

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

30 July 2020 (30.07.2020)



(10) International Publication Number

WO 2020/152466 A1

(51) International Patent Classification:

C12N 15/82 (2006.01) A01H 6/82 (2018.01)

A01H 5/12 (2018.01)

Published:

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

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

(21) International Application Number:

PCT/GB2020/050153

(22) International Filing Date:

23 January 2020 (23.01.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1900940.6 23 January 2019 (23.01.2019) GB

(71) Applicant: **BRITISH AMERICAN TOBACCO (INVESTMENTS) LIMITED** [—/GB]; Globe House, 1 Water Street, London WC2R 3LA (GB).

(72) Inventors: **BEN KHALED**, Sara; c/o British American Tobacco, (Investments) Limited, Globe House, 1 Water Street, London WC2R 3LA (GB). **ANASTACIO DE ABREU E LIMA**, Francisco; c/o British American Tobacco, (Investments) Limited, Globe House, 1 Water Street, London WC2R 3LA (GB).

(74) Agent: **D YOUNG & CO LLP**; 120 Holborn, London EC1N 2DY (GB).

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

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

(54) Title: METHOD FOR DECREASING THE ALKALOID CONTENT OF A TOBACCO PLANT

(57) Abstract: The present invention provides a method for modulating (e.g. decreasing) the alkaloid content of a plant (e.g. a tobacco plant), the method comprising modifying said plant by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein. The present invention also provides for the use of at least one gene encoding a SOUL haem-binding protein for modulating the alkaloid content of a plant, as well as tobacco cells, plants, plant propagation materials, harvested leaves, processed tobaccos, or tobacco products obtainable in accordance with the invention.



WO 2020/152466 A1

METHOD FOR DECREASING THE ALKALOID CONTENT OF A TOBACCO PLANT

FIELD OF THE INVENTION

The present invention relates to methods of modulating the alkaloid content e.g. nicotine content of a plant or part thereof. The invention also extends to methods of modulating the expression and/or activity of polypeptides which modulate alkaloid content within plants. Alternatively, the invention provides methods of modulating the expression and/or activity of genes which encode polypeptides which modulate alkaloid content within plants. The invention also extends to constructs, which can be used to modulate the polypeptides. The invention further relates to plant cells and plants modified to achieve a modulation in alkaloid content. The invention also relates to a processed and harvested leaf from such modulated plants and use thereof in a tobacco industry product, including combustible smoking articles.

BACKGROUND

Alkaloids are a group of naturally occurring compounds which mostly contain basic nitrogen atoms and are produced by a large variety of organisms including bacteria, fungi, plants and animals.

Alkaloids may be classified according to the similarity of the carbon skeleton e.g. indole-, isoquinoline- and pyridine-like. Pyridine derivatives are one class of monomeric alkaloids; this class includes simple derivatives of pyridine, polycyclic condensed and noncondensing pyridine derivatives and sesquiterpene pyridine derivatives. Examples are nicotine, nornicotine, anabasine, myosmine and anatabine.

Most of the known biological functions of alkaloids are related to protection. Neuroactive molecules, like caffeine, cocaine, morphine, and nicotine, act as defence compounds against invading predators. The accumulation of these alkaloids is the result of signal transduction cascades that monitor gene expression, enzyme activities, and alkaloid concentrations. The fine-tuning of alkaloid content in the plant involves negative feedback loops and degradative pathways.

Nicotine occurs naturally in several varieties of plants but is found at the highest level in the tobacco plant. Cultivated tobacco produces 2-4% alkaloids of total dry weight. Nicotine is produced in wild and cultivated *Nicotiana* species and plays an important role in plant defence against herbivores and insects (Voelckel et al. (2001) *Oecologia* 127(2): 274-280, incorporated herein by reference). It accounts for ~90% of the total alkaloid content. The remaining 10% of the alkaloid pool is mostly constituted by the structurally related compounds nornicotine, anatabine, anabasine and pseudoxyntine (PON).

The regulation of alkaloid content in tobacco is complex. Several factors including genotype, environment, fertilization and agronomic practices (e.g. topping) affect alkaloid levels in tobacco

plants. Some key regulators of nicotine biosynthesis are well characterized, for example putrescine N-methyltransferase (PMT), which plays a pivotal role in this pathway, is activated by members of the ethylene responsive factor (ERF) superfamily, the largest transcription factor family in the tobacco genome (Rushton et al. (2008) *Plant Physiol.* 147(1): 280-295 incorporated
 5 herein by reference). Other transcription factors that induce alkaloid biosynthesis belong to the MYC2-like basic helix-loop-helix (bHLH) family. MYC2-like bHLHs regulate alkaloid levels directly, through the Gbox-mediated binding and activation of alkaloid structural genes, and indirectly, through the activation of ERFs.

Modifying alkaloid content in plants (e.g. tobacco) can have several commercial advantages.

10 For example, decreasing total alkaloid content in plants can increase the value of said plant as a biomass resource. For example, modifying alkaloid content may comprise reducing the alkaloid content e.g. nicotine content of tobacco plants. Tobacco plants and products with reduced nicotine may be desirable in view of the potential regulation of "nicotine ceilings" i.e. average upper limits of nicotine in tobacco products. Alternatively, increasing alkaloid content in plants
 15 e.g. tobacco plants, can help to protect plants against insects and herbivores. There remains a need for plants with modulated alkaloid content, for example with modulated nicotine content, with improved commercially desirable traits and methods for making the same.

During the post-harvest leaf curing, reactions between pyridine alkaloids and nitrosating species leads to the formation of tobacco-specific nitrosamines (TSNAs). Nornicotine and PON are the
 20 precursors of the TSNAs NNN and NNK respectively. Reducing the production and accumulation of nornicotine and PON is of high importance. The CYP82E family of nicotine demethylase genes is one of the primary regulators of nicotine to nornicotine conversion, and altering their activity or accumulation may result in a decrease in NNN levels. However, no enzymes or genes responsible for producing PON have been identified thus far.

25 As described in the Examples, the inventors sought to investigate genes responsible for alkaloid synthesis, with the aim of modulating alkaloid content in plants, e.g. decreasing nicotine content in tobacco plants.

SUMMARY OF THE INVENTION

30 It has been surprisingly found that by modulating the activity or expression of a gene encoding a SOUL haem-binding protein the alkaloid content and/or TSNA content or precursor of TSNA content of plants can be modulated. The SOUL haem-binding protein(s) as taught herein, for example Nitab4.5_0013616g0010.2, is a regulator of nicotine, nornicotine, anabasine, PON, myosmine and anatabine in cultivated tobacco. Nitab4.5_0013616g0010.2 encodes a SOUL
 35 haem-binding protein according to the present invention. Nitab4.5_0000652g0130.2, Nitab4.5_0001140g0220.2, Nitab4.5_0003235g0060.2, Nitab4.5_0004868g0020.2,

Nitab4.5_0006614g0010.2, Nitab4.5_0006991g0050.2 and Nitab4.5_0009023g0010.2 encode homologues of Nitab4.5_0013616g0010.2 according to the present invention. SOUL haem-binding proteins according to the present invention contain a conserved three-dimensional structure termed a SOUL motif. SOUL haem-binding proteins according the present invention
5 bind to haem.

According to the present invention, tobacco products with modulated alkaloid content and commercially desirable traits sought after by consumers of tobacco products can be produced. In some instances, consumers may desire a product with low levels of alkaloid content e.g. low levels of TSNAs or low nicotine content.

10 The present invention may be particularly useful in the field of plant molecular farming, where plants (such as tobacco and other *Nicotiana* spp.) are used for the production of proteins, peptides, and metabolites e.g. for the production of therapeutics and pharmaceuticals such as antibiotics, virus like particles, or neutraceuticals or small molecules. Tobacco has been used for the development of an HIV-neutralising antibody in an EU-funded project called PharmPlant
15 and Medicago Inc., Canada have worked on a tobacco-based platform for the production of virus-like particles for flu vaccine manufacture.

Thus, a plant according to the present invention may be used for molecular farming to reduce or eliminate the presence of nicotine, nornicotine, PON and/or other nicotinic alkaloids, such as anabasine or anatabine or myosmine. The use of a low nicotine plant or rootstock is beneficial in
20 molecular farming and would reduce downstream processing costs associated with purification. In other instances, it may be desirable to produce plants with high alkaloid levels e.g. high levels of nicotine content so that nicotine may be purified from the tobacco plant to produce a pure nicotine product for example for use in devices which utilize liquid containing nicotine (e.g. e-cigarettes) or within tobacco heating devices. For example, the production of plants with leaves
25 containing high levels of nicotine could reduce costs of nicotine extraction for the production of e-liquids for e-cigarettes.

The present inventors have surprisingly determined a method for modulating the alkaloid content, e.g. nicotine content, of a plant (e.g. a tobacco plant) by modulating the activity or expression of a gene encoding a SOUL haem-binding protein. The alkaloid content (e.g. the
30 content of one or more of nicotine, nornicotine, anabasine, PON, myosmine or anatabine) of a plant (e.g. tobacco plant) may be decreased by decreasing the activity or expression of a gene encoding a SOUL haem-binding protein or may be increased by increasing the activity or expression of gene encoding a SOUL haem-binding protein (for example by increasing the haem-binding affinity of a SOUL haem-binding protein). Prior to the present invention it had not
35 been known that modulation of the activity or expression of a gene encoding a SOUL haem-binding protein as described herein could be used to modulate alkaloid content.

In one aspect, there is provided a method of modulating (e.g. decreasing) the alkaloid content of a plant or a part thereof, or a cell or cell culture, the method comprising modifying said plant or a cell or cell culture by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein.

- 5 In one aspect, there is provided a method of decreasing the alkaloid content of a plant or a part thereof, or a cell or cell culture, the method comprising modifying said plant or a cell or cell culture by decreasing the activity or expression of at least one SOUL haem-binding protein.
- In another aspect, there is provided a method of modulating (e.g. decreasing) the content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA in a tobacco plant or plant part
- 10 thereof, or cell, the method comprising modifying said plant or a cell or cell culture by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein.

- In another aspect, there is provided a method of decreasing the content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA in a tobacco plant or plant part thereof, or cell, the
- 15 method comprising modifying said plant or a cell or cell culture by decreasing the activity or expression of at least one SOUL haem-binding protein.

In a further aspect there is provided the use of at least one gene encoding a SOUL haem-binding protein for modulating alkaloid content of a cell or plant or part thereof or a cell or cell culture.

- 20 In yet another aspect, there is provided a method for producing a plant or part thereof, a cell or cell culture, a plant propagation material, a leaf, a cut harvested leaf, a processed leaf or a cut and processed leaf which has modulated (e.g. decreased) alkaloid content, the method comprising modifying said plant or cell or cell culture to modulate the activity or expression of at least one gene encoding a SOUL haem-binding protein.

- 25 Suitably, the alkaloid content may be modulated (e.g. decreased) in comparison to a plant or cell or cell culture which has not been modified to modulate the activity or expression of the at least one gene encoding a SOUL haem-binding protein.

- In a further aspect, there is provided a plant or part thereof or a cell or cell culture which has been modified to achieve a modulation (e.g. decrease) in alkaloid content in comparison to an
- 30 unmodified plant or unmodified cell or cell culture, wherein the modification is the reduction of the activity or expression of at least one gene encoding a SOUL haem-binding protein.

