by British Crop Protection Council, page 473), b) imazaquin (see Ch. R. Worthing's "The Pesticide Manual" 8th Edition 1987, by British Crop Protection Council, page 474), and c) imazethamethapyr (chemical name: rac-2-[4,5-dihydro-4-methyl-4-(1-methylethyl)-5-oxo-1H-imidazol-2-yl]-5-methyl-3-pyridine-carboxylic acid; see Weed Techn. 1991 (5), 430-433 and 434-438) and other related compounds.

Host cells

[75] A host cell comprising a subject recombinant nucleic acid is also provided. The host cell may be any cell type, e.g., bacterial (such as *E. coli*, *Bacillus sp.* or *Streptomyces sp.*), fungal (such as a non-filamentous or filamentous fungal), or plant (such as an *Arabidopsis*, corn or soybean plant) host cells. In some embodiments, the host cell may be a cell of a species that has a history of use for production of proteins that has GRAS status, i.e., a Generally Recognized as Safe, by the FDA.

[76] In particular embodiments, the subject host cell may be a fungal cell of the following species: *Trichoderma*, (e.g., *Trichoderma reesei* (previously classified as *T. longibrachiatum* and currently also known as *Hypocrea jecorina*), *Trichoderma viride*, *Trichoderma koningii*, and *Trichoderma harzianum*)); *Penicillium sp.*, *Humicola sp.* (e.g., *Humicola insolens* and *Humicola grisea*); *Chrysosporium sp.* (e.g., *C. lucknowense*), *Gliocladium sp.*, *Aspergillus sp.* (e.g., *Aspergillus oryzae*, *Aspergillus niger*, *Aspergillus nidulans*, *Aspergillus kawachi*, *Aspergillus aculeatus*, *Aspergillus japonicus*, *Aspergillus sojae*, and *Aspergillus awamori*), *Fusarium sp.*, *Mucor sp.*, *Neurospora sp.*, *Hypocrea sp.*, or *Emericella sp.* (See also, Innis et al., (1985) Sci. 228:21-26), among others.

[77] Exemplary bacterial host cells include *Bacillus sp.*, including, but not limited to *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. clausii*, *B. halodurans*, *B. megaterium*, *B. coagulans*, *B. circulans*, *B. lautus*, and *B. thuringiensis*, and *Streptomyces sp.*, including, but not limited to: *S. lividans*, *S. carbophilus* and *S. helveticus*.

[78] Exemplary plant host cells include monocot and dicot plant cells, including, but not limited to corn (*Zea mays*), *Brassica sp.*, rice (*Oryza sativa*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*) and *Arabidopsis thaliana* and tomato (*Lycopersicon esculentum*) host cells. A host cell may be a host cell cultured in vitro, or a host

- 18 -

cell of a multicellular organism, i.e., a plant. Methods of transferring exogenous nucleic acids into such host cells are well known in the art.

[79] In particular embodiments, a subject fungal cell may be a strain of *Trichoderma* and particularly *T. reesei* which include functional equivalents of RL-P37 (Sheir-Neiss et al. (1984) Appl. Microbiol. Biotechnology 20:46 –53). Other useful host strains include; NRRL 15709, ATCC 13631, ATCC 26921 (QM 9414) ATCC 32098, ATCC 32086, and ATCC 56765 (RUT-30). In other embodiments, subject fungal host cell may be of a strain of *Aspergillus* sp., including ATCC 22342, ATCC 44733, ATCC 14331, ATCC 11490, NRRL 3112, and strains derived therefrom.

[80] In some embodiments, a host cell may be one wherein native genes have been deleted or inactivated. For example genes corresponding to protease genes (e.g. aspartyl protease) (Berka et al. (1990) Gene 86:153-162 and USP 6,509,171) or genes corresponding to cellulase genes may be deleted or inactivated, (e.g. *cbh1*, *cbh2* and *egl1*, and *egl2*) such as the quad deleted strain of *T. reesei* disclosed in WO 05/001036 and derivatives thereof.

[81] Introduction of a nucleic acid into a host cell includes techniques such as transformation; electroporation; nuclear microinjection; transduction; transfection, (e.g., lipofection mediated and DEAE-Dextrin mediated transfection); incubation with calcium phosphate DNA precipitate; high velocity bombardment with DNA-coated microprojectiles; and protoplast fusion. General transformation techniques are known in the art (See, e.g., Ausubel et al., (1987), *supra*, chapter 9; and Sambrook (1989) *supra*, and Campbell *et al.*, (1989) *Curr. Genet.* 16:53-56). Reference is also made to WO 05/001036; USP 6,022,725; USP 6,103,490; USP 6,268,328; and published U.S. patent applications 20060041113, 20060040353, 20060040353 and 20050208623.

[82] Transformation and protein expression in *Aspergillus* and *Trichoderma* is further described in, for example U.S. Pat. No. 5,364,770; U.S. Pat. No. 6,022,725; and Nevalainen et al., 1992, The Molecular Biology of *Trichoderma* and its Application to the Expression of Both Homologous and Heterologous Genes, in MOLECULAR INDUSTRIAL MYCOLOGY, Eds. Leon and Berka, Marcel Dekker, Inc. pp. 129 – 148.

[83] As noted above, a subject host cell may further contain a recombinant nucleic acid for expression of a protein of interest in the host cell, in addition to the ALS-encoding nucleic acid. The subject recombinant nucleic acid and the ALS-encoding nucleic acid may be closely linked in *cis*, either in the genome or in a plasmid, such that the ALS inhibitor selects for the recombinant nucleic, and thereby selects for cells that produce the protein.

- 19 -

Protein production

[84] Methods of using the above-described host cell are also provided. In certain embodiments, the subject methods include: culturing the cell comprising a recombinant nucleic acid comprising a first expression cassette for producing a subject ALS enzyme, and a second expression cassette for producing a protein, to produce the protein. In certain embodiments and as discussed above, the protein may be secreted into the culture medium. As such, certain embodiments of the method include the step of recovering the protein from the culture medium.

[85] Cells may be cultured in a standard medium containing physiological salts and nutrients (See, e.g., Pourquie, J. *et al.*, BIOCHEMISTRY AND GENETICS OF CELLULOSE DEGRADATION, eds. Aubert, J. P. *et al.*, Academic Press, pp. 71-86, 1988 and Ilmen, M. *et al.*, (1997) *Appl. Environ. Microbiol.* 63:1298-1306). Common commercially prepared media (e.g., Yeast Malt Extract (YM) broth, Luria Bertani (LB) broth and Sabouraud Dextrose (SD) broth) also find use in the present invention. Preferred culture conditions for a given filamentous fungus are known in the art and may be found in the scientific literature and/or from the source of the fungi such as the American Type Culture Collection (ATCC) and Fungal Genetics Stock Center.

[86] In some embodiments, a subject host cell may be cultured under batch or continuous fermentation conditions. A classical batch fermentation is a closed system, wherein the composition of the medium is set at the beginning of the fermentation and is not subject to artificial alterations during the fermentation. Thus, at the beginning of the fermentation the medium is inoculated with the desired organism(s). In this method, fermentation is permitted to occur without the addition of any components to the system. Typically, a batch fermentation qualifies as a "batch" with respect to the addition of the carbon source and attempts are often made at controlling factors such as pH and oxygen concentration. The metabolite and biomass compositions of the batch system change constantly up to the time the fermentation is stopped. Within batch cultures, cells progress through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase eventually die. In general, cells in log phase are responsible for the bulk of production of end product.

[87] A variation on the standard batch system is the "fed-batch fermentation" system, which also finds use with the present invention. In this variation of a typical batch system, the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Measurement of the actual substrate

- 20 -

concentration in fed-batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO₂. Batch and fed-batch fermentations are common and known in the art.

[88] Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth.

[89] Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth and/or end product concentration. For example, in one embodiment, a limiting nutrient such as the carbon source or nitrogen source is maintained at a fixed rate and all other parameters are allowed to moderate. In other systems, a number of factors affecting growth can be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions. Thus, cell loss due to medium being drawn off must be balanced against the cell growth rate in the fermentation. Methods of modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are known.

[90] A fungal host cell may be cultured in a standard medium containing physiological salts and nutrients (See, e.g., Pourquie, J. et al., BIOCHEMISTRY AND GENETICS OF CELLULOSE DEGRADATION, eds. Aubert, J. P. et al., Academic Press, pp. 71-86, 1988 and Ilmen, M. et al., (1997) Appl. Environ. Microbiol. 63:1298-1306). Common commercially prepared media (e.g., Yeast Malt Extract (YM) broth, Luria Bertani (LB) broth and Sabouraud Dextrose (SD) broth) also find use in the present methods. Preferred culture conditions for fungal host cells are known in the art and may be found in the scientific literature and/or from the source of the fungi such as the American Type Culture Collection (ATCC) and Fungal Genetics Stock Center.

[91] Protein may be recovered from growth media by any convenient method, e.g., by precipitation, centrifugation, affinity, filtration or any other method known in the art. In another embodiment, a culture of cells is provided, where the culture of cells comprises: a) growth medium and b) the above-described host cell.

1818A and 1818B promoters

[92] A promoter that may be used to express a protein in a host cell is also provided. In

- 21 -

one embodiment, the promoter comprises the nucleotide sequence of SEQ ID NOS:7 or 8 (shown in Figs. 5A and 5B, respectively), or a subsequence or functional equivalent thereof that has promoter activity in a host cell. Also provided are recombinant nucleic acids and vectors containing the promoter and host cells containing a recombinant nucleic acid or vector.

5 Methods of producing a protein using the host cells are also provided.

[93] In certain embodiments, the promoter may comprise the nucleotide sequence of: a) SEQ ID NO: 7 or 8; b) a subsequence of SEQ ID NO: 7 or 8 that retains promoter activity; c) a functionally equivalent sequence of SEQ ID NO: 7 or 8 that retains promoter activity or d) a nucleic acid sequence that hybridizes under stringent hybridization conditions with SEQ ID
10 NO: 7 or 8, or the subsequence thereof. In particular embodiments, the nucleotide sequence may be at least 80% identical to the nucleotide sequence of SEQ ID NO: 7 or 8.

[94] In particular embodiments, a subsequence of a subject promoter may contain at least about 100 nucleotides, at least about 200 nucleotides; at least about 250 nucleotides; at least about 300 nucleotides; at least about 350 nucleotides; at least about 400 nucleotides; at least
15 about 450 nucleotides; least about 500 nucleotides; least about 550 nucleotides; least about 600 nucleotides; at least about 650 nucleotides; at least about 700 nucleotides; at least about 800 nucleotides; at least about 850 nucleotides that are contiguous in SEQ ID NO: 7 or 8, the entire contiguous sequence of SEQ ID NO: 7 or 8, or a functional equivalent thereof that retains promoter activity.

[95] In certain embodiments, a functional equivalent promoter may include one or more
20 changes (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, more than 10, up to 20 or 30 or more changes) relative to the nucleotide sequence of SEQ ID NO: 7 or 8, where a change can be a deletion, substitution or insertion, for example. In one exemplary embodiment, the nucleotide sequence of the functional equivalent promoter may include one to five nucleotide differences relative to the
25 nucleotide sequence of the parent promoter such as SEQ ID NO: 7 or 8.

[96] In other embodiments, the promoter may include a nucleotide sequence that hybridizes under stringent hybridization conditions to a polynucleotide having the nucleotide sequence of SEQ ID NO: 7 or 8, where stringent hybridization conditions encompass low, medium, high and very high stringency hybridization conditions, where such conditions are described above.

[97] In another embodiment, a subject promoter may contain a contiguous nucleotide
30 sequence that is 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% identical to SEQ ID NO: 7 or 8, or a subsequence thereof. In one embodiment, the subject promoter may contain a contiguous nucleotide sequence that is at least 95% identical to SEQ ID NO: 7 or 8.

- 22 -

[98] As noted in the Examples section below, the nucleic acid of SEQ ID NOS: 7 and 8 were obtained from *Trichoderma reesei*, a filamentous fungi. As would be readily apparent, functional equivalents of SEQ ID NO: 7 and 8 that retain promoter activity can be identified by identifying sequences that are similar to SEQ ID NO: 7 and 8 in other filamentous fungi. Since the most or all of the genome sequences of other filamentous fungi, e.g., *Aspergillus* (e.g., *Aspergillus fumigatus*, *Aspergillus oryzae* (see, e.g., Machida et al, Nature 2005 438, 1157-1161), *Aspergillus nidulans*, *Aspergillus fumigatus*, *Aspergillus niger*, *Aspergillus flavus*, *Aspergillus terreus*), *Neurospora* (e.g., *Neurospora crassa*), and *Fusarium* (e.g., *Fusarium graminearum*) are available, functional equivalents of SEQ ID NO: 7 and 8 that have promoter activity are readily identifiable.