In a further aspect, there is provided a plant propagation material obtainable from a plant according to the present invention or from a plant or cell or cell culture produced by the method according to the present invention.

- 35 In a further aspect, there is provided a method or use according to the present invention, or a plant or part thereof or a cell or cell culture according to the present invention, or a plant

propagation material according to the present invention, wherein the alkaloid content of the plant is decreased in comparison to a plant or cell or cell culture which has not been modified to modulate the activity or expression of at least one gene encoding a SOUL haem-binding protein.

- 5 Suitably, the activity or expression of at least one gene encoding a SOUL haem-binding protein may be decreased in comparison to a plant or cell or cell culture which has not been modified to modulate the activity or expression of the at least one gene encoding a SOUL haem-binding protein.

- Suitably, the alkaloid content of the plant or cell or cell culture may be increased in comparison
10 to a plant which has not been modified to modulate the activity or expression of the at least one gene encoding a SOUL haem-binding protein.

- Suitably, the plant may be modified to increase the activity or expression of at least one gene encoding a SOUL haem-binding protein and the plant or cell or cell culture exhibits increased alkaloid content in comparison to a plant or cell or cell culture which has not been modified to
15 modulate the activity or expression of the at least one gene encoding a SOUL haem-binding protein.

Suitably, the total alkaloid content of the plant or cell or cell culture may be modulated (e.g. decreased).

- Suitably, the content of one or more alkaloids may be selected from nicotine, nornicotine, PON, anabasine, myosmine and anatabine is modulated (e.g. decreased), preferably the content of
20 nicotine, nornicotine and/or PON is modulated (e.g. decreased). In some embodiments two or more (or three or more) alkaloids selected from nicotine, nornicotine, PON, anabasine, myosmine and anatabine is modulated (e.g. decreased). In some embodiments the total alkaloid content of the plant or cell is modulated (e.g. decreased).

- 25 In one aspect, the plant or plant cell is from the *Solanaceae* family.

Suitably, the plant or plant cell may be from the *Nicotiana* genus.

Suitably, the nicotine content may be modulated. Suitably, the nicotine content may be decreased.

- In one aspect there is provided a method or use according to the present invention, a plant or
30 part thereof or a cell or cell culture according to the present invention, or a plant propagation material according to the present invention wherein the at least one gene encoding a SOUL haem-binding protein encodes a polypeptide which comprises an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof; or

wherein the at least one gene encoding a SOUL haem-binding protein encodes a polypeptide which comprises an amino acid sequence which has at least 80% identity to SEQ ID NO. 1, 4, 7, 10, 13, 16, 19 or 22; or

5 wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof; or

wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

10 In a further aspect there is provided a method or use according to the present invention, a plant or part thereof or cell or cell culture according to the present invention, or a plant propagation material according to the present invention wherein an additional gene encoding a SOUL haem-binding protein is also modulated, wherein the additional gene encodes a polypeptide which comprises an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a
15 functional variant or functional fragment or orthologue thereof; or

wherein the additional gene encodes a polypeptide which comprises an amino acid sequence which has at least 80% identity to SEQ ID NO. 1, 4, 7, 10, 13, 16, 19 or 22; or

20 wherein the additional gene comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof; or

wherein the additional gene comprises a nucleotide sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

25 In one aspect there is provided the use of a plant or part thereof or cell or cell culture according to the present invention, or of a plant produced by the method according to the present invention to breed a plant.

In another aspect there is provided the use of a plant or part thereof or a cell or cell culture according to the present invention, or of a plant produced by the method according to the present invention for production of a product.

30 In yet another aspect there is provided the use of a plant or part thereof according to the present invention, or of a plant produced by the method according to the present invention to grow a crop.

In another aspect there is provided the use of a plant or part thereof according to the present invention, or of a plant produced by the method according to the present invention to produce a leaf.

35 In another aspect there is provided a harvested leaf of a plant according to the present invention, or obtainable from a plant propagated from a propagation material according to the

present invention, or obtainable from a plant obtained by a use according to the present invention, or obtainable from a plant produced by the method according to the present invention. Suitably, the harvested may be a cut harvested leaf.

In another aspect there is provided a processed leaf, preferably a processed tobacco leaf,
5 preferably a non-viable processed tobacco leaf:

obtainable from a plant obtainable from a use according to the present invention;

obtainable by processing a plant according to the present invention;

obtainable from a plant propagated from a plant propagation material according to the present invention; or

10 obtainable by processing a harvested leaf of a plant according to the present invention;
or

obtainable from a plant produced by the method according to the present invention.

Suitably, the leaf may be processed by curing, fermenting, pasteurising or a combination thereof.

15 Suitably, the processed leaf may be a cut processed leaf.

In another aspect there is provided cured tobacco material made from a plant or a part thereof according to the present invention or an extract thereof.

In another aspect there is provided a tobacco blend comprising said cured tobacco material according to the present invention.

20 In another aspect there is provided a tobacco industry product prepared from:

a tobacco plant according to the present invention, or a part thereof or a tobacco cell or cell culture according to the present invention;

a tobacco plant or part thereof propagated from a tobacco plant propagation material according to the present invention;

25 a harvested leaf of a plant according to the present invention, wherein the plant is tobacco;

a processed leaf according to the present invention, wherein the plant is tobacco;

or

a plant produced by the method according to the present invention.

30 Suitably, the tobacco product may be a combustible smoking article.

Suitably, the tobacco product may be a smokeless tobacco product.

Suitably, the tobacco product may be a non-combustible aerosol provision system such as a tobacco heating device or an aerosol-generating device.

In one aspect there is provided the use of a tobacco cell according to the present invention for
35 modulating alkaloid content in cell cultures.

In another aspect there is provided a combustible smoking article, non-combustible aerosol provisioning system, smokeless tobacco product or tobacco heating device comprising a plant or a part thereof according to the present invention or an extract (e.g. a tobacco extract) thereof or a tobacco cell or cell culture according to the present invention; or a cured tobacco material according to the present invention; or a tobacco blend according to the present invention.

In one aspect there is provided the use of a nucleotide sequence of at least one gene encoding a SOUL haem-binding protein to select a plant having modulated (e.g. decreased) alkaloid content and/or modulated (e.g. decreased) content of tobacco specific nitrosamine (TSNA) or a precursor of a TSNA, preferably wherein the sequence of the SOUL haem-binding protein is selected from SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof; or wherein the sequence of the SOUL haem-binding protein is a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In another aspect there is provided a mutant of a plant carrying a heritable mutation in a nucleotide sequence of at least one gene encoding a SOUL haem-binding protein, preferably wherein the gene is selected from SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof; or wherein the gene is selected from a sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24; wherein said heritable mutation modulates (e.g. decreases) the activity or expression of the at least one gene encoding a SOUL haem-binding protein and wherein the mutant plant has modulated (e.g. decreased) alkaloid content and/or modulated content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA relative to a comparable plant which does not carry said heritable mutation.

In another aspect there is provided progeny or seed of a mutant plant which carries the heritable mutation according to the present invention.

In another aspect there is provided a harvested leaf, a processed leaf or cured tobacco material produced from a plant comprising a modification in a nucleotide sequence of at least one gene encoding a SOUL haem-binding protein, wherein the at least one gene is selected from SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof; or wherein the at least one gene is selected from a sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24; wherein said modification modulates (e.g. decreases) the activity or expression of the at

least one gene encoding a SOUL haem-binding protein and wherein said plant has modulated (e.g. decreased) alkaloid content and/or modulated content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA relative to a comparable plant which does not carry said modification in the at least one gene encoding a SOUL haem-binding protein.

- 5 In another aspect there is provided a method, a leaf, a plant, a plant propagation material, a harvested leaf, a processed tobacco, a tobacco product, a use or a combination thereof as described herein with reference to the description and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

- 10 Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 depicts, without wishing to be bound by theory, the effect of overexpressing SOUL haem-binding protein on pyridine alkaloid biosynthesis. (A) Upon perception of wounding or herbivory by surface-localised receptors, immune signalling is induced via phosphorylation of
15 MAPKs and production of reactive oxygen species (ROS). This activates jasmonic acid (JA)-responsive genes and upregulates pyridine alkaloid production. Intracellular ROS levels are kept under tight regulation through haem oxygenase (HO)-mediated haem catabolism. (B) SOUL overexpression reduces the available haem to be broken down by HO. This leads to an increase in intracellular ROS, which upregulates the nicotine biosynthetic pathway. PM: plasma
20 membrane.

Figure 2 shows the vector used to transiently overexpress a SOUL haem-binding protein in tobacco.

- Figure 3** shows the alkaloid (nicotine, nor nicotine, anabasine, PON and anatabine) content of 5-week-old TN90 leaves expressing Nitab4.5_0013616g0010.2 (SEQ ID No. 3). Alkaloid content
25 is represented relative to a control and comprises three biological replicates analysed by t-test. Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.001.

- Figure 4** shows the alkaloid (nicotine, nor nicotine, anabasine, PON and anatabine) content of 5-week-old TN90 leaves expressing a construct which silences Nitab4.5_0013616g0010.2 (SEQ
30 ID No. 3) by virus-induced gene silencing (VIGS). Alkaloid content is represented relative to control and comprises three biological replicates analysed by t-test. Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.001

Figure 5 shows the amino acid sequence of Nitab4.5_0013616g0010.2 – SEQ ID No. 1 – a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 6 shows the genomic sequence of Nitab4.5_0013616g0010.2 – SEQ ID No. 2 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 7 shows the coding sequence of Nitab4.5_0013616g0010.2 – SEQ ID No. 3 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 8 shows the amino acid sequence of Nitab4.5_0000652g0130.2 – SEQ ID No. 4 – a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 9 shows the genomic sequence of Nitab4.5_0000652g0130.2 – SEQ ID No. 5 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 10 shows the coding sequence of Nitab4.5_0000652g0130.2 – SEQ ID No. 6 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 11 shows the amino acid sequence of Nitab4.5_0001140g0220.2 – SEQ ID No. 7 – a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 12 shows the genomic sequence of Nitab4.5_0001140g0220.2 – SEQ ID No. 8 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 13 shows the coding sequence of Nitab4.5_0001140g0220.2 – SEQ ID No. 9 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 14 shows the amino acid sequence of Nitab4.5_0003235g0060.2 – SEQ ID No. 10 – a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 15 shows the genomic sequence of Nitab4.5_0003235g0060.2 – SEQ ID No. 11 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 16 shows the coding sequence of Nitab4.5_0003235g0060.2 – SEQ ID No. 12 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 17 shows the amino acid sequence of Nitab4.5_0004868g0020.2 – SEQ ID No. 13 – a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 18 shows the genomic sequence of Nitab4.5_0004868g0020.2 – SEQ ID No. 14 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 19 shows the coding sequence of Nitab4.5_0004868g0020.2 – SEQ ID No. 15 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 20 shows the amino acid sequence of Nitab4.5_0006614g0010.2 – SEQ ID No. 16 – a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 21 shows the genomic sequence of Nitab4.5_0006614g0010.2 – SEQ ID No. 17 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 22 shows the coding sequence of Nitab4.5_0006614g0010.2 – SEQ ID No. 18 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 23 shows the amino acid sequence of Nitab4.5_0006991g0050.2 – SEQ ID No. 19 – a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 24 shows the genomic sequence of Nitab4.5_0006991g0050.2 – SEQ ID No. 20 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 25 shows the coding sequence of Nitab4.5_0006991g0050.2 – SEQ ID No. 21 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 26 shows the amino acid sequence of Nitab4.5_0009023g0010.2 – SEQ ID No. 22 – a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 27 shows the genomic sequence of Nitab4.5_0009023g0010.2 – SEQ ID No. 23 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 28 shows the coding sequence of Nitab4.5_0009023g0010.2 – SEQ ID No. 24 – encoding a SOUL haem-binding protein from *Nicotiana tabacum* according to the present invention.