[99] As noted above, a subject promoter may have promoter activity in a host cell. Promoter activity may be detected using any suitable assay. In certain embodiments, a subject promoter may be operably linked to a polynucleotide, and transcription of the polynucleotide may be detected using any suitable method, e.g., Northern blotting or RT-PCR, etc. In other embodiments, the promoter may be operably linked to a polynucleotide that encodes a protein, e.g., a reporter protein, and the activity of the promoter can be evaluated by detecting the protein. In these embodiments, if necessary, a 5' untranslated region may be linked to the promoter such that the resultant transcript has a 5' UTR followed by a coding sequence. As would be recognized, the results obtained from such an assay may be compared to results compared to a suitable control, e.g., a negative or positive control, to determine the significance of results obtained. Any host cell, e.g., a bacterial host cell such as *E. coli*, *Bacillus* or *Streptomyces* host cell, or a filamentous fungal cell, e.g., an *Aspergillus* ssp., *Trichoderma* ssp. or *Fusarium* ssp. host cell may be employed. There is no requirement for a subject promoter to be contained within a particular host cell. In certain cases, the promoter may be tested for promoter activity in a *Trichoderma reesei* host cell.

[100] A recombinant nucleic acid comprising the subject promoter is also provided. In certain cases, the recombinant nucleic acid may comprise a subject promoter and a polynucleotide, where the promoter and the polynucleotide are operably linked such that the promoter causes transcription of the polynucleotide in a cell. In certain cases, the promoter and polynucleotide are not normally linked in nature, i.e., are heterologous to each other. In certain cases, the polynucleotide may contain a coding sequence for a protein. The protein may be an enzyme, a reporter or a therapeutic protein (e.g., an antibody protein), as discussed above, for example. In certain embodiments, the protein may be a fusion protein which may, in certain cases, contain a signal sequence or carrier portion for secretion of the protein.

- 23 -

[101] A nucleic acid vector comprising the subject recombinant nucleic acid is also provided, as well as a host cell containing the same. In certain embodiments, the recombinant nucleic acid may be present in the genome of the host cell. In other embodiments, the recombinant nucleic acid may be present in a vector that replicates in the cell. The host cell may be any of a variety of different host cells, including bacterial, fungal, yeast, plant and mammalian host cells. In one embodiment, the host cell may be a filamentous fungal host cell, and in another embodiment, the host cell may be a bacterial cell.

[102] A culture of cells comprising culture medium and a subject host cell is also provided.

[103] A method of producing a protein is also provided. In general terms, this method includes maintaining a subject culture of cells under conditions suitable to produce the protein. This method may further include recovering the protein from culture medium.

[104] In order to further illustrate the present invention and advantages thereof, the following specific examples are given with the understanding that they are being offered to illustrate the present invention and should not be construed in any way as limiting its scope.

EXAMPLE 1

Isolation of ALS genes from chlorimuron ethyl-resistant cells

[105] Chlorimuron ethyl (Chem Service Inc., West Chester, PA), freshly prepared and dissolved in DMF to a concentration of 20 mg/ml, was added at various concentrations to molten Vogels agar immediately prior to pouring the plates to produce media containing 25, 50, 100, 200, 300, 400 or 500 ppm chlorimuron ethyl.

[106] Approximately 25 million spores *T.reesei* strain QM6a (ATCC13631) were plated onto each plate. After 10 days of growth at 28 °C, five colonies were isolated, two from 50 ppm plates and one each from each of the 25, 200 and 300 ppm plates. The colonies were further isolated by restreaking them onto fresh Vogels agar plates containing 200 ppm chlorimuron ethyl.

[107] Genomic DNA was prepared from each of the five chlorimuron ethyl resistant strains in order to amplify and sequence the acetolactate synthase (*als*) gene from those strains.

Herculase DNA polymerase (Stratagene, La Jolla, CA) was used for amplification following manufacturers instructions and the following primers: Forward (5'-3')

GGCGCGCCTGAGACAATGGCCGGCAATGGTAAAAA (SEQ ID NO:5) and Reverse (5'-3') GCGATCGCCATCCCGTCGCGTCAAAAACACTGC (SEQ ID NO:6)

[108] Unique restriction sites were added to the 5' ends of the primers for subsequent

- 24 -

manipulation. The resulting 4.1 kb fragments were isolated, sequenced and compared to the sequence of the native *T. reesei* acetolactate synthase gene (from the JGI genome website).

[109] Three unique point mutations in the *als* gene that confer resistance to chlorimuron ethyl were identified.

5 [110] Compared to the sequence of the wild-type *T. reesei* acetolactate synthase protein (Fig. 1; SEQ ID NO:1), the chlorimuron ethyl-resistant acetolactate synthase proteins each had one of the following amino acid substitutions: A190D, K241E or R372H. The amino acids at positions 190, 241 and 372 of the wild-type *T. reesei* acetolactate synthase protein of SEQ ID NO:1 are underlined.

10 [111] Compared to the sequence of the wild-type *T. reesei* acetolactate synthase deduced cDNA (Fig. 2; SEQ ID NO:2), the chlorimuron ethyl-resistant acetolactate synthase cDNAs each had one of the following nucleotide substitutions: C569A (this corresponds to the A190D amino acid substitution), A721G (this corresponds to the K241E amino acid substitution) or G1115A (this corresponds to the R372H amino acid substitution). Each of the altered codons
15 (GCC, which is altered to GAC; AAG, which is altered to GAG; and CGT, which is altered to CAT) is indicated on Fig. 2.

[112] Compared to the sequence of the wild-type *T. reesei* acetolactate synthase gene (Fig. 3; SEQ ID NO:3), the chlorimuron ethyl-resistant acetolactate synthase genes each had one of the following nucleotide substitutions: C1023A (this corresponds to the A190D amino acid
20 substitution), A1175G (this corresponds to the K241E amino acid substitution) or G1569A (this corresponds to the R372H amino acid substitution).

[113] Because of the degeneracy of the genetic code, other mutations in the *als* coding sequence can encode the A190D, K241E or R372H amino acid substitutions.

25 EXAMPLE 2

Transformation of *T.reesei* with A190D acetolactate synthase gene

[114] The vector pTrex-glucoamylase was created to express a glucoamylase derived from *Trichoderma reesei* in *T.reesei*. Figure 4 depicts a general diagram of the vector. Three
30 different vector constructs have been made with different promoters; 1818A, 1818B and the *T. reesei cbh1* promoter. Fig.5A illustrates the sequences of the promoter designated 1818A and Fig. 5B illustrates the sequences of the promoter 1818B.

[115] The entire nucleotide sequence of the pTrex-glucoamylase including the *cbh1* promoter is shown in Fig. 6. The pTrex-glucoamylase vector is based on vector pTrex3g as described in

- 25 -

detail in WO 05/001036. In brief, the pTrex3g is based on the *E. coli* vector pSL1180 (Pharmacia, Inc., Piscataway, NJ) which is a pUC118 phagemid based vector with an extended multiple cloning site containing 64 hexamer restriction enzyme recognition sequences. It was designed as a Gateway destination vector (Hartley, J.L. et al., (2000) Genome Research 10:1788 – 1795) to allow insertion using Gateway Technology (Invitrogen) of any desired open reading frame between the promoter and terminator regions of the *T. reesei* cbh1 gene.

[116] In the pTrex-glucoamylase vector, the A190D ALS gene is under control of its native promoter and terminator and is used to replace the fungal selectable marker *amdS* which is used in pTrex3g.

[117] The vector was transformed into the quad deleted ($\Delta chb1$, $\Delta cbh2$, $\Delta egl1$, and $\Delta egl2$) *T. reesei* strain (WO 05/001036) originally derived from RL-P37 (Sheir-Neiss et al., (1984) Appl. Microbiol. Biotechnol. 20:46 – 53; USP 4,797,361) using the procedure outlined below.

[118] A suspension of spores (approximately 5×10^8 spores/ml) from the *Trichoderma* strain was prepared. 100ul – 200ul of spore suspension was spread onto the center of plates of modified Vogels medium with 200 ppm chlorimuron ethyl and allowed to dry.

[119] Modified Vogels had the following composition: 2.5 g/L $\text{Na}_3\text{Citrate} \cdot 2\text{H}_2\text{O}$, 5.0 g/L KH_2PO_4 , 2.0 g/L NH_4NO_3 , 0.2 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.1 g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 5 mL/L Modified Vogels Trace Elements Solution, 2.5 mL/L Modified Vogels Biotin Solution, 20 g/L Agar. Modified Vogels trace elements solution contained 50 g/L Citric Acid, 50 g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 10 g/L $\text{Fe}(\text{NH}_4)_2\text{SO}_4 \cdot 6\text{H}_2\text{O}$, 2.5 g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.5 g/L $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$, 0.5 g/L H_3BO_3 , 0.5 g/L $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$.

[120] Modified Vogels Biotin Solution contained 0.1 g/L d-Biotin. After autoclave the following additions are made prior to pouring plates: 20 mL/L of 50% glucose, 10 mL/L of 20 mg/mL chlorimuron ethyl dissolved DMF.

[121] Transformation of the *Trichoderma* strain by the biolistic transformation method was accomplished using a Biolistic® PDS-1000/he Particle Delivery System from Bio-Rad (Hercules, CA) following the manufacturers instructions (see, WO 05/001036 and US 2006/0003408).

[122] Transformants were isolated after 3 to 4 days of growth at 28°C. Transformants were serially passed twice on fresh Vogels plates with 200 ppm chlorimuron ethyl in order to isolate stable monokaryons. Transformants were cultured, and culture supernatant was tested by SDS page electrophoresis and using an enzyme assay. An SDS PAGE gel is shown in Fig. 7 illustrating the expression of the *T. reesei* glucoamylase.

- 26 -

EXAMPLE 3**Sequencing and cloning of the 1818A and 1818B promoters**

5 [123] The 1818A and 1818B promoters (as shown in Figs 5A and 5B, respectively) were identified by mining the US Department of Energy Joint Genome Institute's *Trichoderma reesei* genome sequence database (as found at the world wide website of jgi.doe.gov) for sequences upstream of highly represented ESTs.

10 [124] The promoters were amplified by PCR and cloned behind a glucoamylase coding sequence in accordance with the vector map shown in Fig. 8. The entire nucleotide sequence of this construct is shown in Fig. 9. The vector used is based on the vector pTrex3g, which is described in detail in Example 6 of WO 05/001036. In brief, the pTrex3g is based on the *E. coli* vector pSL1180 (Pharmacia, Inc., Piscataway, NJ) which is a pUC118 phagemid based vector with an extended multiple cloning site containing 64 hexamer restriction enzyme recognition
15 sequences. It was designed as a Gateway destination vector (Hartley, J.L. et al., (2000) Genome Research 10:1788 – 1795) to allow insertion using Gateway Technology (Invitrogen) of any desired open reading frame between the promoter and terminator regions of the *T. reesei* *cbh1* gene. In the vector used in this example, the *cbh1* promoter was replaced with either the 1818A or the 1818B promoter, the *amdS* selectable marker was replaced with a *Trichoderma reesei*
20 acetolactate synthase marker, driven by its native promoter and terminator.

EXAMPLE 4**Transformation of *T. reesei* host cells with vectors including the 1818A and 1818B promoters**

25 [125] The vector was transformed into the quad deleted ($\Delta chb1$, $\Delta cbh2$, $\Delta egl1$, and $\Delta egl2$) *T. reesei* strain (WO 05/001036) originally derived from RL-P37 (Sheir-Neiss et al., (1984) Appl. Microbiol. Biotechnol. 20:46 – 53 and USP 4,797,361) using the procedure outlined below.