Figure 29 shows SEQ ID No. 25, a 300-nucleotide cDNA fragment used in Example 2.

Figure 30 shows SEQ ID No. 26, sequence TRV1 used in Example 2.

Figure 31 shows SEQ ID No. 27, sequence TRV2 used in Example 2.

Figure 32 shows the nor nicotine content of 5-week-old TN90 leaves transiently expressing the indicated constructs: OE = expressing Nitab4.5_0013616g0010.2; or As = an antisense construct which silences Nitab4.5_0013616g0010.2. Content is represented relative to control and comprises three biological replicates analysed by one-way ANOVA and Tukey's multiple-comparison post-test. Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.001.

Figure 33 shows the nor nicotine content of 5-week-old TN90 leaves transiently expressing constructs which encode the indicated homologue sequences (see Example 3). Content is represented as percentage relative to control (100%) and comprises two biological replicates analysed by t-test. Values are shown as means $\pm$ SEM.

5 **Figure 34** shows the nor nicotine content of greenhouse-grown TN90 plants at 12 leaf stage overexpressing Nitab4.5_0013616g0010.2 or control plants. (n=24).

Figure 35 shows in panel (A) the nor nicotine content of field-grown control (wild type WT) TN90 leaves and Nitab4.5_0013616g0010.2 overexpressors (OX). Content is represented relative to control. Results were analysed by t-test. Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.001. In panel (B) the NNN content of field-grown control (WT) TN90 leaves and Nitab4.5_0013616g0010.2 overexpressors. Content is represented relative to control. Results were analysed by t-test. Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.01.

15 **SEQUENCE LISTING**

A summary of sequence identifiers used throughout the subject specification and the corresponding sequence listing is provided wherein:

SEQ ID No. 1 corresponds to the amino acid sequence of Nitab4.5_0013616g0010.2.

SEQ ID No. 2 corresponds to the genomic sequence of Nitab4.5_0013616g0010.2.

20 **SEQ ID No. 3** corresponds to the coding sequence of Nitab4.5_0013616g0010.2.

SEQ ID No. 4 corresponds to the amino acid sequence of Nitab4.5_0000652g0130.2.

SEQ ID No. 5 corresponds to the genomic sequence of Nitab4.5_0000652g0130.2.

SEQ ID No. 6 corresponds to the coding sequence of Nitab4.5_0000652g0130.2.

SEQ ID No. 7 corresponds to the amino acid sequence of Nitab4.5_0001140g0220.2.

25 **SEQ ID No. 8** corresponds to the genomic sequence of Nitab4.5_0001140g0220.2.

SEQ ID No. 9 corresponds to the coding sequence of Nitab4.5_0001140g0220.2.

SEQ ID No. 10 corresponds to the amino acid sequence of Nitab4.5_0003235g0060.2.

SEQ ID No. 11 corresponds to the genomic sequence of Nitab4.5_0003235g0060.2.

SEQ ID No. 12 corresponds to the coding sequence of Nitab4.5_0003235g0060.2.

30 **SEQ ID No. 13** corresponds to the amino acid sequence of Nitab4.5_0004868g0020.2.

SEQ ID No. 14 corresponds to the genomic sequence of Nitab4.5_0004868g0020.2.

SEQ ID No. 15 corresponds to the coding sequence of Nitab4.5_0004868g0020.2.

SEQ ID No. 16 corresponds to the amino acid sequence of Nitab4.5_0006614g0010.2.

SEQ ID No. 17 corresponds to the genomic sequence of Nitab4.5_0006614g0010.2.

35 **SEQ ID No. 18** corresponds to the coding sequence of Nitab4.5_0006614g0010.2.

SEQ ID No. 19 corresponds to the amino acid sequence of Nitab4.5_0006991g0050.2.

SEQ ID No. 20 corresponds to the genomic sequence of Nitab4.5_0006991g0050.2.

SEQ ID No. 21 corresponds to the coding sequence of Nitab4.5_0006991g0050.2.

SEQ ID No. 22 corresponds to the amino acid sequence of Nitab4.5_0009023g0010.2.

SEQ ID No. 23 corresponds to the genomic sequence of Nitab4.5_0009023g0010.2.

5 **SEQ ID No. 24** corresponds to the coding sequence of Nitab4.5_0009023g0010.2.

SEQ ID No. 25 is a 300-nucleotide cDNA fragment used in Example 2.

SEQ ID No. 26 is sequence TRV1 used in Example 2.

SEQ ID No. 27 is sequence TRV2 used in Example 2.

10 Some sequences disclosed herein contain "X" or "N" in nucleotide sequences. "X" or "N" can be any nucleotide or a deletion or insertion of one or more nucleotides. For example, in some cases a string of "X"s or "N"s are shown. The number of "X"s or "N"s does not necessarily correlate with the actual number of nucleotides at that position. There may be more or fewer nucleotides than shown as "X" or "N" in the sequence.

15 **DETAILED DESCRIPTION**

For the first time the present inventors have shown that by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein in a plant (e.g. a tobacco plant), the alkaloid and/or TSNA content of the plant (or processed plant) can be modulated.

20 The present invention provides a method of modulating (e.g. decreasing) the alkaloid content of a plant or a part thereof, the method comprising modifying said plant by modulating (e.g. decreasing) the activity or expression of at least one gene encoding a SOUL haem-binding protein.

Also provided is a method of modulating (e.g. decreasing) the content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA in a tobacco plant or plant part thereof, the method comprising modifying said plant by modulating (e.g. decreasing) the activity or expression of at least one gene encoding a SOUL haem-binding protein.

25 The at least one gene encoding a SOUL haem-binding protein may be selected from at least one gene encoding a SOUL haem-binding protein which comprises an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

35 Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, more than one SOUL haem-binding proteins may be modified. In one embodiment at least one SOUL haem-binding protein gene is modified selected from the group comprising genes encoding polypeptides comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In one embodiment at least two SOUL haem-binding protein genes are modified selected from the group comprising genes encoding polypeptides comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In one embodiment at least three SOUL haem-binding protein genes are modified selected from the group comprising genes encoding polypeptides comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In one embodiment at least four SOUL haem-binding protein genes modified selected from the group comprising genes encoding polypeptides comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2,

3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In one embodiment at least five SOUL haem-binding protein genes are modified selected from the group comprising genes encoding polypeptides comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In one embodiment at least six SOUL haem-binding protein genes are modified selected from the group comprising genes encoding polypeptides comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In one embodiment at least seven SOUL haem-binding protein genes are modified selected from the group comprising genes encoding polypeptides comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In one embodiment at least eight SOUL haem-binding protein genes are modified selected from the group comprising genes encoding polypeptides comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2,

3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

Suitably, the SOUL haem-binding protein of the present invention may comprise a SOUL motif.

In one aspect, the at least one SOUL haem-binding protein gene encodes a polypeptide which
 5 comprises an amino acid sequence as set out in: SEQ ID No. 1 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2 or 3 or a functional variant or functional
 10 fragment or orthologue of SEQ ID No. 2 or 3; or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2 or 3. Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

In one aspect, the activity or expression of at least one further gene is modulated. Suitably, at least two (or at least three or at least four or at least five or at least six or at least seven or at least eight or at least nine) additional genes selected from SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12,
 15 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 may also be modulated. Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

20 The “activity” of a gene encoding a SOUL haem-binding protein may refer to the ability of the SOUL haem-binding protein produced by the gene to bind other molecules, such as other proteins. Suitably, the “activity” of a gene encoding a SOUL haem-binding protein refers to the ability of the SOUL haem-binding protein produced by the gene to bind haem.

The “expression” of a gene encoding a SOUL haem-binding protein may refer to the level of
 25 transcription, translation i.e. protein expression.

Measurement of the level or amount of a gene product may be carried out by any suitable method, for example comparison of mRNA transcript levels, protein or peptide levels, and/or phenotype of a plant, between a modified plant and comparable plant which has not been modified according to the present invention.

30 The term “a comparable product” as defined herein would be one derived from a plant (e.g. a tobacco plant) which had not been modified according to the present invention, but in which all other relevant features were the same (e.g. plant species, growing conditions, method of processing the plant, e.g. tobacco, etc.). The comparable product according to the present invention may mean a plant (e.g. a tobacco plant) or a part thereof, such as a leaf (e.g. a
 35 tobacco leaf), a harvested leaf (e.g. a harvested tobacco leaf), a cut harvested leaf (e.g. a cut harvested tobacco leaf), a processed leaf (e.g. a processed tobacco leaf) or plant propagation

material (e.g. tobacco plant propagation material), or a product comprising said plant or part therefore, e.g. a tobacco product or combinations thereof obtainable or obtained from a plant which has not been modified in accordance with the present invention, e.g. to modulate the activity or expression of gene encoding a SOUL haem-binding protein. In one embodiment a comparable product is one which does not comprise gene encoding a SOUL haem-binding protein whose activity or expression has been modulated.

The term “modifying” or “modified” as used herein means a plant (e.g. a tobacco plant) or nucleic acid sequence that has been altered or changed. The present invention comprises the modification of plants using techniques for genetic modification of plants or non-genetic modification of plants. Such methods are well known in the art and examples of genetic modification techniques include transformation, transgenics, cisgenics, and gene editing methods. Examples of non-genetic modification techniques include fast-neutron mutagenesis, chemical mutagenesis e.g. ethyl methanesulfonate (EMS) mutagenesis and modern population analysis approaches.

The term “unmodified plant” as defined herein would be a plant (e.g. a tobacco plant) which had not been modified according to the present invention, e.g. to modulate the activity or expression of a gene encoding a SOUL haem-binding protein or to modify the nucleic acid sequence of at least one gene encoding a SOUL haem-binding protein; and in which all other relevant features were the same (e.g. plant species, growing conditions, method of processing tobacco, etc.). In one embodiment an unmodified plant is one which does not comprise a gene encoding an SOUL haem-binding protein whose activity or expression has been modulated. In one embodiment, an unmodified plant is one which does not comprise a modified nucleic acid sequence which encodes at least one gene encoding a SOUL haem-binding protein.

SOUL haem-binding protein

A “SOUL haem-binding protein” as used herein has its usual meaning in the art and refers to a protein which comprises a SOUL motif, which is a hydrophobic cleft flanked by an α -helix and a β -loop.

SOUL haem-binding proteins are found in bacteria, plants and animals. SOUL haem-binding proteins bind tetrapyrroles such as haem (Takahashi et al. (2008) Photochem. Photobiol. Sci. 7, 1216-1224). It has been suggested that SOUL proteins in plants may be involved in transporting haem to other haem-containing proteins such as cytochrome P450s and in binding free haem (Freire et al. (2009) Acta Cryst. F65, 723-726).