30 [126] A suspension of spores (approximately 5×10^8 spores/ml) from the *Trichoderma* strain was prepared. 100ul – 200ul of spore suspension was spread onto the center of plates of modified Vogels medium with 200 ppm chlorimuron ethyl. Modified Vogels had the following composition: 2.5 g/L $\text{Na}_3\text{Citrate} \cdot 2\text{H}_2\text{O}$, 5.0 g/L KH_2PO_4 , 2.0 g/L NH_4NO_3 , 0.2 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, .1 g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 5 mL Modified Vogels Trace Elements Solution, 2.5 mL

- 27 -

Modified Vogels Biotin Solution, 20 g/L Agar Modified Vogels trace elements solution contained 50 g/L Citric Acid, 50 g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 10 g/L $\text{Fe}(\text{NH}_4)_2\text{SO}_4 \cdot 6\text{H}_2\text{O}$, 2.5 g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.5 g/L $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$, 0.5 H_3BO_3 , 0.5 g/L $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$. Modified Vogels Biotin Solution contained 0.1 g/L d-Biotin. After autoclave the following additions are made
5 prior to pouring plates: 20 mL/L of 50% glucose, 10 mL/L of 20 mg/mL chlorimuron ethyl dissolved DMF. The spore suspension was allowed to dry on the surface of the Modified Vogels plates.

[127] Transformation of the *Trichoderma* strain by the biolistic transformation method was accomplished using a Biolistic® PDS-1000/he Particle Delivery System from Bio-Rad
10 (Hercules, CA) following the manufacturers instructions (see, WO 05/001036 and US 2006/0003408).

[128] Transformants were isolated.

EXAMPLE 5

15 Screening transformants for expression of glucoamylase activity driven by 1818A and 1818B promoters

[129] Stable transformants were grown on Modified Vogels Lactose Agar plates with starch (9cm diameter petri-plates). After about 4 days growth at 28°C, 10 ml of a 1mg/ml solution of
20 4-methyl-umbelliferyl- α -D-glucose was poured over the grown colonies. After 30 minutes at room temperature, strains expressing glucoamylase were visualized as fluorescent blue colonies when viewed illuminated by a long wavelength UV lamp. The untransformed, parent *T. reesei* control strain did not show blue fluorescence.

[130] Modified Vogels Lactose Agar Plates with starch is the same recipe as Modified Vogels
25 Agar except that 25 ml/L of 20% α -lactose solution (added after autoclaving) is substituted for glucose solution and 20 g/L of Pure Food Powder cornstarch is added prior to autoclaving.

[131] The glucoamylase substrate 4-methyl-umbelliferyl- α -D-glucose was prepared as follows: 200 mg of 4-methyl-umbelliferyl- α -D-glucose (Sigma-Aldrich Co.) is dissolved in 5 ml of DMSO and 195 ml of 75 mM Potassium Phosphate Buffer pH 6.3 is added.

EXAMPLE 6

Shake Flask Experiments with *Trichoderma reesei* transformants

[132] Individual fungal transformants will be grown up in shake flask culture to determine the

- 28 -

level of glucoamylase protein expression. The experiments will be conducted essentially as described in example 1 of U.S. Patent 5,874,276 with the following modification: 16 g/L of alpha-lactose was substituted for cellulose in TSF medium.

[133] In general, the fermentation protocol as described in Foreman *et al.* (Foreman *et al.* (2003) *J. Biol. Chem* 278:31988-31997) will be followed. Vogels minimal medium (Davis *et al.*, (1970) *Methods in Enzymology* 17A, pg 79 - 143 and Davis, Rowland, *NEUROSPORA, CONTRIBUTIONS OF A MODEL ORGANISM*, Oxford University Press, (2000)) containing 5% glucose will be inoculated with 1.5 ml frozen spore suspension. After about 48 hours, each culture will be transferred to 6.2L of the same medium in a 14L BiolaFitte fermenter. The fermenter will be run at 25°C, 750 RPM and 8 standard liters per minute airflow. About one hour after the initial glucose is exhausted, a 25% (w/w) lactose feed will be started and fed in a carbon limiting fashion to prevent lactose accumulation. The concentrations of glucose and lactose will be monitored. Samples will be obtained at regular intervals to monitor the progress of the fermentation. Collected samples will be spun in a 50ml centrifuge tube at 3/4 speed in an International Equipment Company (Needham Heights, MA) clinical centrifuge. Shake flask grown supernatant samples will be run on BIS-TRIS SDS -PAGE gels (Invitrogen), under reducing conditions with MOPS (morpholinepropanesulfonic acid) SDS running buffer and LDS sample buffer.

- 29 -

CLAIMS

1. An isolated polynucleotide encoding an acetolactate synthase protein that provides resistance to toxic sulphonylurea compounds, wherein said acetolactate
5 synthase protein comprises an amino acid sequence that is at least 95% identical to the full length of SEQ ID NO: 1.
2. The isolated polynucleotide of claim 1, wherein said protein contains an acidic amino acid at position 190.
- 10 3. The isolated polynucleotide of claim 1, wherein said protein contains a basic amino acid at position 241.
4. The isolated polynucleotide of claim 1, wherein said protein contains a basic
15 amino acid at position 372.
5. The isolated polynucleotide of claim 1, wherein said isolated polynucleotide is operably linked to a promoter and a terminator.
- 20 6. The isolated polynucleotide of claim 1, wherein said isolated polynucleotide comprises introns.
7. The isolated polynucleotide of claim 6, wherein said isolated polynucleotide comprises a nucleotide sequence that is at least 95% identical to the full length of
25 SEQ ID NO:4.
8. The isolated polynucleotide of claim 1, wherein said isolated polynucleotide comprises a single open reading frame.

- 30 -

9. The isolated polynucleotide of claim 1, wherein said isolated polynucleotide comprises a nucleotide sequence that is at least 95% identical to the full length of SEQ ID NO:2.

5 10. The isolated polynucleotide of claim 1, wherein said isolated polynucleotide is codon-optimized for expression of said acetolactate synthase protein in a particular host cell.

11. A vector comprising the isolated polynucleotide of claim 1.

10 12. The vector of claim 11, wherein said vector further comprises an expression cassette for expression of a recombinant protein.

13. A host cell comprising the isolated polynucleotide of claim 1, wherein said
15 host cell is resistant to toxic sulphonylurea compounds.

14. The host-cell of claim 13, wherein said host cell is a fungal host cell.

15. The host cell of claim 14, wherein said host cell is a filamentous fungal host
20 cell.

16. The host cell of claim 15, wherein said host cell is a *Trichoderma reesei* host cell.

25 17. The host cell of claim 13, wherein said host cell is a plant cell.

18. The host cell of claim 13, wherein isolated polynucleotide is present in a genome of said host cell.

30 19. The host cell of claim 13, wherein said isolated polynucleotide is present in a vector that autonomously replicates in said host cell.

- 31 -

20. A method of selecting a cell resistant to a toxic sulphonylurea or imidazoline compound comprising:

introducing an isolated polynucleotide of claim 1 into a plurality of cells;

5 contacting said plurality of cells with the toxic sulphonylurea or imidazolinone compound;

and culturing said cells to provide for selection of said cell.

21. The method of claim 20, wherein said cell is a fungal cell.

10

22. The method of claim 21, wherein said cell is a *Trichoderma reesei* cell.

23. The method of claim 20, wherein said cell is a plant cell.

15 24. The method of claim 20, further comprising culturing said cell in liquid medium.

25 25. An isolated acetolactate synthase protein comprising an amino acid sequence that is at least 95% identical to the full length of SEQ ID NO: 1, wherein the protein provides resistance to toxic sulphonylurea compounds.

26. The isolated protein of claim 25, wherein the protein comprises the full length amino acid sequence of SEQ ID NO: 1.

25 27. The isolated protein of claim 25, wherein said protein contains an acidic amino acid at position 190.

28. The isolated protein of claim 25, wherein said protein contains a basic amino acid at position 241.

30

- 32 -

29. The isolated protein of claim 25, wherein said protein contains a basic amino acid at position 372.

1/15

MLRSRQVTARAVRALGQARAFTSTTKPVMIQSSQRKQANASAAPQVRPVP
SPAFNAEDKDRSHVQPLVNPSKPDMDSEFIGKTGGEIFHEMMLRQGVKHI
FGYPGGAILPVFDAIYNSKHFDLPRHEQGAGHMAEGYARASGKPGVVL
VTSGPGATNVITPMQDALSDGTPLVVFCGQVPTTAIGSDAFQEADVVGIS
RACTKWNVMVKSVAELPRRINEAFEIATSGRPGPVLVDLPKDVTAGILRR
AIPETALPSLPSAASRAAMELSSKQLNASIKRAADLINIAKKPVIYAGQ
GVIQSEGGVELLKQLADKASIPVTTTLHGLGAFDELDEKSLHMLGMHGSA
YANMAMQQADLI IALGSRFDDRVTLNVSKFAPAARQAAAEGRGGIIHFEI
MPKNINKV IQATEAVEGDVATNLKHLIPQIAEKSMADRGGEWFGLINEWKK
KWPLSNYQRAERAGLIKPTVMEEISNLTANRKDKTYIATGVGQHQMWA
QHFRWRHPRSMITSGGLGTMGYGLPAAIGAKVAQPDALVIDVDGDASFNM
TLTELSTAAQFNIGVKVVVLNNEEQGMVTQWQNLFYEDRYAHTHQKNPDF
MKLADAMGVQHQRVTEPEKLVDALTWLINTDGPALLEVVTDKKVPVLPV
PAGSALHEFLVFEPEKDKQRRELMKERTKGVHS

Fig. 1

2/15

ATGCTCCGAAGTCGCCAAGTCACAGCCAGGGCCGTCCGGGCTCTGGGCCAGGCGCG
CGCCTTTACCTCGACGACCAAGCCTGTGATGATCCAGAGCAGCCAGAGGAAACAGG
CCAACGCCAGCGCTGCTCCACAAGTCCGCCCTGTACCGAGCCCTGCTTTCAACGCT
GAAGACAAAGACCGCAGCCATGTGCAGCCTCTGGTCAACCCGTCGAAGCCCGACAT
GGATGAATCATTCATTGGCAAAACCGGAGGCGAAATCTTCCACGAGATGATGCTGC
GACAGGGTGTCAAGCACATTTTCGGATACCCTGGCGGCGCTATCCTGCCCGTCTTC
GACGCCATCTACAACTCAAAACACTTCGACTTCATCCTGCCCCGTGATGAGCAGGG
AGCTGGCCATATGGCCGAGGGCTATGCCCGTGCCCTCGGGCAAACCCGGTGTCGTCC
TGGTGAATTCGGGCCCCGGTGCTACCAATGTCATCACGCCCATGCAGGATGCCCTG
TCGGACGGAACGCCCTTGGTGTCTTCTGCGGCCAGGTCCCCACCACGGCCATCGG
CAGCGATGCCCTTCCAAGAGGGCCGACGTGCTGGGCATCTCGCGGGCCTGCACCAAGT
GGAACGTGATGGTCAAGAGCGTTGCTGAGCTGCCGCGGAGAATCAACGAGGCCTTT
GAGATTGCCACCAGCGGCCGCCCTGGCCCCGTCTCTGTCGACCTGCCCCAAGGATGT
CACGGCTGGTATCCTGAGGAGAGCCATCCCTACGGAGACTGCTCTGCCGTCTCTGC
CCAGTGCCGCCTCCCGCGCCGCCATGGAGCTGAGCTCCAAGCAGCTCAACGCCTCC
ATCAAGCGTGCCGCCGACCTCATCAACATCGCCAAGAAGCCCGTCATCTACGCCGG
TCAGGGTGTGATCCAGTCCGAGGGCGGCGTTGAGCTCCTGAAGCAGCTGGCGGACA
AGGCCTCCATCCCCGTGACCACCACCCTCCATGGCCTGGGTGCCTTTGATGAGCTG
GACGAGAAGTCGCTGCACATGCTGGGCATGCACGGCTCGGCGTATGCCAACATGGC
CATGCAGCAGGCCGACCTCATCATCGCCCTCGGCAGCCGATTGACGACCGTGTTA
CTCTGAATGTCTCCAAATTTGCGCCTGCAGCCAGGCAAGCTGCTGCCGAGGGCCGC
GGCGGCATCATTCACCTTTGAGATCATGCCCAAGAACATCAACAAGGTCATCCAGGC
GACCGAGGCCGTGAGGGCGACGTCGCCACCAACCTGAAGCACCTCATTTCCCAGA
TTGCCGAAAAGTCCATGGCGGACCGAGGAGAGTGGTTCGGCCTCATCAATGAGTGG
AAGAAGAAGTGGCCCCCTGTCAAACCTACCAGCGCGCGGAGCGGGCTGGCCTCATCAA
GCCGCAGACGGTCATGGAGGAGATTAGCAACCTGACGGCCAACCGAAAGGACAAGA
CGTACATTGCCACGGGTGTGCGCCAGCACCAGATGTGGGTTGCCAGCACTTCCGC
TGGAGGCACCTCGATCCATGATTACCTCTGGTGGTCTGGGCACCATGGGCTACGG
TCTCCCCGCGGCCATTGGCGCCAAGGTGGCCCAGCCCGACGCTCTCGTAATTGACG
TTGATGGCGATGCCTCGTTTAACATGACGCTGACGGAGCTGTCGACTGCTGCACAG
TTCAACATTGGCGTCAAGGTGGTTGTGCTCAACAACGAGGAGCAGGGCATGGTGAC
GCAGTGGCAGAACCTCTTTTACGAGGACCGATATGCCACACGCACCAGAAGAACC
CCGACTTCATGAAGCTGGCCGACGCCATGGGCGTTTCAACACCAGCGCGTGACGGAG
CCGGAGAAGCTGGTCGATGCCCTGACGTGGCTGATCAACACCGATGGCCCCGGCCCT
GTTGGAGGTTGTACGGACAAGAAGGTGCCTGTCTGCCCATGGTGCCCGCCGGAT
CGGCCCTGCACGAGTTCCTCGTCTTTGAACCTGAAAAGGATAAGCAGCGCCGTGAG
CTGATGAAGGAGAGAACAAAGGGTGTGCACTCCTAA

Fig. 2

3/15

TGAGACAATGGCCGGCAATGGTAAAAAGGACCAAGATGTACTAGGTAGTTGCAATGTGGCTTAT
TACCTACCTACTACCTGGTAGGCACCTACTAGGTACTTGGGTAGACGGACAATGAAATTTGAAG
TCGGGGTTGCAGGAAAGCAGGGCGCTGGACACATTGTGCTTCAGGCGGTACCCGTCTCATCGT
CAGCCAATGTTCGAGGCCCGGCAGCCCGAGGAGCGAGACAACCTTGGCCGGAGGAGCCCGCAGGT
ACCTGCCAAAGCGCGGCTGGTACCTCTCAACCTCTCAGGCCCTGTTGGATGCCCTATGACATGC
CCTGGGGGATGCAGCTGTTGCCCCGGCCCCGCACTTTTCGGGTGACCGCGAGGCTGCTGATTGGC
TGGTTGCCACGGGCTGGGCGGTCCCTGAAGTTGTTGCCATCTGAACTCTGTCGGCGCTGGCGTC
GGCTGCGCCCAATGGGAGGCGAGACAACCTCAGGGTACTAGAATCACTGACAGAAGAAGAGAATC
GAAAGTAGGTAGACAGCCAATTCTGTTGCATGGCAGGCAACCGCACAGGAGAAAAATTGACTACC
CCACAATCAGGCACAGTAAGTAGGGCACAGTACGTATGTACAGACAAGGCGCAAGCGATACTGC
GCGACCCGGTACCTCGCCGGCTTGACACGTGCGACAGGCTACTTTACTAGTATTCGCAGCGGCG
GGTCGCGCATTATTACATGTACTGTGCCGCCATTTGATGACTGGGCTGCTGCAGTATTAGTAGA
TCTGCCCGGCATCGCCCTTCCATGGGCGCGACCCGGGACTGGACCTCTGACTCTACCTACATG
TACCTAGGCCGGGCGGGCTTGGTGACTTTTGTCCGATCAGGTCGTTTCGCTGGCTACCTATTA
TTTCTCTTTCTTCTTCTCCATCCTGCTTCTGGCCTTGCAATTCTTCTTCGCCACTCCTCCCTCT
TCCCCCGCGATACCCTTGAATTCGTACAGAGGAAAAGACGAGAAAAAAAAGGGCAGCAGAGA
CGTCGGTCTGGCTCACGTGCTGCATCTCTGCGCACTCTCATTTTTTTTTATTGTCCGACCCCTCC
CTCAACCTTCTCCTTCGTTGACAGGCTAAGCCTTGCTTCGACGCTCTCTCTTTGAATTTTTCTA
CTTCTACCTTCTTTTCTTGCGTGTTACCCACCATAGCTCGATTACGATGCTCCGAAGTCGCCA
AGTCACAGCCAGGGCCGTCCGGGCTCTGGGCCAGGCGCGCGCCTTTACCTCGACGACCAAGCCT
GTCATGATCCAGAGCAGCCAGAGGAAACAGGCCAACGCCAGCGCTGCTCCGTAAGTCGCCCAATT
GCCATTGCATCTTCTGTTTGATATATACTTCTGCTGCTTGCGTGCGCTGCTCTCTCGGTTATG
CGTGTCAAGGACCAGGTGTGTTTCGCATCGTGGTTTTCCAGCGCCGATTACCGGGGGACGAATTT
TTGGCTGCTCAACTCGCGCGCGCGCATTCTGATTCTTCGTTTTCAATCTTGAGCGACAACCTGGC
TAACATAATGGCCATTGGCAATTGCTTCACACAGACAAGTCCGCCCTGTACCGAGCCCTGCTTT
CAACGCTGAAGACAAAGACCGCAGCCATGTGCAGCCTCTGGTCAACCCGTCGAAGCCCGACATG
GATGAATCGTATGTCCACGTCCCTCGTCCCGCCCTACAAAATGAACACGATTACACCAGAATT
TTTGCAACAATCGACACTTCTATAACAGACCAATTGAGCTTTGTTCTGACCAATCATGTTGCTC
TAGATTCAATTGGCAAACCGGAGGCGAAATCTTCCACGAGATGATGCTGCGACAGGGTGTCAG
CACATTTGTAGGTTCCGATGCCGGCCGCCACACGGGCTCCATCCTTGCTCCATCTCTCCAGCT
AGGCAAATCTCGCTAACCTTGAGTCACCATCCAGTCGGATACCCTGGCGGCGCTATCCTGCCCG
TCTTCGACGCCATCTACAACTCAAAACACTTCGACTTCATCCTGCCCCGTCATGAGCAGGGAGC
TGGCCATATGGCCGAGGGCTATGCCCGTGCTCGGGCAAACCCGGTGTCGTCCTGGTGACTTCC
GGCCCCGGTGCTACCAATGTCATCACGCCCATGCAGGATGCCCTGTTCGGACGGAACGCCCTTGG
TCGTCTTCTGCGGCCAGGTCCCCACCACGGCCATCGGCAGCGATGCCTTCCAAGAGGCCGACGT
CGTGGGCATCTCGCGGGCCTGCACCAAGTGGAACGTCATGGTCAAGAGCGTTGCTGAGCTGCCG
CGGAGAATCAACGAGGCCCTTTGAGATTGCCACCAGCGGCCCGCCCTGGCCCCGTCCTCGTCGACC
TGCCCAAGGATGTCACGGCTGGTATCCTGAGGAGAGCCATCCCTACGGAGACTGCTCTGCCGTC
TCTGCCCAGTGCCGCCTCCCGCGCCGCCATGGAGCTGAGCTCCAAGCAGCTCAACGCCTCCATC
AAGCGTGCCGCCGACCTCATCAACATCGCCAAGAAGCCCGTCATCTACGCCGGTCAGGGTGTC
TCCAGTCCGAGGGCGGCGTTGAGCTCCTGAAGCAGCTGGCGGACAAGGCCTCCATCCCCGTCAC
CACCACCCTCCATGGCCTGGGTGCCTTTGATGAGCTGGACGAGAAGTCGCTGCACATGCTGGGC
ATGCACGGCTCGGCGTATGCCAACATGGCCATGCAGCAGGCCGACCTCATCATCGCCCTCGGCA

4/15

GCCGATTCGACGACCGTGTTACTCTGAATGTCTCCAAATTTGCGCCTGCAGCCAGGCAAGCTGC
TGCCGAGGGCCGCGGCGGCATCATTCACTTTGAGATCATGCCCAAGAACATCAACAAGGTCATC
CAGGCGACCGAGGCCGTCGAGGGCGACGTCGCCACCAACCTGAAGCACCTCATTCCCCAGATTG
CCGAAAAGTCCATGGCGGACCGAGGAGAGTGGTTCGGCCTCATCAATGAGTGGAAGAAGAAGTG
GCCCCGTGTCAAACCTACCAGCGCGCGGAGCGGGCTGGCCTCATCAAGCCGCAGACGGTCATGGAG
GAGATTAGCAACCTGACGGCCAACCGAAAGGACAAGACGTACATTGCCACGGGTGTCGGCCAGC
ACCAGATGTGGGTTGCCCAGCACTTCCGCTGGAGGCACCCTCGATCCATGATTACCTCTGGTGG
TCTGGGCACCATGGGCTACGGTCTCCCCGCGGCCATTGGCGCCAAGGTGGCCAGCCCGACGCT
CTCGTAATTGACGTTGATGGCGATGCCTCGTTTAACATGACGCTGACGGAGCTGTCGACTGCTG
CACAGTTCAACATTGGCGTCAAGGTGGTTGTGCTCAACAACGAGGAGCAGGGCATGGTGACGCA
GTGGCAGAACCTCTTTTACGAGGACCGATATGCCCACACGCACCAGAAGAACCCCGACTTCATG
AAGCTGGCCGACGCCATGGGCGTTCAGCACCAGCGCGTGACGGAGCCGGAGAAGCTGGTCGATG
CCCTGACGTGGCTGATCAACACCGATGGCCCGGCCCTGTTGGAGGTTGTCACGGACAAGAAGGT
GCCTGTCCTGCCCATGGTGCCCGCCGGATCGGCCCTGCACGAGTTCCTCGTCTTTGAACCTGGT
GAGTCTACTTCAGACATATTGCTTGCGCATTGCAGATACTAACACTCTCACAGAAAAGGATAAG
CAGCGCCGTGAGCTGATGAAGGAGAGAACAAAGGGTGTGCACTCCTAAAGCGATGATGTCTGCG
AGGGGTTCTTCGTTGAACCCTAGTTCAAGGCACCATCTTACCCTCTTATTTTTTCCCGTGGGCTT
TCATTTTGTGTCATCCGAGCATGACGTTGTAGGGTTGGAGTTTCTTCCTTTTTATCTTGTCATT
TACTGGTACCCATAGGCGCGAGACTAGGCTTCCATGTTTTGTTTTGCGACTTTCAAAAAGTACT
TTTAGTGGTTTGGGGCACGACGAGGGGGGGCAACCTCTTCTGTGCAAAAAGGTGGCTGGATGGA
TGAGATGAGATGAGATGAGGGTGAAGATAGATACCTGCAGTGTTTTTGACGCGACGGGATG

5/15

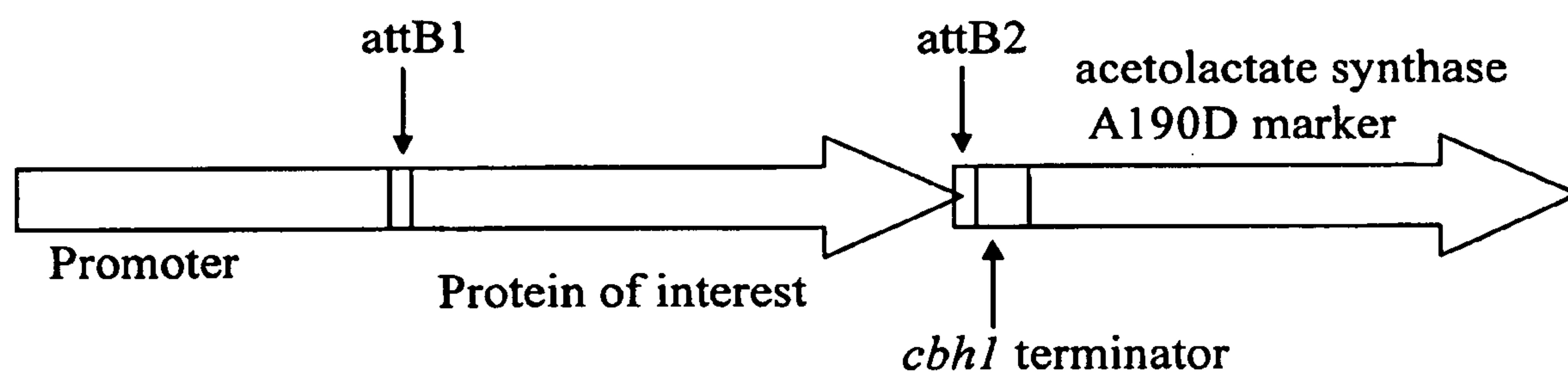


FIG. 4

6/15

Fig. 5A

contig1818 A version promoter (868 base pairs):