The three-dimensional structure of a SOUL haem-binding protein from mouse has been determined. The murine SOUL structure consists of a nine-stranded twisted β -barrel, two α -helices and a hydrophobic cleft. The hydrophobic cleft is flanked by one of the α -helices and the

β 8-9 loop, forming a conserved "SOUL motif", which is essential for haem-binding (Dias et al. (2006) J. Biol. Chem. 276, 18161-18168). A similar structure is predicted for Arabidopsis SOUL haem-binding proteins (Takahashi et al. (2008) Photochem. Photobiol. Sci. 7, 1216-1224).

In one embodiment the SOUL haem-binding protein comprises a SOUL motif. As used herein
5 the term "SOUL motif" refers to a conserved structural motif. SOUL haem-binding proteins and SOUL motifs may be identified by comparison of the predicted structure of a protein with known protein structures. For example, SOUL haem-binding proteins and SOUL motifs may be identified by comparison of the predicted structure of a protein against the predicted structure of SEQ ID No. 1, wherein the presence of a hydrophobic cleft flanked by an α -helix and β -loop
10 identifies the protein as a SOUL haem binding protein.

In one embodiment, a SOUL motif is a region of a protein which corresponds to a hydrophobic cleft flanked by an α -helix and β -loop. The SOUL motif may be involved in haem-binding.

Without wishing to be bound by theory, it is hypothesized that increasing content of a SOUL haem-binding protein in a plant cell or increasing activity, such as haem-binding activity, of a
15 SOUL haem-binding protein in a plant would lead to increased binding of free haem. The reduction in free haem levels would mean a reduction in haem degradation by haem oxygenase, leading to reduced scavenging of reactive oxygen species (ROS). ROS stimulate activation of nicotine biosynthesis genes. Increased activity or expression of a SOUL haem-binding protein thereby leads to increased levels of pyridine alkaloids. Conversely, according to
20 this hypothesis, decreased levels or activity, such as haem-binding activity, of a SOUL haem-binding protein in a plant cell would increase the amount of free haem, leading to increased ROS scavenging and therefore reduced activation of nicotine biosynthesis genes, thereby decreasing levels of pyridine alkaloids in the cell. This hypothesis is depicted diagrammatically in Figure 1.

25 In one embodiment a SOUL haem-binding protein comprises an amino acid sequence shown as SEQ ID No. 1 or a sequence which has at least 80% identity thereto, or a homologue thereof. Suitably, the SOUL haem-binding protein may comprise a SOUL motif. Suitably, a homologue of SEQ ID No. 1 may be selected from the group comprising: SEQ ID No. 4, 7, 10, 13, 16, 19 or 22, or a sequence which has at least 80% identity thereto. Suitably, a homologue of SEQ ID No.
30 1 may be selected from the group comprising: SEQ ID No. 4, 7, 10, 13, 16, 19 or 22 wherein said sequence comprises a SOUL motif, or a sequence which has at least 80% identity to SEQ ID No. 4, 7, 10, 13, 16, 19 or 22 and comprises a SOUL motif.

In one embodiment a SOUL haem-binding protein comprises an amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a sequence which has at least 80% identity thereto
35 (preferably at least 85%, at least 90%, at least 95%, at least 97%, or at least 99% identity thereto). In one embodiment a SOUL haem-binding protein comprises an amino acid sequence

shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a sequence which has at least 80% identity thereto (preferably at least 85%, at least 90%, at least 95%, at least 97%, or at least 99% identity thereto) and comprises a SOUL motif.

Suitably, a SOUL haem-binding protein according to the present invention may comprise an amino acid sequence shown as SEQ ID No. 1, or a sequence which has at least 80% identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, a SOUL haem-binding protein according to the present invention may comprise an amino acid sequence shown as SEQ ID No. 4, or a sequence which has at least 80% identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, a SOUL haem-binding protein according to the present invention may comprise an amino acid sequence shown as SEQ ID No. 7, or a sequence which has at least 80% identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, a SOUL haem-binding protein according to the present invention may comprise an amino acid sequence shown as SEQ ID No. 10, or a sequence which has at least 80% identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, a SOUL haem-binding protein according to the present invention may comprise an amino acid sequence shown as SEQ ID No. 13, or a sequence which has at least 80% identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, a SOUL haem-binding protein according to the present invention may comprise an amino acid sequence shown as SEQ ID No. 16, or a sequence which has at least 80% identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, a SOUL haem-binding protein according to the present invention may comprise an amino acid sequence shown as SEQ ID No. 19, or a sequence which has at least 80% identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at

least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, a SOUL haem-binding protein according to the present invention may comprise an amino acid sequence shown as SEQ ID No. 22, or a sequence which has at least 80% identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

In one embodiment the SOUL haem-binding protein according to the present invention comprises or consists of an amino acid sequence selected from: SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22.

Suitably, the protein may be from *Nicotiana tabacum*.

In one embodiment the SOUL haem-binding protein is encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24; or a sequence which has at least 80% sequence identity thereto. Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 2, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 3, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 5, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 6, or a sequence which has at least 80% sequence identity

thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 8, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 9, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 11, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 12, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 14, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 15, or a sequence which has at least 80% sequence identity

thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 17, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 18, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 20, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 21, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 23, or a sequence which has at least 80% sequence identity thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

Suitably, the SOUL haem-binding protein for use according to the present invention may be encoded by a polynucleotide sequence wherein the gene (prior to mutation) comprises the sequence shown as SEQ ID No. 24, or a sequence which has at least 80% sequence identity

thereto (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, at least 97%, or at least 99% identity thereto). Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

In one embodiment the SOUL haem-binding protein is encoded by a polynucleotide sequence wherein the gene (prior to mutation) is selected from: SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

Suitably, the protein for use according to the present invention may be encoded by a polynucleotide sequence from *Nicotiana tabacum*.

In one aspect the present invention provides a method of decreasing the alkaloid content of a plant or part thereof or cell (e.g. plant cell), the method comprising modifying said plant by decreasing or inhibiting the activity or expression of at least one gene encoding a SOUL haem-binding protein.

In one aspect the present invention provides a method of decreasing the alkaloid content of a plant or part thereof or plant cell, the method comprising modifying said plant by decreasing or inhibiting the activity or expression of at least one gene encoding a SOUL haem-binding protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a sequence which has at least 80% identity thereto or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24. Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

In one aspect the present invention provides a method of decreasing the content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA a plant or part thereof (e.g. leaf), the method comprising modifying said plant by decreasing or inhibiting the activity or expression of at least one gene encoding a SOUL haem-binding protein.

In one aspect the present invention provides a method of decreasing the content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA in a plant or part thereof (e.g. leaf), the method comprising modifying said plant by decreasing or inhibiting the activity or expression of at least one gene encoding a SOUL haem-binding protein comprising the amino acid sequence shown as SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a sequence which has at least 80% identity thereto, or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a nucleic

acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

The term “decreasing” or “inhibiting” (e.g. inhibiting the activity or expression of gene encoding a SOUL haem-binding repeat protein) as used herein means that the activity or expression of the gene encoding the SOUL haem-binding repeat protein is lower or decreased compared with the activity or expression of the gene in a comparable product.

In one aspect the present invention provides a method of increasing the alkaloid content of a plant or part thereof or cell (e.g. plant cell), the method comprising modifying said plant by increasing or enhancing the activity or expression of at least one gene encoding a SOUL haem-binding protein.

In one aspect the present invention provides a method of increasing the alkaloid content of a plant or part thereof or plant cell, the method comprising modifying said plant by increasing or enhancing the activity or expression of at least one gene encoding a SOUL haem-binding protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a sequence which has at least 80% identity thereto or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24. Suitably, the SOUL haem-binding protein may comprise a SOUL motif.

In one aspect the present invention provides a method of increasing the content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA a plant or part thereof (e.g. leaf), the method comprising modifying said plant by increasing or enhancing the activity or expression of at least one gene encoding a SOUL haem-binding protein.

In one aspect the present invention provides a method of increasing the content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA in a plant or part thereof (e.g. leaf), the method comprising modifying said plant by increasing or enhancing the activity or expression of at least one gene encoding a SOUL haem-binding protein comprising the amino acid sequence shown as SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a sequence which has at least 80% identity thereto, or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

The term “increasing” or “enhancing” (e.g. increasing the activity or expression of gene encoding a SOUL haem-binding repeat protein) as used herein means that the activity or expression of the gene encoding the SOUL haem-binding repeat protein is higher or increased compared with the activity or expression of the gene in a comparable product.

- 5 According to the present invention, the activity or expression of a gene encoding a SOUL haem-binding protein is modulated.

In one aspect the present invention provides a method of modulating (i.e. increasing or decreasing) the alkaloid content of a plant or part thereof or cell (e.g. plant cell), the method comprising modifying said plant by modulating (i.e. increasing or decreasing) the activity of at
10 least one gene encoding a SOUL haem-binding protein.

The term “activity” refers to any functionality of the SOUL haem-binding protein encoded by the at least one gene. Examples of activity include enzymatic activity or localization of the SOUL haem-binding protein. Suitably, the activity is the ability of the SOUL haem-binding protein to interact with another molecule or molecules. In some embodiments the invention provides a
15 method of modulating (i.e. increasing or decreasing) the alkaloid content of a plant or part thereof or cell (e.g. plant cell), the method comprising modifying said plant by modulating (i.e. increasing or decreasing) the ability of a SOUL haem-binding protein to interact with another molecule.

Suitably, the ability of the SOUL haem-binding protein to interact with the other molecule is the
20 ability to bind the other molecule. The other molecule may be a protein. The other molecule may be a tetrapyrrole. The other molecule may be haem. In some embodiments the invention provides a method of modulating (i.e. increasing or decreasing) the alkaloid content of a plant or part thereof or cell (e.g. plant cell), the method comprising modifying said plant by modulating (i.e. increasing or decreasing) the ability of a SOUL haem-binding protein to bind haem.

25 Suitably, the other molecule is more than one molecule, such as one or more molecules, such as two or more molecules, such as three or more molecules. Where the other molecule is more than one molecule, the other molecules may be the same molecule or may be different molecules. Where the other molecule is more than one molecule, at least one of the other molecules may be haem.

30 Modulation of the activity of a gene encoding a SOUL haem-binding protein may entail increasing or decreasing the activity of the SOUL haem-binding protein.

Increasing the activity of a SOUL haem-binding protein refers to enhancing or improving the ability of the SOUL haem-binding protein to carry out a particular function in comparison to a SOUL haem-binding protein in a plant that has not been modified in accordance with the
35 invention.

Decreasing the activity of a SOUL haem-binding protein refers to reducing, inhibiting or disrupting the ability of the SOUL haem-binding protein to carry out a particular function in comparison to a SOUL haem-binding protein in a plant that has not been modified in accordance with the invention. The activity of a SOUL haem-binding protein may be reduced to
5 such an extent that the activity is prevented or eliminated.

In some embodiments the activity of a SOUL haem-binding protein may be modulated (i.e. increased or decreased) by at least about 10% 20% 30%, or 40%, suitably at least about 50%, 60%, 70%, more suitably at least about 80%, 90%, 95% or 100% in comparison to the activity of a gene encoding a SOUL haem-binding protein in a plant (e.g. a tobacco plant) which has not
10 been modified in accordance with the present invention.