TAGGCTGGCGGGTTGGTGCTTTTGGCCCTTTGTCTGCCGCGACCTCTTGGTGACGCTGCATCGTGTTGGCTGTAGCTGCAC
GCCCCGGCCTCTTTTCTGGCATCTCGCACAGACTGTACGCAGAACATGTGTAAGTGACTTTCTATTGTAGGACTTGGATGG
ATGGGAGATGGCTATAGACAACAAGAGCGACAACAAGGATTATAGACAGTTCCATTACCTGGGCGACCACCATCGCACTT
CACCTTGCTCGCCCCGGCAGGACTGGCAGGCGTTGGCAATCTTGAGTCGCTCCCTGTACAGCGACAAGGGCGCCTGAGCCT
GCACCTGAGCCTGGGCCTGAGTCTGAGCCTGAAGCCGTTGATCCTCGTCTTGACGGGCGGACATGACGCTTTACGTGAGC
TGAGTCCAGGGCTGTCGTCCGCAAGTGATTCCGGCAGCCGGCAGGCACAGACGAAAGATAAGAAGCGGCTCGTTCACCAGA
TCCGAAGGCGGAATGCAAGAATGTCCGATGATGTCCGGGAAGAGAAGGCGATGCCTCGTCTATCTGGGAGGCGAGGGAAG
AAGGGAAAGCAACTCCCGCCTCGGCACAGCTACTCATTGGTCAGAGCGGACTTCGGAGTATAACCACACGGCTGACAAAC
CTGCATCGACCGTGATGGCGTTCTTTTGGTCCAATCATGGCTGAATGCGACGTTGGTTGTGTCCGCCCCCTGCAGCCAATG
AGAAGCTGGTCCGATCCTGGGGACGAGGGAGAGGCAAGCATATCGTATAAGTCCAAGCTCAAGCTTGGCGGCAGGTACCG
GTTGATGTGGTCGCTGTATATAAGTAAAAGAACTCGATTCCGCTTCGGAGTTTTCTCTGTCACCCTTC

Fig. 5B

contig1818 B version promoter (928 base pairs):

TAGGCTGGCGGGTTGGTGCTTTTGGCCCTTTGTCTGCCGCGACCTCTTGGTGACGCTGCATCGTGTTGGCTGTAGCTGCAC
GCCCCGGCCTCTTTTCTGGCATCTCGCACAGACTGTACGCAGAACATGTGTAAGTGACTTTCTATTGTAGGACTTGGATGG
ATGGGAGATGGCTATAGACAACAAGAGCGACAACAAGGATTATAGACAGTTCCATTACCTGGGCGACCACCATCGCACTT
CACCTTGCTCGCCCCGGCAGGACTGGCAGGCGTTGGCAATCTTGAGTCGCTCCCTGTACAGCGACAAGGGCGCCTGAGCCT
GCACCTGAGCCTGGGCCTGAGTCTGAGCCTGAAGCCGTTGATCCTCGTCTTGACGGGCGGACATGACGCTTTACGTGAGC
TGAGTCCAGGGCTGTCGTCCGCAAGTGATTCCGGCAGCCGGCAGGCACAGACGAAAGATAAGAAGCGGCTCGTTCACCAGA
TCCGAAGGCGGAATGCAAGAATGTCCGATGATGTCCGGGAAGAGAAGGCGATGCCTCGTCTATCTGGGAGGCGAGGGAAG
AAGGGAAAGCAACTCCCGCCTCGGCACAGCTACTCATTGGTCAGAGCGGACTTCGGAGTATAACCACACGGCTGACAAAC
CTGCATCGACCGTGATGGCGTTCTTTTGGTCCAATCATGGCTGAATGCGACGTTGGTTGTGTCCGCCCCCTGCAGCCAATG
AGAAGCTGGTCCGATCCTGGGGACGAGGGAGAGGCAAGCATATCGTATAAGTCCAAGCTCAAGCTTGGCGGCAGGTACCG
GTTGATGTGGTCGCTGTATATAAGTAAAAGAACTCGATTCCGCTTCGGAGTTTTCTCTGTCACCCTTCATGGACGAAGAG
CTATTGAGCCGCATCTTTGCTTCGTTGTCCAGCGTGAATTCTCCTACA

FIGS. 5A and 5B

7/15

AAGCTTACTAGTACTTCTCGAGCTCTGTACATGTCCGGTCGCGACGTACGCGTATCGATGGCGCCAGCT
GCAGGCGGCCGCTGCAGCCACTTGCAGTCCCGTGGAAATTCTCACGGTGAATGTAGGCCTTTTGTAGGG
TAGGAATTGTCACCTCAAGCACCCCAACCTCCATTACGCCTCCCCCATAGAGTTCCCAATCAGTGAGTC
ATGGCACTGTTCTCAAATAGATTGGGGAGAAGTTGACTTCCGCCAGAGCTGAAGGTCGCACAACCGCA
TGATATAGGGTCGGCAACGGCAAAAAAGCACGTGGCTCACCGAAAAGCAAGATGTTTGCATCTAACAT
CCAGGAACCTGGATACATCCATCATCACGCACGACCCTTTGATCTGCTGGTAAACTCGTATTCGCCCT
AAACCGAAGTGCGTGGTAAATCTACACGTGGGCCCTTTTCGGTATACTGCGTGTGTCTTCTCTAGGTGC
CATTTCTTTCCCTTCTCTAGTGTGAATTGTTTGTGTTGGAGTCCGAGCTGTAACCTCTGAATCT
CTGGAGAATGGTGGACTAACGACTACCGTGCACCTGCATCATGTATATAATAGTGATCCTGAGAAGGGG
GGTTTGGAGCAATGTGGGACTTTGATGGTCATCAAACAAAGAACGAAGACGCCTCTTTTGCAAAGTTT
GTTTCGGCTACGGTGAAGAACTGGATACTTGTGTGTCTTCTGTGTATTTTTGTGGCAACAAGAGGCCA
GAGACAATCTATTCAAACACCAAGCTTGCTCTTTTGAGCTACAAGAACCTGTGGGGTATATATCTAGAG
TTGTGAAGTCGGTAATCCCGCTGTATAGTAATACGAGTCGCATCTAAATACTCCGAAGCTGCTGCGAAC
CCGGAGAATCGAGATGTGCTGGAAAGCTTCTAGCGAGCGGCTAAATTAGCATGAAAGGCTATGAGAAAT
TCTGGAGACGGCTTGTGAATCATGGCGTTCCATTCTTCGACAAGCAAAGCGTTCCGTGCGAGTAGCAG
GCACTCATTCCCGAAAAAACTCGGAGATTCTTAAGTAGCGATGGAACCGGAATAATATAATAGGCAATA
CATTGAGTTGCCTCGACGGTTGCAATGCAGGGGTACTGAGCTTGGACATAACTGTTCCGTACCCACCT
CTTCTCAACCTTTGGCGTTTCCCTGATTACGCGTACCCGTACAAGTCGTAATCACTATTAACCCAGACT
GACCGGACGTGTTTGGCCCTTCATTTGGAGAAATAATGTCATTGCGATGTGTAATTTGCCTGCTTGACC
GACTGGGGCTGTTGGAAGCCCGAATGTAGGATTGTTATCCGAACCTCTGCTCGTAGAGGCATGTTGTGAA
TCTGTGTCGGGCAGGACACGCCTCGAAGGTTACGGCAAGGGAAACCACCGATAGCAGTGTCTAGTAGC
AACCTGTAAAGCCGCAATGCAGCATCACTGGAAAATACAAACCAATGGCTAAAAGTACATAAGTTAATG
CCTAAAGAAGTCATATAACAGCGGCTAATAATTGTACAATCAAGTGGCTAAACGTACCGTAATTTGCCA
ACGGCTTGTGGGGTTGCAGAAGCAACGGCAAAGCCCCACTTCCCCACGTTTGTCTTCACTCAGTCCA
ATCTCAGCTGGTGATCCCCCAATTGGGTGCTTGTGTTTCCGGTGAAGTGAAAGAAGACAGAGGTAAG
AATGTCTGACTCGGAGCGTTTGCATACAACCAAGGGCAGTGATGGAAGACAGTGAAATGTTGACATTC
AAGGAGTATTTAGCCAGGGATGCTTGAGTGTATCGTGTAAGGAGGTTTGTCTGCCGATACGACGAATAC
TGTATAGTCACTTCTGATGAAGTGGTCCATATTGAAATGTAAAGTCGGCACTGAACAGGCAAAAGATTG
AGTTGAAACTGCCTAAGATCTCGGGCCCTCGGGCCCTTCGGCTTTGGGTGTACATGTTTGTGCTCCGGG
CAAATGCAAAGTGTGGTAGGATCGAACACACTGCTGCCTTTACCAAGCAGCTGAGGGTATGTGATAGGC
AAATGTTTCAGGGGCCACTGCATGGTTTTCGAATAGAAAGAGAAGCTTAGCCAAGAACAAATAGCCGATAAA
GATAGCCTCATTAACCGAATGAGCTAGTAGGCAAAGTCAGCGAATGTGTATATATAAAGGTTTCGAGGT
CCGTGCCTCCCTCATGCTCTCCCCATCTACTCATCAACTCAGATCCTCCAGGAGACTTGTACACCATCT
TTTGAGGCACAGAAACCAATAGTCAACCATCACAAGTTTGTACAAAAAGCAGGCCTATGCACGTCCTG
TCGACTGCGGTGCTGCTCGGCTCCGTTGCCGTTCAAAGGTCCTGGGAAGACCAGGATCAAGCGGTCTG
TCCGACGTACCAAGAGGTCTGTTGACGACTTCATCAGCACCGAGACGCCTATTGCACTGAACAATCTT
CTTTGCAATGTTGGTCTTGATGGATGCCGTGCATTCCGGCACATCAGCTGGTGCGGTGATTGCATCTCCC
AGCACAATTGACCCGGACTGTAAGTTGGCCTTGATGAACCATATCATATATCGCCGAGAAGTGGACCGC
GTGCTGAGACTGAGACAGACTATTACATGTGGACGCGAGATAGCGCTCTTGTCTTCAAGAACCTCATCG
ACCGCTTCACCGAAACGTACGATGCGGGCCTGCAGCGCCGCATCGAGCAGTACATTACTGCCCAGGTCA
CTCTCCAGGGCCTCTCTAACCCTCGGGCTCCCTCGCGGACGGCTCTGGTCTCGGCGAGCCCAAGTTTG
AGTTGACCCTGAAGCCTTTACCGGCAACTGGGGTCGACCGCAGCGGGATGGCCCAGCTCTGCGAGCCA
TTGCCTTGATTGGATACTCAAAGTGGCTCATCAACAACAACTATCAGTCGACTGTGTCCAACGTCATCT
GGCCTATTGTGCGCAACGACCTCAACTATGTTGCCAGTACTGGTCAGTGCTTGCTTGCTCTTGAATTA
CGTCTTTGCTTGTGTGTCTAATGCCTCCACCACAGGAACCAACCGGCTTTGACCTCTGGGAAGAAGTC
AATGGGAGCTCATTCTTTACTGTTGCCAACACGACCGAGGTATGAAGCAAATCCTCGACATTCGCTGC
TACTGCACATG

8/15

AGCATTGTTACTGACCAGCTCTACAGCACTTGTGCGAGGGCGCCACTCTTGCTGCCACTCTTGGCCAGTC
GGGAAGCGCTTATTCATCTGTTGCTCCCCAGGTTTTGTGCTTTCTCCAACGATTCTGGGTGTCGTCTGG
TGGATACGTCGACTCCAACAGTATGTCTTTTCACTGTTTATATGAGATTGGCCAATACTGATAGCTCGC
CTCTAGTCAACACCAACGAGGGCAGGACTGGCAAGGATGTCAACTCCGTCTGACTTCCATCCACACCT
TCGATCCCAACCTTGGCTGTGACGCGAGGCACCTTCCAGCCATGCAGTGACAAAGCGCTCTCCAACCTCA
AGGTTGTTGTGCGACTCCTTCCGCTCCATCTACGGCGTGAACAAGGGCATTCTGCGCGGTGCTGCCGTGCG
CCATTGGCCCGGTATGCAGAGGATGTGTACTACAACGGCAACCCTTGGGTATCTTGCTACATTTGCTGCTG
CCGAGCAGCTGTACGATGCCATCTACGTCTGGAAGAAGACGGGCTCCATCACGGTGACCGCCACCTCCC
TGGCCTTCTTCCAGGAGCTTGTTCCTGGCGTGACGGCCGGGACCTACTCCAGCAGCTCTTCGACCTTTA
CCAACATCATCAACGCCGTCTCGACATACGCCGATGGCTTCTCAGCGAGGCTGCCAAGTACGTCCCCG
CCGACGGTTCGCTGGCCGAGCAGTTTGACCGCAACAGCGGCACTCCGCTGTCTGCGCTTCACCTGACGT
GGTCGTACGCCTCGTTCTTGACAGCCACGGCCCGTCCGGCTGGCATCGTGCCCCCTCGTGGGCCAACA
GCAGCGCTAGCACGATCCCCTCGACGTGCTCCGGCGCGTCCGTGGTCCGATCCTACTCGCGTCCCACCG
CCACGTCAATCCCTCCGTGCGAGACGCCCAAGCCTGGCGTGCTTCCGGTACTCCCTACACGCCCTGCG
CCTGCGCGACCCCAACCTCCGTGGCCGTACCTTCCACGAGCTCGTGTGACACAGTTTGGCCAGACGG
TCAAGGTGGCGGGCAACGCCGCGGCCCTGGGCAACTGGAGCACGAGCGCCCGCTGGCTCTGGACGCCG
TCAACTATGCCGATAACCAACCCCTGTGGATTGGGACGGTCAACCTCGAGGCTGGAGACGTGCTGGAGT
ACAAGTACATCAATGTGGGCCAAGATGGCTCCGTGACCTGGGAGAGTGATCCCAACCACACTTACACGG
TTCTGCGGTGGCTTGTGTGACGCGAGGTTGTCAAGGAGGACACCTGGCAGTCGTAAACCCAGCTTTCTT
GTACAAAGTGGTGATCGCGCCAGCTCCGTGCGAAAGCCTGACGACCCGGTAGATTCTTGGTGAGCCCGT
ATCATGACGGCGGCGGGAGCTACATGGCCCCGGGTGATTTATTTTTTTTGTATCTACTTCTGACCCTTT
TCAAATATACGGTCAACTCATCTTTCACTGGAGATGCGGCCTGCTTGGTATTGCGATGTTGTCAGCTTG
GCAAATTGTGGCTTTCGAAAACACAAAACGATTCCCTTAGTAGCCATGCATTTTAAGATAACGGAATAGA
AGAAAGAGGAAATTAAAAAACAACATCCCGTTCATAACCCGTAGAATCGCCGCTGGCG
CGCCTGAGACAATGGCCGGCAATGGTAAAAGGACCAAGATGTACTAGGTAGTTGCAATGTGGCTTATT
ACCTACCTACTACCTGGTAGGCACCTACTAGGTACTTGGGTAGACGGACAATGAAATTTGAAGTCGGGG
TTGCAGGAAAGCAGGGCGCTGGACACATTGTGCTTCAGGCGGTACCCGTGTCATCGTCAGCCAATGTC
GAGGCCCGGCAGCCCGAGGAGCGAGACAACCTTGGCCGGAGGAGCCCGCAGGTACCTGCCAAAGCGCGG
CTGGTACCTCTCAACCCTCTCAGGCCTGTTGGATGCCCTATGACATGCCCTGGGGGATGCAGCTGTTGC
CCCGGCCCGCACTTTTCGGGTGACCGCGAGGCTGCTGATTGGCTGGTTGCCACGGGCTGGGCGGTCCCT
GAAGTTGTTGCCATCTGAACTCTGTGCGCGCTGGCGTCCGCTGCGCCCAATGGGAGGCGAGACAACCTCA
GGGTACTAGAATCACTGACAGAAGAAGAGAATCGAAAGTAGGTAGACAGCCAATTCTGTTGCATGGCAGG
CAACCGCACAGGAGAAAAATTGACTACCCCACAATCAGGCACAGTAAGTAGGGCACAGTACGTATGTAC
AGACAAGGCGCAAGCGATACTGCGCGACCCGGTACCTCGCCGGCTTGACACGTGCGACAGGCTACTTTA
CTAGTATTCGCAGCGGCGGGTCGCGCATTATTACATGTACTGTGCCGCCATTTGATGACTGGGCTGCTG
CAGTATTAGTAGATCTGCCCGGCATCGCCCTTCCATGGGCGCGACCCGGGACTGGACCTCTGACTCTA
CCTACATGTACCTAGGCCGGGCGGGCTTGGTGACTTTTGTCCGATCAGGTGTTTCGCTGGCTACCTA
TTATTTCTCTTTCTTCTTCTCCATCCTGCTTCTGGCCTTGCAATTCTTCTTCGCCACTCCTCCCTCTTC
CCCCGCGATACCTTTGAATTCGTGAGAGGAAAAGACGAGAAAAAAGGGCAGCAGAGACGTCCGT
CTGGCTCACGTGCTGCATCTCTGCGCACTCTCATTTTCTTTTATTGTCCGACCCCTCCCTCAACCTTCTC
CTTCGTTGACAGGCTAAGCCTTGCTTCGACGCTCTCTCTTTGAATTTTCTACTTCTACCTTCTTTTCT
TGCGTGTTACCCACCATAGCTCGATTACAGATGCTCCGAAGTCGCCAAGTCACAGCCAGGGCCGTCCGG
GCTCTGGGCCAGGCGCGCGCCTTTACCTCGACGACCAAGCCTGTGATGATCCAGAGCAGCCAGAGGAAA
CAGGCCAACGCCAGCGCTGCTCCGTAAGTCGCCCATTGCCATTGCATCTTCTGTTTGATATATACTTCC
TGCTGCTTGCGTGGCGTCG

9/15

TCTCTCGGTTATGCGTGTCAAGGACCAGGTGTGTTTCGCATCGTGGTTTTCCAGCGCCGATTACCGGG
GGACGAATTTTTGGCTGCTCAACTCGCGCGCGCGCATTCTGATTCTTCGTTTTCAATCTTGAGCGAC
AACTGGCTAACATAATGGCCATTGGCAATTGCTTTCACACAGACAAGTCCGCCCTGTACCGAGCCCTG
CTTTCAACGCTGAAGACAAAGACCGCAGCCATGTGCAGCCTCTGGTCAACCCGTGGAAGCCCGACAT
GGATGAATCGTATGTCCACGTCCCCTCGTCCCGCCCTACAAAATGAACACGATTACACCAGAATTTT
TGCAACAATCGACACTTCTATAACAGACCAATTGAGCTTTGTTCTGACCAATCATGTTGCTCTAGAT
TCATTGGCAAAACCGGAGGCGAAATCTTCCACGAGATGATGCTGCGACAGGGTGTCAAGCACATTTG
TAGGTTCCGATGCCGGCCGCCACACGGGCTCCATCCTTGCTCCATCTCTCCAGCTAGGCAAATCTC
GCTAACCTTGAGTCACCATCCAGTCGGATACCCTGGCGGCGCTATCCTGCCCGTCTTCGACGCCATC
TACAACTCAAAACACTTCGACTTCATCCTGCCCCGTGATGAGCAGGGAGCTGGCCATATGGCCGAGG
GCTATGCCCGTGCTCGGGCAAACCCGGTGTGCTCCTGGTGACTTCCGGCCCCGGTGCTACCAATGT
CATCACGCCCATGCAGGATGCCCTGTGCGACGGAACGCCCTTGCTCGTCTTCTGCGGCCAGGTCCCC
ACCACGGCCATCGGCAGCGATGACTTCCAAGAGGCCGACGTGCTGGGCATCTCGCGGGCCTGCACCA
AGTGGAACGTCATGGTCAAGAGCGTTGCTGAGCTGCCGCGGAGAATCAACGAGGCCTTTGAGATTGC
CACCAGCGGCCGCCCTGGCCCCGTCTCGTCGACCTGCCCAAGGATGTACGGCTGGTATCCTGAGG
AGAGCCATCCCTACGGAGACTGCTCTGCCGTCTCTGCCCAGTGCCGCTCCCGCGCCGCCATGGAGC
TGAGCTCCAAGCAGCTCAACGCCTCCATCAAGCGTGCCGCCGACCTCATCAACATCGCCAAGAAGCC
CGTCATCTACGCCGGTCAGGGTGTGTCATCCAGTCCGAGGGCGGCGTTGAGCTCCTGAAGCAGCTGGCG
GACAAGGCCTCCATCCCCGTCAACACCACCCTCCATGGCCTGGGTGCCTTTGATGAGCTGGACGAGA
AGTCGCTGCACATGCTGGGCATGCACGGCTCGGCGTATGCCAACATGGCCATGCAGCAGGCCGACCT
CATCATCGCCCTCGGCAGCCGATTCGACGACCGTGTTACTCTGAATGTCTCCAAATTTGCGCCTGCA
GCCAGGCAAGCTGCTGCCGAGGGCCGCGCGGCATCATTCACTTTGAGATCATGCCCAAGAACATCA
ACAAGGTGATCCAGGCGACCGAGGCCGTCGAGGGCGACGTCGCCACCAACCTGAAGCACCTCATTC
CCAGATTGCCGAAAAGTCCATGGCGGACCGAGGAGAGTGGTTCGGCCTCATCAATGAGTGGAAGAAG
AAGTGCCCCCTGTCAAATAACAGCGCGCGGAGCGGGCTGGCCTCATCAAGCCGCAGACGGTCATGG
AGGAGATTAGCAACCTGACGGCCAACCGAAAGGACAAGACGTACATTGCCACGGGTGTGCGCCAGCA
CCAGATGTGGGTGCCCAGCACTTCCGCTGGAGGCACCCTCGATCCATGATTACCTCTGGTGGTCTG
GGCACCATGGGCTACGGTCTCCCCGCGGCCATTGGCGCCAAGGTGGCCCAGCCCCGACGCTCTCGTAA
TTGACGTTGATGGCGATGCCTCGTTTAAACATGACGCTGACGGAGCTGTGACTGCTGCACAGTTCAA
CATTGGCGTCAAGGTGGTTGTGCTCAACAACGAGGAGCAGGGCATGGTGACGCAGTGGCAGAACCTC
TTTTACGAGGACCGATATGCCACACGCACCAGAAGAACCCCGACTTCATGAAGCTGGCCGACGCCA
TGGGCGTTTACGACACAGCGCGTGACGGAGCCGAGAGAAGCTGGTCGATGCCCTGACGTGGCTGATCAA
CACCGATGGCCCGGCCCTGTTGGAGGTTGTACGGACAAGAAGGTGCCTGTCTGCCCATGGTGCCC
GCCGGATCGGCCCTGCACGAGTTCCTCGTCTTTGAACCTGGTGAGTCTACTTCAGACATATTGCTTG
CGCATTGCAGATACTAACACTCTCACAGAAAAGGATAAGCAGCGCCGTGAGCTGATGAAGGAGAGAA
CAAAGGGTGTGCACTCCTAAAGCGATGATGTCTGCGAGGGGTTCTTCGTTGAACCCTAGTTTCAAGCA
CCATCTTACCCTCTTATTTTTTCCCGTGGGCTTTTCATTTTGTGTGTCATCCGAGCATGACGTTGTAGGG
TTGGAGTTTCTTCTTTTATCTTGTGTCATTTACTGGTACCCATAGGCGCGAGACTAGGCTTCCATGT
TTTGTTTTGCGACTTTCAAAAAGTACTTTTAGTGGTTTGGGGCACGACGAGGGGGGGCAACCTCTTC
TGTCGAAAAAGGTGGCTGGATGGATGAGATGAGATGAGATGAGGGTGAAGATAGATACCTGCAGTGT
TTTTGACGCGACGGGATGGCGATCGCTCAGGGTTGCGTTTCCAGTCTAGACACGTATAACGGCACAA
GTGTCTCTCACCAAATGGGTTATATCTCAAATGTGATCTAAGGATGGAAAGCCCAGAATATCGATCG
CGCGCAGATCCATATATAGGGCCCCGGGTTATAATTACCTCAGGTGACGTCCTCATGGCCATTCGAAT
TCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAG

10/15

.TGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAG
CTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCTCTTCCGCTTCCTC
GCTCACTGACTCGCTGCGCTCGGTCTGCTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAA
TACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAA
AAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACCAGGCGTTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCT
TCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTACGCCGACCGCTGCGCCTTATCCGGTAACTATCGTC
TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAA
GGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTA
AGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTT
TTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCA
CCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC
CAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCC
GCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTTCGCCAGTTAATAGTTTGCG
CAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTCAGCT
CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTC
GGTCCTCCGATCGTTGTGCAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCA
TAATTCTCTTACTGTGATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCA
CATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTT
ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT
TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
CGGAAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCT
CATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCACATTTCCCC
GAAAAGTGCCACCTGACGTCTAAGAAACCATTTATTATCATGACATTAACCTATAAAAATAGGCGTATC
ACGAGGCCCTTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGA
GACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG
TTGGCGGGTGTCGGGGCTGGCTTAACTATGCCGCATCAGAGCAGATTGTAAGTACTGAGAGTGACCATAAA
ATTGTAAACGTTAATATTTTGTAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTTTAACCA
ATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTT
CAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTAT
CAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGC
ACTAAATCGGAACCTTAAAGGGAGCCCCCGATTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGA
GAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGC
GTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTACTATGGTTGCTTTGACGTATG
CGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGCCATTTCGCCATTACGGCT
GCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGAT
GTGCTGCAAGGCGATTAAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCC
AGTGCCC

11/15

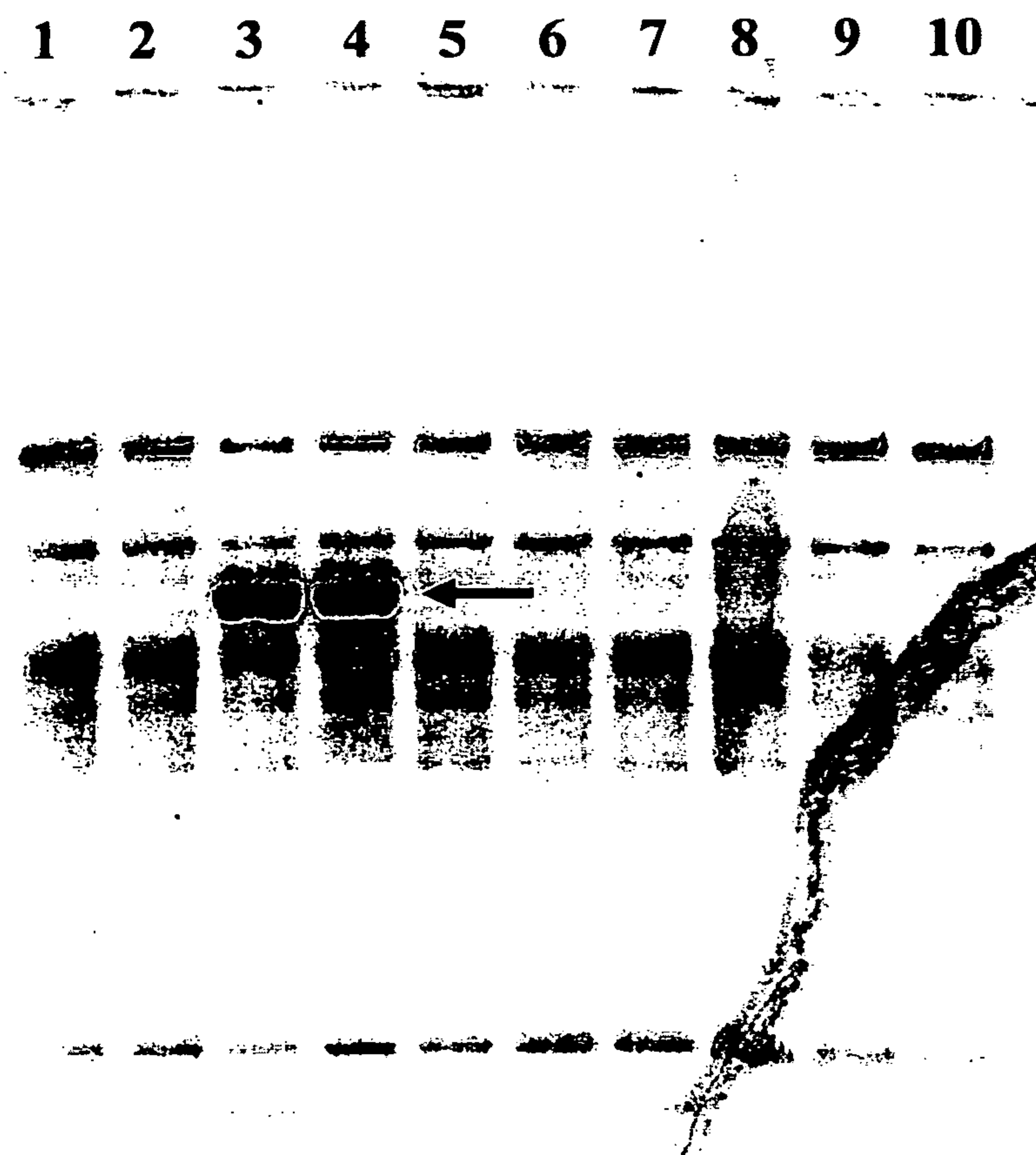


Fig. 7

12/15

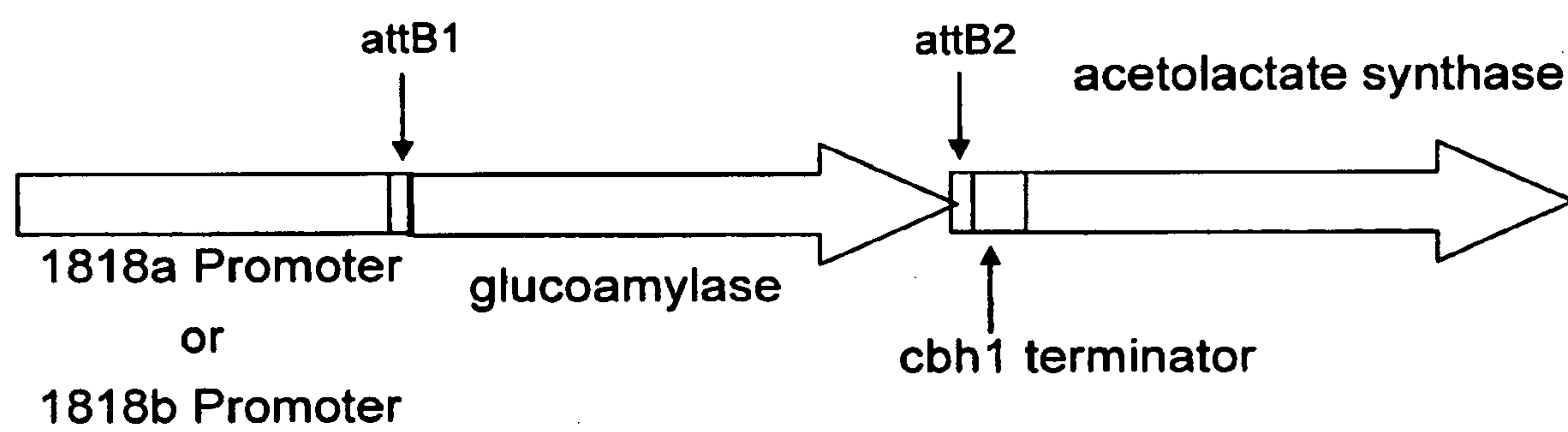


Fig. 8

13/15

CATTACGCCTCCCCCATAGATAGGTCTAGGCTGGCGGGTTGGTGCTTTTGCCCTTTGTCTGCCGCGACC
TCTTGGTGACGCTGCATCGTGTGGCTGTAGCTGCACGCCCCGGCCTCTTTTCTGGCATCTCGCACAGAC
TGTAACGAGAACATGTGTAAGTGACTTTCTATTGTAGGACTTGGATGGATGGGAGATGGCTATAGACAA
CAAGAGCGACAACAAGGATTATAGACAGTTCCATTACCTGGGCGACCACCATCGCACTTCACCTTGCTC
GCCCCGCAGGACTGGCAGGCGTTGGCAATCTTGAGTCGCTCCCTGTACAGCGACAAGGGCGCCTGAGCC
TGCACCTGAGCCTGGGCCTGAGTCTGAGCCTGAAGCCGTTGATCCTCGTCTTGACGGGCGGACATGACG
CTTTACGTGAGCTGAGTCCAGGGCTGTCTGTCGGCAAGTGATTCGGCAGCCGGCAGGCACAGACGAAAGA
TAAGAAGCGGCTCGTTCACCAGATCCGAAGGCGGAATGCAAGAATGTCCGATGATGTCGGGGAAGAGAA
GGCGATGCCTCGTCTATCTGGGAGGCGAGGGAAGAAGGGAAGCAACTCCCGCCTCGGCACAGCTACTC
ATTGGTCAGAGCGGACTTCGGAGTATAACCACACGGCTGACAAACCTGCATCGACCGTGATGGCGTTCT
TTTGGTCCAATCATGGCTGAATGCGACGTTGGTTGTGTCTCGCCCCCTGCAGCCAATGAGAAGCTGGTCCG
ATCCTGGGGACGAGGGAGAGGCAAGCATATCGTATAAGTCCAAGCTCAAGCTTGGCGGCAGGTACCGGT
TGATGTGGTCGCTGTATATAAGTAAAAGAACTCGATTCCGCTTCGGAGTTTTCTCTGTCAACCTTCACA
AGTTTGTACAAAAAGCAGGCTATGCACGTCCTGTCTGACTGCGGTGCTGCTCGGCTCCGTTGCCGTTCA
AAAGGTCCTGGGAAGACCAGGATCAAGCGGTCTGTCCGACGTCACCAAGAGGTCTGTTGACGACTTCAT
CAGCACCGAGACGCCTATTGCACTGAACAATCTTCTTTGCAATGTTGGTCCTGATGGATGCCGTGCATT
CGGCACATCAGCTGGTGCGGTGATTGCATCTCCAGCACAATTGACCCGGACTGTAAGTTGGCCTTGAT
GAACCATATCATATATCGCCGAGAAGTGGACCGCGTGCTGAGACTGAGACAGACTATTACATGTGGACG
CGAGATAGCGCTCTTGTCTTCAAGAACCTCATCGACCGCTTCACCGAAACGTACGATGCGGGCCTGCAG
CGCCGCATCGAGCAGTACATTACTGCCCAGGTCACCTCTCCAGGGCCTCTCTAACCCTCGGGCTCCCTC
GCGGACGGCTCTGGTCTCGGCGAGCCCAAGTTTGAGTTGACCTGAAGCCTTTCACCGGCAACTGGGGT
CGACCGCAGCGGGATGGCCAGCTCTGCGAGCCATTGCCTTGATTGGATACTCAAAGTGGCTCATCAAC
AACAACATATCAGTCGACTGTGTCCAACGTCTGTCCTATTGTGCGCAACGACCTCAACTATGTTGCC
CAGTACTGGTCAGTGCTTGCTTGCTCTTGAATTACGTCTTTGCTTGTTGTGTCTAATGCCTCCACCACAG
GAACCAAACCGGCTTTGACCTCTGGGAAGAAGTCAATGGGAGCTCATTCTTTACTGTTGCCAACCAGCA
CCGAGGTATGAAGCAAATCCTCGACATTCGCTGCTACTGCACATGAGCATTGTTACTGACCAGCTCTAC
AGCACTTGTCGAGGGCGCCACTCTTGCTGCCACTCTTGCCAGTCGGGAAGCGCTTATTCATCTGTTGC
TCCCCAGGTTTTTGTGCTTTCTCCAACGATTCCTGGGTGTCGTCCTGGTGGATACGTCGACTCCAACAGTAT
GTCTTTTCACTGTTTATATGAGATTGGCCAATACTGATAGCTCGCCTCTAGTCAACACCAACGAGGGCA
GGACTGGCAAGGATGTCAACTCCGTCTGACTTCCATCCACACCTTCGATCCCAACCTTGGCTGTGACG
CAGGCACCTTCCAGCCATGCAGTGACAAAGCGCTCTCCAACCTCAAGGTTGTTGTGCTGACTCCTTCCGCT
CCATCTACGGCGTGAACAAGGGCATTCCCTGCCGGTGCTGCCGTGCGCATTGGCCGGTATGCAGAGGATG
TGTAATAACAACGGCAACCCTTGGTATCTTGCTACATTTGCTGCTGCCGAGCAGCTGTACGATGCCATCT
ACGTCTGGAAGAAGACGGGCTCCATCACGGTGACCGCCACCTCCCTGGCCTTCTTCCAGGAGCTTGTTT
CTGGCGTGACGGCCGGGACCTACTCCAGCAGCTCTTCGACCTTTACCAACATCATCAACGCCGTCTCGA
CATACGCCGATGGCTTCTCAGCGAGGCTGCCAAGTACGTCCCCGCCGACGGTTCGCTGGCCGAGCAGT
TTGACCGCAACAGCGGCACTCCGCTGTCTGCGCTTCACCTGACGTGGTCGTACGCCTCGTTCTTGACAG
CCACGGCCCCGTCGGGCTGGCATCGTGCCCCCTCGTGGGCCAACAGCAGCGCTAGCACGATCCCTCGA
CGTGCTCCGGCGCGTCCGTGGTCCGATCCTACTCGCGTCCCACCGCCACGTCATTCCCTCCGTGCGAGA
CGCCCAAGCCTGGCGTGCTTCCGGTACTCCCTACACGCCCTGCCCTGCGCGACCCCAACCTCCGTGG
CCGTACCTTCCACGAGCTCGTGTGACACAGTTTGGCCAGACGGTCAAGGTGGCGGGCAACGCCGCGG
CCCTGGGCAACTGGAGCACGAGCGCCGCGGTGGCTCTGGACGCCGTCAACTATGCCGATAACCACCCCC
TGTGGATTGGGACGGTCAACCTCGAGGCTGGAGACGTCTG