In some embodiments the modulated SOUL haem-binding protein exhibits increased or decreased activity compared to an unmodified SOUL haem-binding protein. The modulated SOUL haem-binding protein may exhibit at least about 1%, at least about 3%, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least
15 about 50%, at least about 60%, at least about 70%, at least about 80% or at least about 90% increased or decreased activity compared to an unmodified SOUL haem-binding protein.

Suitably, the modulation of activity is an increase or decrease in the ability of the SOUL haem-binding protein to interact with (such as bind) another molecule or molecules (such as haem).

In some embodiments the invention provides a method of modulating the alkaloid content of a plant or part thereof or cell, the method comprising modifying said plant by modulating the ability
20 of a SOUL haem-binding protein to bind haem. Suitably the ability of a SOUL haem-binding protein to bind haem is increased. Suitably the ability of a SOUL haem-binding protein to bind haem is decreased. An increase or decrease in the ability of the SOUL haem-binding protein to bind another molecule may be expressed as an increase or decrease in the binding affinity of
25 the SOUL haem-binding protein. In some embodiments the invention provides a method of modulating the alkaloid content of a plant or part thereof or cell, the method comprising modifying said plant by modulating the binding affinity of a SOUL haem-binding protein for haem.

Techniques are known in the art for measuring protein activities. For example, assays are
30 known for measuring the enzymatic activity of a protein and the localization of a protein can be identified using microscopy techniques.

In particular, the ability of a SOUL haem-binding protein to bind another molecule may be measured using techniques known in the art. Examples of such techniques include immunoprecipitation, isothermal calorimetry, surface plasmon resonance and microscale
35 thermophoresis. For example, the ability of a modulated or mutated SOUL haem-binding protein to bind other molecules may be determined for example by performing co-immunoprecipitation

experiments using a modulated or mutated SOUL haem-binding protein and a corresponding unmodified or unmutated SOUL haem-binding protein. If the modulation or mutation in the SOUL haem-binding protein reduces, inhibits or eliminates the ability of the SOUL haem-binding protein bind other molecules, the co-immunoprecipitation will show that the modulated or mutated SOUL haem-binding protein binds fewer other molecules.

According to the present invention, the activity or expression of a SOUL haem-binding protein is modulated.

In one aspect the present invention provides a method of modulating (i.e. increasing or decreasing) the alkaloid content of a plant or part thereof or cell (e.g. plant cell), the method comprising modifying said plant by modulating (i.e. increasing or decreasing) the expression of at least one gene encoding a SOUL haem-binding protein.

The "expression" of a gene refers to the degree to which the information encoded in the gene is converted to a functionality. The level of expression of a gene may be equated with the amount of the product of that gene present in a cell or organism. A modification that modulates (i.e. increases or decreases) the expression of a gene is one that increases the amount of the product of that gene in a plant or cell in comparison to an unmodified plant or cell.

In some embodiments the expression of a SOUL haem-binding protein is modulated (i.e. increased or decreased) in comparison to the expression of a gene encoding a SOUL haem-binding protein in a plant (e.g. a tobacco plant) which has not been modified in accordance with the present invention.

In some embodiments the expression of a SOUL haem-binding protein may be modulated (i.e. increased or decreased) by at least about 10% 20% 30%, or 40%, suitably at least about 50%, 60%, 70%, more suitably at least about 80%, 90%, 95% or 100% in comparison to the expression of a gene encoding a SOUL haem-binding protein in a plant (e.g. a tobacco plant) which has not been modified in accordance with the present invention.

In some embodiments the modulated SOUL haem-binding protein exhibits increased or decreased expression compared to an unmodified SOUL haem-binding protein. The modulated SOUL haem-binding protein may exhibit at least about 1%, at least about 3%, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80% or at least about 90% increased or decreased expression compared to an unmodified SOUL haem-binding protein.

Typically, genes are transcribed to mRNA, which is translated to protein, the final gene product. Proteins may be sequestered in cellular stores and/or degraded. The expression of a gene may be modulated by modulating any or all of these steps. Accordingly, in some embodiments the modification modulates expression of at least one SOUL haem-binding protein gene in one of the following ways:

modulating transcription from the at least one SOUL haem-binding protein gene;
modulating translation of the mRNA from the at least one SOUL haem-binding protein gene;
modulating release of the SOUL haem-binding protein from intracellular stores; and/or
modulating the rate of degradation of the SOUL haem-binding protein.

- 5 The expression of specific genes encoding SOUL haem-binding proteins can be measured by measuring transcription and/or translation of the gene. Methods for measuring transcription are well known in the art and include, amongst others, northern blot, RNA-Seq, in situ hybridization, DNA microarrays and RT-PCR. Alternatively, the expression of a gene may be measured indirectly by measuring the level of the gene product for example the protein encoded by said
- 10 gene. For example, the expression of a SOUL haem-binding protein may be determined by measuring the presence of the protein using an antibody specific for the SOUL haem-binding protein (for example antibodies specific for a SOUL motif) by western blot.

Modifying

- 15 The term “modifying” as used herein refers to alteration of the genetic material of a plant or cell. The plant or cell may be modified in any way that modulates activity or expression of at least one gene encoding a SOUL haem-binding protein. Types of modifications to plants and cells that modulate activity or expression of genes, as well as techniques to achieve those modifications, are known in the art.

- 20 In some embodiments the present invention provides a method of decreasing the alkaloid content of a plant or part thereof or cell (e.g. plant cell), the method comprising modifying said plant by decreasing or inhibiting the activity or expression of at least one gene encoding a SOUL haem-binding protein.

- In some embodiments the present invention provides a method of decreasing the content of a
- 25 tobacco specific nitrosamine (TSNA) or a precursor of a TSNA in a tobacco plant or plant part thereof, the method comprising modifying said plant or a cell or cell culture by decreasing the activity or expression of at least one gene encoding a SOUL haem-binding protein.

Any method known in the art for decreasing or inhibiting the activity or expression of a gene may be used in the methods according to the present invention.

- 30 Suitably, the activity or expression of the SOUL haem-binding protein gene may be reduced, partly inactivated, inhibited, eliminated, knocked out or lost such that the protein activity, expression or function of the SOUL haem-binding protein gene is not detectable.

In one aspect, the at least one SOUL haem-binding protein gene is knocked out. In other words, the SOUL haem-binding protein gene has been rendered completely inoperative.

- 35 By way of example, the present method may comprise:

- providing a mutation in a nucleic acid sequence which encodes a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or an amino acid sequence which has at least 80% sequence identity thereto;
- providing a mutation in a regulatory region (e.g. a promoter or an enhancer) which contributes to controlling the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto;
- providing an antisense RNA, siRNA or miRNA which reduces the level of nucleic acid sequence encoding a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto.

Each of the above approaches results in the reduction or prevention of activity or expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

As used herein, the term "mutation" encompasses a natural genetic variant or an engineered variant. In particular, the term "mutation" refers to a variation in the nucleotide sequence encoding the amino acid sequence or in the amino acid sequence compared to the sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% (preferably at least 85%, preferably at least 90%, preferably at least 93%, preferably at least 95%, preferably at least 98%, preferably at least 99%) sequence identity thereto.

In one embodiment the mutation decreases the alkaloid content of a plant. In another embodiment, the mutation decreases the content of at least one TSNA or a precursor of a TSNA in tobacco.

In one embodiment, a method according to the present invention may comprise providing a nucleic acid sequence to a plant or part thereof or plant cell, wherein said nucleic acid results in the reduction or elimination of the activity or expression of at least one gene encoding a SOUL haem-binding protein.

In one embodiment, a method according to the present invention may comprise providing a nucleic acid sequence to a plant or part thereof or plant cell, wherein said nucleic acid results in the modification of the nucleic acid sequence of at least one gene encoding a SOUL haem-binding protein.

Suitably said nucleic acid sequence may be introduced to the plant or part thereof or cell. Suitably an endogenous nucleic acid sequence in the plant or part thereof or cell may be modified to encode the polypeptide according to the present invention (e.g. by gene editing). For example, an endogenous nucleotide sequence may be modified to decrease the activity or expression of at least one gene encoding a SOUL haem-binding protein.

In a preferred embodiment, each copy of a nucleic acid sequence encoding a protein comprising a sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a sequence which has at least 80% sequence identity thereto or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 which is present in the plant is modified e.g. mutated as defined herein (e.g. each genomic copy of a gene encoding said protein in a plant is mutated). For example, each copy of the gene in the allotetraploid genome of *Nicotiana tabacum* may be mutated.

In a preferred embodiment, some or all of the homologues of the SOUL haem-binding protein described herein are modified e.g. inhibited or mutated. Suitably, some or all of SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or corresponding sequences which have at least 80% sequence identity thereto are modified e.g. inhibited or mutated.

In some embodiments the plant or plant cell according to the present invention is homozygous. Suitably, the plant or plant cell may be homozygous for the modification e.g. inhibition or mutation.

In some embodiments the plant or plant cell according to the present invention expresses only the modified e.g. mutated nucleic acid encoding the SOUL haem-binding protein. In other words, in some embodiments no endogenous (or endogenous and functional protein) is present in the plant according to the present invention. In other words, if any endogenous protein is present it is preferably in an inactive form.

In one embodiment the present method may comprise providing a mutation in the nucleic acid sequence shown as SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a nucleic acid sequence which has at least 80% identity thereto.

The mutation may alter the plant genome such that a nucleic acid sequence encoding a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or an amino acid sequence which has at least 80% sequence identity thereto is completely or partially deleted or otherwise modified to inhibit or eliminate the ability of the SOUL haem-binding protein to interact with other molecules, such as bind to other molecules, in comparison with a protein shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a sequence which has at least 80%

sequence identity thereto. In some embodiments the mutation does not alter the level or expression of the protein but reduces inhibits or eliminates the ability of the SOUL haem-binding protein to interact with other molecules, such as bind to other molecules, in comparison with a protein shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a sequence which has at least 80% sequence identity thereto. Suitably, the mutation inhibits or eliminates ability of the SOUL haem-binding protein to interact with other proteins. Suitably, the mutation inhibits or eliminates the ability of the SOUL haem-binding protein to bind other molecules. Suitably, the mutation inhibits or eliminates the ability of the SOUL haem-binding protein to bind other proteins. Suitably, the mutation inhibits or eliminates the ability of the SOUL haem-binding protein to bind a tetrapyrrole or tetrapyrroles. Suitably, the mutation inhibits or eliminates the ability of the SOUL haem-binding protein to bind haem. The expression "inhibits or eliminates" means that the interaction between the SOUL haem-binding protein and the other molecule is reduced, suitably to the extent that the interaction is prevented (i.e. the SOUL haem-binding protein does not interact with the other molecule at all). Suitably inhibition or elimination of binding of the SOUL haem-binding protein to another molecule means a decrease in the binding affinity of the SOUL haem-binding protein for the other molecule. In some embodiments, the mutation reduces the binding affinity of the SOUL haem-binding protein for haem.

Suitably, the mutation may be in a domain of the SOUL haem-binding protein that mediates interaction with other molecules. The mutation may be in a domain of the SOUL haem-binding protein that mediates binding to haem. The mutation may be in the SOUL motif of the SOUL haem-binding protein. In some embodiments, the SOUL motif may be mutated thereby modifying the ability of the SOUL haem-binding protein to interact with other molecules. In other words, the SOUL motif may be mutated, resulting in a modification of the binding affinity of the SOUL haem-binding protein for other molecules. Suitably, the SOUL motif is mutated such that the binding affinity of the SOUL haem-binding protein for other molecules is increased. Suitably, the SOUL motif is mutated such that the binding affinity of the SOUL haem-binding protein for other molecules is decreased.

The mutation may interrupt the nucleic acid sequence which encodes a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto.

The interruption may cause the nucleic acid sequence to not be transcribed and/or translated.

The nucleic acid sequence may be interrupted, for example, by deleting or otherwise modifying the ATG start codon of the nucleic acid sequence such that translation of the protein is reduced or prevented.

The nucleic acid sequence may comprise one or more nucleotide change(s) that reduce or prevent expression of the protein or affect protein trafficking. For example, expression of the

protein may be reduced or prevented by introduction of one or more pre-mature stop codons, a frame shift, a splice mutation or a non-tolerated amino acid substitution in the open reading frame.

A premature stop codon refers to a mutation which introduces a stop codon into the open reading frame and prevents translation of the entire amino acid sequence. The premature stop codon may be a TAG ("amber"), TAA ("ochre"), or TGA ("opal" or "umber") codon.

A frame-shift mutation (also called a framing error or a reading frame shift) is a mutation caused by indels (insertions or deletions) of a number of nucleotides in a nucleic acid sequence that is not divisible by three. Due to the triplet nature of gene expression by codons, the insertion or deletion can change the reading frame, resulting in a completely different translation from the original. A frameshift mutation will often cause the reading of the codons after the mutation to code for different amino acids. The frameshift mutation will commonly result in the introduction of a premature stop codon.

A splice mutation inserts, deletes or changes a number of nucleotides in the specific site at which splicing takes place during the processing of precursor messenger RNA into mature messenger RNA. The deletion of the splicing site results in one or more introns remaining in mature mRNA and may lead to the production of abnormal proteins.

A non-tolerated amino acid substitution refers to a mutation which causes a non-synonymous amino acid substitution in the protein which results in reduced or ablated function of the protein.

Any method known in the art for providing a mutation in a nucleic acid sequence may be used in the method according to the present invention. For example, homologous recombination may be used, in which a vector is created in which the relevant nucleic acid sequence(s) are mutated and used to transform plants or plant cells. Recombinant plants or plant cells expressing the mutated sequence may then be selected.

In one embodiment the mutation introduces a non-tolerated amino acid substitution in a protein comprising an amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a sequence which has at least 80% sequence identity thereto.

In some embodiments, the SOUL motif may contain a mutation which decreases the expression of the at least one gene encoding a SOUL haem-binding protein.

The mutation may be a deletion, a splice mutant or codon encoding a non-tolerated amino acid substitution.

In one embodiment, the nucleic acid sequence encoding the SOUL haem-binding protein may be wholly or partially deleted. The deletion may be continuous, or may comprise a plurality of sections of sequence. The deletion preferably removes a sufficient amount of nucleotide sequence such that the nucleic acid sequence no longer encodes a functional SOUL haem-binding protein. The deletion may be total, in which case 100% of the coding portion of the

nucleic acid sequence is absent, when compared to the corresponding genome of a comparable unmodified plant. The deletion may, for example, remove at least 50, 60, 70, 80 or 90% of the coding portion of the nucleic acid sequence. Suitably, at least part of the protein may be deleted. The deletion may, for example, remove at least 10, 20, 30, 40, 50, 60, 70, 80 or 90% of the coding portion of the protein.

The deletion may remove at least part of the SOUL motif.

The deletion may, for example, remove at least 10, 20, 30, 40, 50, 60, 70, 80 or 90% of the SOUL motif.

Suitably, the deletion may remove at least 5 amino acids, at least 10 amino acids, at least 15, at least 20, at least 25, at least 30 amino acids from the SOUL motif. Suitably, the deletion may remove at least 5 amino acids, at least 10 amino acids, at least 15, at least 20, at least 25, at least 30 amino acids from the SOUL motif.

In one embodiment, the deletion may remove at least 100 amino acids, at least 150, at least 200, at least 250, at least 300, at least 350 amino acids from the C terminus of the SOUL haem-binding protein.

Suitably, the mutated protein may be a truncated protein which lacks at least about 100 amino acids, at least about 150 amino acids corresponding to the amino acids from the C-terminal of SEQ ID No. 1 or a sequence which has at least 80% (preferably at least 85%, at least, at least 90%, at least 95%, at least 98%) sequence identity thereto to a truncated protein which lacks at least about 100 amino acids, at least about 150 amino acids corresponding to the amino acids from the C-terminal of SEQ ID No. 1.

Methods for deletion of nucleic acid sequences in plants are known in the art. For example, homologous recombination may be used, in which a vector is created in which the relevant nucleic acid sequence(s) are missing and used to transform plants or plant cells. Recombinant plants or plant cells expressing the new portion of sequence may then be selected.

Plant cells transformed with a vector as described herein may be grown and maintained in accordance with well-known tissue culturing methods such as by culturing the cells in a suitable culture medium supplied with the necessary growth factors such as amino acids, plant hormones, vitamins, etc.

Modification of the nucleic acid sequence may be performed using targeted mutagenesis methods (also referred to as targeted nucleotide exchange (TNE) or oligo-directed mutagenesis (ODM)). Targeted mutagenesis methods include, without limitation, those employing zinc finger nucleases, TALENs (see WO2011/072246 and WO2010/079430), Cas9-like, Cas9/crRNA/tracrRNA, Cas9/gRNA, or other CRISPR systems (see WO 2014/071006 and WO2014/093622), meganucleases (see WO2007/047859 and WO2009/059195), or targeted mutagenesis methods employing mutagenic oligonucleotides, possibly containing chemically

modified nucleotides for enhancing mutagenesis with sequence complementarity to the gene, into plant protoplasts (e.g., KeyBase® or TALENs).

Alternatively, mutagenesis systems such as TILLING (Targeting Induced Local Lesions IN Genomics; McCallum et al. (2000) Nat. Biotech. 18:455, and McCallum et al. (2000) Plant Physiol. 123, 439-442, both incorporated herein by reference) may be used to generate plant lines which comprise a gene encoding a protein having a mutation. TILLING uses traditional chemical mutagenesis (e.g. ethyl methanesulfonate (EMS) mutagenesis, which produces random mutations) followed by high-throughput screening for mutations. Thus, plants, seeds, cells and tissues comprising a gene having the desired mutation may be obtained.

The method may comprise the steps of mutagenizing plant seeds (e.g. EMS mutagenesis), pooling of plant individuals or DNA, PCR amplification of a region of interest, heteroduplex formation and high-throughput detection, identification of the mutant plant, sequencing of the mutant PCR product. It is understood that other mutagenesis and selection methods may equally be used to generate such modified plants. Seeds may, for example, be radiated or chemically treated and the plants may be screened for a modified phenotype.

Fast neutron deletion mutagenesis may be used in a reverse genetics sense (i.e. with PCR) to identify plant lines carrying a deletion in the endogenous gene. See for example Ohshima et al. (1998) Virology 213:472-481; Okubara et al. (1994) Genetics 137:867-874; and Quesada et al. (2000) Genetics 154:421-4315 which are incorporated herein by reference.

In another approach, dominant mutants may be used to trigger RNA silencing due to gene inversion and recombination of a duplicated gene locus. See for example Kusaba et al. (2003) Plant Cell 15:1455-1467 (incorporated herein by reference).

Modified plants may be distinguished from non-modified plants, i.e., wild type plants, by molecular methods, such as the mutation(s) present in the DNA, and by the modified phenotypic characteristics. The modified plants may be homozygous or heterozygous for the modification. Preferably modified plants are homozygous for the modification.

In one embodiment the method of reducing or preventing the activity or expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto does not comprise treating the plant with a chemical (e.g. an agrochemical).

Other ways of reducing or preventing the expression will be apparent to one skilled in the art and include the use of virus-induced gene silencing (VIGs), micro RNA silencing, RNAi, antisense, tDNA insertions, or dominant negative constructs (or antimorphic mutations).

In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by virus-induced gene silencing.

In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by microRNAs.

5 In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by RNAi.

In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by antisense suppression.

10 In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by sense suppression.

15 In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by tDNA insertions.

In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by dominant negative constructs (or antimorphic mutations).

20 In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by a targeted mutagenesis based system.

25 In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by CRISPR based system.

30 In one embodiment the expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or an amino acid sequence which has at least 80% sequence identity thereto may be reduced or eliminated by zinc finger nuclease, TALENs, meganucleases, mutagenic oligonucleotides or TILLING.

In some embodiments the present invention provides a method of increasing the alkaloid content of a plant or part thereof or cell (e.g. plant cell), the method comprising modifying said plant by increasing or enhancing the activity or expression of at least one gene encoding a SOUL haem-binding protein.

35 Any method known in the art for increasing or enhancing the activity or expression of a gene may be used in the methods according to the present invention.

In some embodiments the method may comprise overexpressing at least one gene encoding a SOUL haem-binding protein. Suitably the method may comprise expressing one or more additional copies of the at least one gene encoding a SOUL haem-binding protein in the plant or cell. Suitably the method may comprise modifying the endogenous copy of the at least one gene
5 encoding a SOUL haem-binding protein such that its expression is increased. The method may comprise mutating the coding sequence of the at least one gene encoding a SOUL haem-binding protein. The method may comprise mutating a regulatory sequence that regulates expression of the at least one gene encoding a SOUL haem-binding protein.

Suitably the method may comprise transforming a cell of a plant (e.g. a tobacco plant) with a
10 genetic construct which encodes at least one SOUL haem-binding protein comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11,
15 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24; or which comprises a nucleotide sequence which encodes a protein which is capable of promoting or augmenting at least one endogenous SOUL haem-binding protein gene. It will be
20 appreciated that each of these options would result in an increased activity and expression of the polypeptide encoded by the at least one SOUL haem-binding protein gene. The method may comprise regenerating the plant from the transformed cell. There is provided use of genetic construct which is capable of increasing the activity and/or expression of a polypeptide encoded by at least one SOUL haem-binding protein gene for increasing the alkaloid content (e.g.
25 nicotine content) and or TSNA content (or precursor thereto) in a plant or part thereof or cell transformed with the construct.

The genetic construct may encode a polypeptide comprising the amino acid SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein
30 the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

35 In another embodiment, the invention relates to a method of increasing the alkaloid content of a plant or part thereof and/or TSNA content (or precursor thereto) in a plant or plant part thereof,

comprising modifying said plant by increasing the activity of at least one gene encoding a SOUL haem-binding protein.

In one embodiment the activity of at least one gene encoding a SOUL haem-binding protein may be increased by introducing (or providing) a mutation to at least one gene encoding an SOUL motif.

Suitably, the activity of at least one gene encoding a SOUL haem-binding protein may be increased by introducing a mutation to at least one gene encoding a SOUL haem-binding protein which comprises an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

In some embodiments a modification which increases the activity or expression of at least one SOUL haem-binding protein gene and thereby increases alkaloid content and/or TSNA content (or precursor thereof) by one of the following:

modulating transcription from the at least one SOUL haem-binding protein gene;

modulating translation of the mRNA from the at least one SOUL haem-binding protein gene;

modulating release of the SOUL haem-binding protein from intracellular stores; and/or

modulating the rate of degradation of the SOUL haem-binding protein.

Alkaloid content

In one embodiment the present invention provides a method of modulating the alkaloid content of a plant (e.g. a tobacco plant) or a part thereof, the method comprising modifying said plant by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein.

The term “modulating” is used herein to mean either increasing or decreasing.

The term “increasing alkaloid content” is used herein to mean that the alkaloid content in the product of the present invention (e.g. plant, part thereof (e.g. leaf), processed leaf or a product made from the plant (e.g. a tobacco product)) is higher compared with a comparable product which has not been modified in accordance with the present invention.

The term “decreasing alkaloid content” is used herein to mean that alkaloid content in the product of the present invention (e.g. plant, part thereof (e.g. leaf), processed leaf or a product

made from the plant (e.g. a tobacco product)) is lower compared with a comparable product which has not be modified in accordance with the present invention.

In some embodiments, the modulation of alkaloid content refers to an increase in alkaloid content wherein the activity or expression of at least one SOUL haem-binding protein gene is increased (or in other words the protein is overexpressed).

In some embodiments, the modulation of alkaloid content refers to a decrease in alkaloid content wherein the expression of at least one gene encoding a SOUL haem-binding protein is decreased or inhibited or eliminated.

In a further aspect, the alkaloid content is measured from leaves. In one aspect the alkaloid content is measured from green leaves. In a further aspect, the alkaloid content is measured from cured leaves, e.g. air-cured, flue-cured, fire-cured or sun-cured leaves. In a further aspect, the alkaloid content is measured from flue-cured leaves. In a further aspect, the alkaloid content is measured from air-cured leaves.

The term "alkaloid content" is used herein to mean the concentration and/or total amount of the entire group of compounds classified as alkaloids or the concentration and/or total amount of one or more compounds classified as alkaloids. Alkaloids typically present in tobacco include nicotine, nornicotine, PON, anatabine, anabasine and myosmine. In some embodiments the content of one or more alkaloids, such as two or more alkaloids, such as three or more alkaloids, such as four or more alkaloids, such as five or more alkaloids, such as all six alkaloids, selected from nicotine, nornicotine, PON, anatabine, anabasine and myosmine is modulated. In some embodiments the content of one or more alkaloids, such as two or more alkaloids, such as three or more alkaloids, such as four or more alkaloids, such as five or more alkaloids, such as all six alkaloids, selected from nicotine, nornicotine, PON, anatabine, anabasine and myosmine is increased. In some embodiments the content of one or more alkaloids, such as two or more alkaloids, such as three or more alkaloids, such as four or more alkaloids, such as five or more alkaloids, such as all six alkaloids, selected from nicotine, nornicotine, PON, anatabine, anabasine and myosmine is decreased. In some embodiments the total alkaloid content of the plant or cell is modulated. In some embodiments the total alkaloid content is increased. In some embodiments the total alkaloid content is increased.

Suitably nicotine content may be modulated. In one embodiment, the nicotine content is reduced. In another embodiment the nicotine content is not modulated but the content of one or more alkaloids selected from nornicotine, PON, anatabine, anabasine and myosmine is modulated. In another embodiment the nicotine content is not modulated but the content of one or more alkaloids selected from nornicotine, PON, anatabine and anabasine is modulated.

Any method known in the art for determining the concentration and/or total content of alkaloids may be used. One preferred method for analysing alkaloid content involves the analysis by gas chromatography-flame ionization detection method (GC-FID).

In one embodiment there is provided a method for producing a plant (e.g. a tobacco plant) or part thereof, a plant propagation material (e.g. a tobacco plant propagation material), a cell (e.g. a tobacco cell), a leaf (e.g. a tobacco leaf), a harvested leaf (e.g. a harvested tobacco leaf), a cut harvested leaf (e.g. a cut harvested tobacco leaf), a processed leaf (e.g. a processed tobacco leaf), a cut and processed leaf (e.g. a cut and processed tobacco leaf), a product comprising said plant or part thereof (e.g. a tobacco product) or combinations thereof obtainable or obtained by a plant of the invention which has modulated alkaloid content, the method comprising modifying said plant to modulate the activity or expression of a gene encoding an SOUL haem-binding protein. The modulated alkaloid content may be determined by comparing the alkaloid content in the plant (e.g. tobacco plant) or part thereof, plant propagation material (e.g. tobacco plant propagation material), a cell (e.g. a tobacco cell), leaf (e.g. tobacco leaf), harvested leaf (e.g. a harvested tobacco leaf), cut harvested leaf (e.g. a cut harvested tobacco leaf), processed leaf (e.g. processed tobacco leaf), cut and processed leaf (e.g. cut and processed tobacco leaf), a product comprising a plant or part thereof of the present invention, e.g. a tobacco product, or combinations thereof with a comparable product.

Suitably the alkaloid content may be modulated in a plant, e.g. a tobacco plant e.g. modified tobacco plant. Suitably the alkaloid content may be modulated in a leaf (e.g. a tobacco leaf e.g. a tobacco leaf from a modified tobacco plant). Suitably the alkaloid content may be modulated in a harvested leaf (e.g. a harvested tobacco leaf from a modified tobacco plant). Suitably the alkaloid content may be modulated in a cut harvested leaf (e.g. a cut harvested tobacco leaf from a modified tobacco plant). Suitably the alkaloid content may be modulated in a processed leaf (e.g. a processed tobacco leaf e.g. a processed tobacco leaf from a modified tobacco plant). Suitably the alkaloid content may be modulated in a cut and processed leaf (e.g. a cut and processed tobacco leaf e.g. a cut and processed tobacco leaf from a modified tobacco plant). Suitably the alkaloid content may be modulated in a cured leaf (e.g. cured a tobacco leaf from a modified tobacco plant). Suitably the alkaloid content may be modulated in an extract of a green leaf (e.g. a green tobacco leaf from a modified tobacco plant). Suitably the alkaloid content may be modulated in a product comprising the plant of the present invention or part thereof (e.g. a tobacco product, for example a tobacco product produced from a modified tobacco plant or part thereof). Suitably the alkaloid content may be modulated in any one of the above products or combinations thereof. Suitably the modulation of alkaloid content described above may be an increase in alkaloid content. Suitably the modulation of alkaloid content described above may be a decrease in alkaloid content (e.g. a decrease in nicotine content).

In one embodiment the content of one or more alkaloids selected from nicotine, nornicotine, PON, anatabine, anabasine and myosmine is decreased. In one embodiment the content of one or more alkaloids selected from nicotine, nornicotine, PON, anatabine and anabasine is decreased.

- 5 Suitably the modulation of alkaloid content described above may be a decrease in nicotine content. Suitably the modulation of alkaloid content described above may be a decrease in the content of one or more alkaloids selected from nornicotine, PON, anatabine, anabasine and myosmine but not a decrease in the content of nicotine. Suitably the modulation of alkaloid content described above may be a decrease in the content of one or more alkaloids selected
10 from nornicotine, PON, anatabine and anabasine but not a decrease in the content of nicotine.

In one embodiment the nicotine content of a modified plant (e.g. tobacco plant), plant propagation material (e.g. tobacco plant propagation material), leaf (e.g. tobacco leaf), harvested leaf (e.g. harvested tobacco leaf), cut harvested leaf (e.g. cut harvested tobacco leaf), processed leaf (e.g. processed tobacco leaf), cut and processed leaf (e.g. cut and processed
15 tobacco leaf) or tobacco product from a modified tobacco plant is decreased.

In one embodiment the alkaloid content of a plant (e.g. tobacco plant) or part thereof may be modulated by at least 0.5, 1.5 or 2 fold when compared to the alkaloid content of a plant (e.g. tobacco plant) or part thereof, respectively, which has not been modified to modulate the activity or expression of at least one gene encoding a SOUL haem-binding protein and which has been
20 grown under similar growth conditions. Suitably the alkaloid content may be modulated by about 0.5 fold to about 2 fold. Suitably the modification may be an increase or a decrease in alkaloid content. Suitably the modulation may be of one or more alkaloids selected from nicotine, nornicotine, PON, anatabine, anabasine and myosmine. Suitably the modulation may be of one or more alkaloids selected from nicotine, nornicotine, PON, anatabine and anabasine. Suitably,
25 the nicotine content is modulated.

In one embodiment of the invention the alkaloid content of a plant (e.g. a tobacco plant) or part thereof may be modulated by at least 1%, 2%, 5%, 8%, 10%, 12%, 15%, 20%, 25%, 30%, 40%, 50%, 60 %, 70%, 80%, 90% or 100% in comparison to a plant (e.g. a tobacco plant) or part thereof which has not been modified according to the present invention. In one embodiment the
30 alkaloid content may be modulated by at least 30% in comparison to an unmodified plant or part thereof. In one embodiment the alkaloid content may be modulated by at least 40% in comparison to an unmodified plant or part thereof. In one embodiment the alkaloid content may be modulated by at least 50% in comparison to an unmodified plant or part thereof. In one embodiment the alkaloid content may be modulated by at least 60% in comparison to an
35 unmodified plant or part thereof. The modulation may be an increase or a decrease in alkaloid content when compared to an unmodified plant (e.g. a tobacco plant) or part thereof. Suitably

the modulation may be of total alkaloid content. Suitably the modulation may be of one or more alkaloids selected from nicotine, nor nicotine, PON, anatabine, anabasine and myosmine. Suitably the modulation may be of one or more alkaloids selected from nicotine, nor nicotine, PON, anatabine and anabasine. Suitably the modulation is of nicotine content, such as
5 decrease in nicotine content. Suitably the modulation is of nor nicotine content, such as decrease in nor nicotine content. Suitably the modulation is of anabasine content, such as decrease in anabasine content. Suitably the modulation is of PON content, such as decrease in
10 PON content. Suitably the modulation is of anatabine content, such as decrease in anatabine content. Suitably the modulation is of more than one alkaloid, such as two or more alkaloids, such as three or more alkaloids, such as four or more alkaloids, such as five or more alkaloids, such as all six alkaloids, selected from nicotine, nor nicotine, PON, anatabine, anabasine and myosmine.

In some embodiments the alkaloid content of the plant may be modulated by between about 5% and about 100%, by between about 10% and about 90%, by between about 20% and about
15 80%, by between about 30% and about 70%, by between about 40% and 60%, or by between about 40% and 50%.

In one embodiment the alkaloid content of the plant may be modulated by between about 50% and about 100%, preferably between about 60% and 90%, preferably 80% or more.

Tobacco-specific nitrosamine (TSNA) content

20 In one embodiment the present invention provides a method of modulating (i.e. increasing or reducing) the content of tobacco-specific nitrosamine (TSNA) or a precursor of a TSNA in a plant (e.g. a tobacco plant) or a part thereof. Suitably, the method may comprise modifying said plant by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein. Suitably, the content of TSNA or a precursor of a TSNA may be reduced.
25 Suitably, the total TSNA content of the plant or part thereof may be modulated.

The TSNA may be measured in a processed tobacco, e.g. cured tobacco or reconstituted tobacco. In one embodiment the TSNA content is measured and/or modified (e.g. reduced) in a cured tobacco plant or part thereof (e.g. in cured tobacco leaf).

The term "tobacco-specific nitrosamine" or "TSNA" as used herein has its usual meaning in the
30 art, namely a nitrosamine which is found only in tobacco products or other nicotine-containing products. Suitably the at least one tobacco-specific nitrosamine may be 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), N'-nitrosonor nicotine (NNN), N'-nitrosoanatabine (NAT) or N-nitrosoanabasine (NAB).

The term "precursor thereto" when used in relation to at least one tobacco-specific nitrosamine
35 refers to one or more chemicals or compounds of a tobacco plant that give rise to the formation of a tobacco-specific nitrosamine or are involved in the nitrosation reaction leading to tobacco-

specific nitrosamine production. Suitably the term “precursor thereto” may refer to nitrate, nitrite or nitric oxide.

The term “modulating” is used herein to mean either increasing or decreasing.

In one embodiment the TSNA is N’nitrosonornicotine (NNN) and/or the precursor is nornicotine.

- 5 In one embodiment the TSNA may be one or more of group selected from: N’-nitrosonornicotine (NNN), N’nitrosoanatabine (NAT), N’-nitrosoanabasine (NAB) and 4-(methyl nitrosamino)-1-(3-pyridyl)-1-butanone (NNK). Suitably the at least one tobacco-specific nitrosamine may be NNK or NNN. In one embodiment the tobacco-specific nitrosamine is NNN.

- 10 In one embodiment the precursor of the TSNA is one or more of the group selected from nornicotine, anabasine, anatabine, and an oxidised derivative of nicotine such as pseudooxynicotine (PON).

In a preferred embodiment the precursor of the TSNA is nornicotine.

- 15 In one embodiment, the precursor of the TSNA may be PON. The precursor of the TSNA (e.g. NNN, NNK, NAB and/or NAT) may be measured in green tobacco leaf, e.g. prior to processing, e.g. prior to curing. In one embodiment the precursor of the TSNA (e.g. NNN, NNK, NAB and/or NAT) is measured and/or modified (e.g. reduced) in a green tobacco leaf, e.g. prior to processing, e.g. prior to curing.

- 20 In one embodiment carrying out a method and or use of the invention results in a reduction of at least one TSNA or a precursor thereto in the modified tobacco plant (or part thereof) when compared to a tobacco plant (or part thereof) which has not been modified in accordance with the present invention.

- 25 The terms “reducing at least one TSNA or precursor thereto” or “reduction of at least one TSNA or precursor thereto” are used herein to mean that the concentration and/or total content of the at least one TSNA or precursor thereto in the product, method or use of the invention is lower in relation to a comparable product, method or use. For example, a comparable tobacco industry product would be derived from a tobacco plant which had not been modified according to the present invention, but in which all other relevant features were the same (e.g. plant species, growing conditions, method of processing tobacco, etc.).

- 30 Any method known in the art for determining the concentration and/or levels of at least one TSNA or precursor thereto may be used. In particular a method such may comprise the addition of deuterium labelled internal standard, an aqueous extraction and filtration, followed by analysis using reversed phase high performance liquid chromatography with tandem mass spectrometry (LC-MS/MS) may be used. Other examples for determining the concentration and/or level of a precursor to a tobacco-specific nitrosamine include a method such as the one
35 detailed in CORESTA recommended method CRM-72: Determination of Tobacco Specific Nitrosamines in Tobacco and Tobacco Products by LC-MS/MS; CRM being developed into

ISO/DIS 21766 or Wagner et al. (2005) Analytical Chemistry 77(4), 1001-1006 all of which are incorporated herein by reference.

Suitably the concentration and/or total content of the at least one tobacco-specific nitrosamine or precursor thereto may be reduced by carrying out a method and/or use of the present invention. Suitably the concentration and/or level of the at least one tobacco-specific nitrosamine or precursor thereto may be reduced in a tobacco plant of the invention (e.g. obtainable or obtained by a method and/or use of the invention) when compared to the concentration and/or level of the at least one tobacco-specific nitrosamine(s) or precursor thereto in a tobacco plant which has not been modified in accordance with present invention.

10 The concentration and/or total content of the at least one tobacco-specific nitrosamine(s) or precursor thereto may be reduced in a tobacco leaf, harvested leaf, processed tobacco leaf, tobacco industry product or combinations thereof obtainable or obtained from a tobacco plant (or part of a tobacco plant or a tobacco cell or cell culture) of the invention when compared with a tobacco leaf, harvested leaf, processed tobacco leaf, tobacco industry product or combinations thereof obtainable or obtained from a tobacco plant (or part of a tobacco plant or a tobacco cell or cell culture) which has not been modified in accordance with the present invention.

Suitably the concentration and/or total content of the at least one tobacco-specific nitrosamine or precursor thereto may be reduced in a processed tobacco leaf.

20 Suitably the concentration and/or level of the at least one tobacco-specific nitrosamine or precursor thereto may be reduced in a tobacco industry product.

In one embodiment the at least one tobacco-specific nitrosamine or precursor thereto may be reduced by at least about 1%, at least about 3%, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40% or at least about 50%. In some embodiments the at least one tobacco-specific nitrosamine or precursor thereto may be reduced by between about 5% and about 50%, by between about 10% and about 50%, by between about 20% and about 50%, by between about 30% and about 50%, or by between about 40% and 50%.

In relation to processed (e.g. cured) tobacco leaf (e.g. cured or reconstituted), the at least one tobacco-specific nitrosamine or precursor thereto may be reduced by between about 5000 ng/g and about 50 ng/g, by between about 4000 ng/g and about 100 ng/g, by between about 3000 ng/g and 500 ng/g or by between 2000 ng/g and 1000 ng/g. In some embodiments the at least one tobacco-specific nitrosamine or precursor thereto may be reduced by at least about 5000 ng/g, at least about 4000 ng/g, at least about 3000 ng/g, at least about 2000 ng/g, at least about 1000 ng/g, at least about 500 ng/g, at least about 100 ng/g or at least about 50 ng/g.

Biomass production

In one aspect, the present invention provides a method of producing a biomass comprising: growing a cell which has been engineered to modulate (e.g. decrease) the activity or expression of a gene encoding a SOUL haem-binding protein under conditions to produce a biomass.

- 5 In one embodiment, the present invention provides a method of producing a biomass having modified (e.g. decreased) concentration and/or total content of nicotine, comprising growing a cell which has been engineered to decrease the activity or expression of at least one gene encoding a SOUL haem-binding protein comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24, or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

The cell may be engineered by any method known in the art to modify the activity or expression of at least one gene encoding a SOUL haem-binding protein. Suitably, the cell may be engineered to express an exogenous gene encoding a SOUL haem-binding protein. Suitably, the cell may be engineered to overexpress a gene encoding a SOUL haem-binding protein. Suitably, the cell may be engineered to decrease the activity or expression of a gene encoding a SOUL haem-binding protein.

Suitably, the biomass may contain a lower concentration and/or total content of nicotine compared with the biomass produced by a comparable cell which has not been modified in accordance with the present invention.

- 25 Suitably the cell for use in biomass production may be a plant cell, such as a tobacco cell. Suitably the cell for use in biomass production may be a yeast cell.

In one embodiment the cell (e.g. yeast cell) may be further modified to comprise one or more sequences that increases nicotinic alkaloid biosynthesis. Suitably these one or more sequences may be incorporated into a nucleic acid construct that is suitable for cell (e.g. yeast cell) transformation. The one or more sequences may be overexpressed in the cell (e.g. yeast cell). The sequences may be selected from one or more of the following genes: MPO (or Methylputrescine Oxidase or MPO1 or MPO2); A622 (or Isoflavone reductase-like protein or Isoflavone reductase homolog or Isoflavone reductase-like protein); BBL (or Berberine bridge enzyme or Berberine bridge enzyme-like or BBE or NBB1); PMT (or Putrescine N-Methyltransferase or putrescine methyltransferase or S-adenosyl-L-methionine:putrescine N-methyltransferase or PMT or PMT1 or PMT2 or PMT3 or PMT4) and QPT (or quinolinate

phosphoribosyltransferase). In one embodiment the sequences may be selected from one or more of the following genes: BBL, A622, PMT and MPO (MPO1 or MPO2). Genes suitable for modification in this way may be taught in US2016032299 for example, which is incorporated herein by reference.

5

Commercially desirable traits

In one embodiment the plants of the present invention have modified (i.e. increased or decreased) total alkaloid content and/or modified (i.e. increased or decreased) content of one or more alkaloids selected from nicotine, nor nicotine, anabasine, myosmine and anatabine and/or reduced nicotine, whilst the flavour characteristics and/or other commercially desirable traits are at least maintained. In one embodiment the plants of the present invention produce leaves of a similar grade and/or quality to plants which have not been modified according to the invention.

In one embodiment the plants of the present invention have reduced nicotine content without a significant change in the flavour characteristics of the plant (e.g. compared with the same plant which has not been modified in accordance with the present invention).

In one embodiment the plants of the present invention have modified (i.e. increased or decreased) alkaloid and/or TSNA content without a significant change (e.g. decrease) in other commercially desirable traits of the plant (e.g. compared with the same plant which has not been modified in accordance with the present invention). In particular the yield of the modified plant is preferably not reduced compared with the same plant which has not been modified in accordance with the present invention.

Therefore in one embodiment the methods and uses of the present invention relate to modifying (i.e. increasing or reducing) total alkaloid content and/or modifying (i.e. increasing or reducing) one or more alkaloids selected from nicotine, nor nicotine, anabasine and anatabine and/or modifying (i.e. increasing or reducing) nicotine content and/or TSNA content, whilst maintaining the flavour characteristics and/or other commercially desirable traits (e.g. yield).

The term "commercially desirable traits" as used herein will include traits such as yield, mature plant height, harvestable leaf number, average node length, cutter leaf length, cutter leaf width, quality (e.g. leaf quality, suitably cured leaf quality), abiotic (for instance drought) stress tolerance, herbicide tolerance and/or biotic (for instance insect, bacteria or fungus) stress tolerance.

Leaf quality may be measured based on colour, texture and aroma of the cured leaf, for example according to United States Department of Agriculture (USDA) grades and standards.

Tobacco grades are evaluated based on factors including, but not limited to, the leaf stalk position, leaf size, leaf colour, leaf uniformity and integrity, ripeness, texture, elasticity, sheen (related with the intensity and the depth of coloration of the leaf as well as the shine),

hygroscopicity (the faculty of the tobacco leaves to absorb and to retain the ambient moisture), and green nuance or cast.

Leaf grade can be determined using standard methods known in the art, for example, using an Official Standard Grade published by the Agricultural Marketing Service of the US Department of Agriculture (7 U.S.C. §511). See, e.g., Official Standard Grades for Burley Tobacco (U.S. Type 31 and Foreign Type 93), effective November 5, 1990 (55 F.R. 40645); Official Standard Grades for Flue-Cured Tobacco (U.S. Types 11, 12, 13, 14 and Foreign Type 92), effective March 27, 1989 (54 F.R. 7925); Official Standard Grades for Pennsylvania Seedleaf Tobacco (U.S. Type 41), effective January 8, 1965 (29 F.R. 16854); Official Standard Grades for Ohio Cigar-Leaf Tobacco (U.S. Types 42, 43, and 44), effective December 8, 1963 (28 F.R. 11719 and 28 F.R. 11926); Official Standard Grades for Wisconsin