Fig. 9

14/15

TGGAGTACAAGTACATCAATGTGGGCCAAGATGGCTCCGTGACCTGGGAGAGTGATCCCAACCACACT
TACACGGTTCCCTGCGGTGGCTTGTGTGACGCAGGTTGTCAAGGAGGACACCTGGCAGTCGTAAACCCA
GCTTTCTTGTACAAAGTGGTGATCGCGCCAGCTCCGTGCGAAAGCCTGACGCACCGGTAGATTCTTGG
TGAGCCCGTATCATGACGGCGGCGGGAGCTACATGGCCCCGGGTGATTTATTTTTTTTGTATCTACTT
CTGACCCTTTTCAAATATACGGTCAACTCATCTTCACTGGAGATGCGGCCTGCTTGGTATTGCGATG
TTGTCAGCTTGGCAAATTGTGGCTTTCGAAAACACAAAACGATTCCCTTAGTAGCCATGCATTTTAAGA
TAACGGAATAGAAGAAAGAGGAAATTAAAAAACAACATCCCGTTCATAACCCGTAGA
ATCGCCGCTGGCGCGCCTGAGACAATGGCCGGCAATGGTAAAAAGGACCAAGATGTACTAGGTAGTTG
CAATGTGGCTTATTACCTACCTACTACCTGGTAGGCACCTACTAGGTACTTGGGTAGACGGACAATGA
AATTTGAAGTCGGGGTTGCAGGAAAGCAGGGCGCTGGACACATTGTGCTTCAGGCGGTACCCGTGCTC
ATCGTCAGCCAATGTCGAGGCCCGGCAGCCCGAGGAGCGAGACAACCTTGGCCGGAGGAGCCCGCAGG
TACCTGCCAAAGCGCGGCTGGTACCTCTCAACCCTCTCAGGCCTGTTGGATGCCCTATGACATGCCCT
GGGGGATGCAGCTGTTGCCCGGCCCGCACCTTTCGGGTGACCGCGAGGCTGCTGATTGGCTGGTTGC
CACGGGCTGGGCGGTCCCTGAAGTTGTTGCCATCTGAACTCTGTGCGCGCTGGCGTCGGCTGCGCCCA
ATGGGAGGCGAGACAACCTCAGGGTACTAGAATCACTGACAGAAGAAGAGAATCGAAAGTAGGTAGACA
GCCAATTCTGTTGCATGGCAGGCAACCGCACAGGAGAAAAATTGACTACCCACAATCAGGCACAGTAA
GTAGGGCACAGTACGTATGTACAGACAAGGCGCAAGCGATACTGCGCGACCCGGTACCTCGCCGGCTT
GACACGTGCGACAGGCTACTTTACTAGTATTCGCAGCGGCGGGTTCGCGCATTATTACATGTACTGTGC
CGCCATTTGATGACTGGGCTGCTGCAGTATTAGTAGATCTGCCCCGCATCGCCCTTCCATGGGCGCGA
CCCGGGACTGGACCCTCTGACTCTACCTACATGTACCTAGGCCGGGCGGGCTTGGTGACTTTTGTCC
GATCAGGTCGTTTCGCCTGGCTACCTATTATTTCTTTCTTCTTCTCCATCCTGCTTCTGGCCTTGCA
ATTCTTCTTCGCCACTCCTCCCTCTTCCCCCGCGATAACCCTTGAATTCGTGAGAGAGGAAAAGACGA
GAAAAAAGGGGCAGCAGAGACGTGGTCTGGCTCACGTGCTGCATCTCTGCGCACTCTCATTTTTTT
TATTGTCCGACCCCTCCCTCAACCTTCTCCTTCGTTGACAGGCTAAGCCTTGCTTCGACGCTCTCTCT
TTGAATTTTTCTACTTCTACCTTCTTTTCTTGCGTGTACCCACCATAGCTCGATTCACGATGCTCCG
AAGTCGCCAAGTCACAGCCAGGGCCGTCCGGGCTCTGGGCCAGGCGCGCGCTTTACCTCGACGACCA
AGCCTGTGATGATCCAGAGCAGCCAGAGGAAACAGGCCAACGCCAGCGCTGCTCCGTAAGTCGCCCAT
TGCCATTGCATCTTCTGTTTGATATATACTTCCTGCTGCTTGCGTGGCGTCGTCTCTCGGTTATGCGT
GTCAAGGACCAGGTGTGTTTCGCATCGTGGTTTTCCAGCGCCGATTACCGGGGGACGAATTTTTGGCTG
CTCAACTCGCGCGCGCGCATCTGATTCTTCGTTTTCAATCTTGAGCGACAACCTGGCTAACATAATGG
CCATTGGCAATTGCTTCACACAGACAAGTCCGCCCTGTACCGAGCCCTGCTTTCAACGCTGAAGACAA
AGACCGCAGCCATGTGCGAGCCTCTGGTCAACCCGTCGAAGCCCGACATGGATGAATCGTATGTCCACG
TCCCCTCGTCCCGCCCTACAAAATGAACACGATTACACCAGAATTTTTGCAACAATCGACACTTCTAT
AACAGACCAATTGAGCTT

Fig. 9

15/15

TGTTCTGACCAATCATGTTGCTCTAGATTTCATTGGCAAACCGGAGGCGAAATCTTCCACGAGATGA
TGCTGCGACAGGGTGTCAAGCACATTTGTAGGTTCCGATGCCGGCCGCCACACGGGCTCCATCCTT
GCTCCATCTCTCCAGCTAGGCAAATCTCGCTAACCTTGAGTCACCATCCAGTCGGATACCCTGGCGG
CGCTATCCTGCCCCGTCTTCGACGCCATCTACAACTCAAACACTTCGACTTCATCCTGCCCCGT
GAGCAGGGAGCTGGCCATATGGCCGAGGGCTATGCCCGTGCCTCGGGCAAACCCGGTGTCTGCTCCTGG
TGACTTCCGGCCCCGGTGCTACCAATGTCATCAGCCCATGCAGGATGCCCTGTGCGACGGAACGCC
CTTGGTCTGTCTTCTGCGGCCAGGTCCCCACCACGGCCATCGGCAGCGATGACTTCCAAGAGGCCGAC
GTCGTGGGCATCTCGCGGGCCTGCACCAAGTGGAACGTCATGGTCAAGAGCGTTGCTGAGCTGCCGC
GGAGAATCAACGAGGCCCTTTGAGATTGCCACCAGCGGCCGCCCTGGCCCCGTCTCTGTCGACCTGCC
CAAGGATGTCACGGCTGGTATCCTGAGGAGAGCCATCCCTACGGAGACTGCTCTGCCGTCTCTGCCC
AGTGCCGCCTCCCGCGCCGCCATGGAGCTGAGCTCCAAGCAGCTCAACGCCTCCATCAAGCGTGCCG
CCGACCTCATCAACATCGCCAAGAAGCCCGTCATCTACGCCGGTCAGGGTGTATCCAGTCCGAGGG
CGGCGTTGAGCTCCTGAAGCAGCTGGCGGACAAGGCCTCCATCCCCGTCAACCACCACCCTCCATGGC
CTGGGTGCCTTTGATGAGCTGGACGAGAAGTCGCTGCACATGCTGGGCATGCACGGCTCGGCGTATG
CCAACATGGCCATGCAGCAGGCCGACCTCATCATCGCCCTCGGCAGCCGATTTCGACGACCGTGTTAC
TCTGAATGTCTCAAATTTGCGCCTGCAGCCAGGCAAGCTGCTGCCGAGGGCCGCGGCGGCATCAT
CACTTTGAGATCATGCCCAAGAACATCAACAAGGTCATCCAGGCGACCGAGGCCGTGAGGGCGACG
TCGCCACCAACCTGAAGCACCTCATTTCCCAGATTGCCGAAAAGTCCATGGCGGACCGAGGAGAGTG
GTTCCGGCCTCATCAATGAGTGGAAGAAGAAGTGGCCCTGTCAAACCTACCAGCGCGCGGAGCGGGCT
GGCCTCATCAAGCCGCGAGACGGTCATGGAGGAGATTAGCAACCTGACGGCCAACCGAAAGGACAAGA
CGTACATTGCCACGGGTGTGCGGCCAGCACCAGATGTGGGTTGCCAGCACTTCCGCTGGAGGCACCC
TCGATCCATGATTACCTCTGGTGGTCTGGGCACCATGGGCTACGGTCTCCCCGCGGCCATTGGCGCC
AAGGTGGCCCAGCCCCGACGCTCTCGTAATTGACGTTGATGGCGATGCCTCGTTTAACATGACGCTGA
CGGAGCTGTCGACTGCTGCACAGTTCAACATTGGCGTCAAGGTGGTTGTGCTCAACAACGAGGAGCA
GGGCATGGTGACGCAGTGGCAGAACCCTCTTTTACGAGGACCGATATGCCACACGCACCAGAAGAAC
CCCGACTTCATGAAGCTGGCCGACGCCATGGGCGTTTCAACACCAGCGCGTGACGGAGCCGGAGAAGC
TGGTCGATGCCCTGACGTGGCTGATCAACACCGATGGCCCGGCCCTGTTGGAGGTTGTCACGGACAA
GAAGGTGCCTGTCTGCCCATGGTGCCCGCCGGATCGGCCCTGCACGAGTTCCTCGTCTTTGAACCT
GGTGAGTCTACTTCAGACATAATTGCTTGCGCATTGCAGATACTAACACTCTCACAGAAAAGGATAAG
CAGCGCCGTGAGCTGATGAAGGAGAGAACAAAGGGTGTGCACTCCTAAAGCGATGATGTCTGCGAGG
GGTCTTTCGTTGAACCCTAGTTTCAAGGCACCATCTTACCCTCTTATTTTTTCCCGTGGGCTTTTCA
TGTGTCATCCGAGCATGACGTTGTAGGGTTGGAGTTTCTTCTTTTATCTTGTCAATTTACTGGTAC
CCATAGGCGCGAGACTAGGCTTCCATGTTTTGTTTTGCGACTTTCAAAAAGTACTTTTAGTGGTTTG
GGGCACGACGAGGGGGGGCAACCTCTTCTGTGAAAAAGGTGGCTGGATGGATGAGATGAGATGAGA
TGAGGGTGAAGATAGATACCTGCAGTGTTTTTGACGCGACGGGATGGCGATCGC

Fig. 9

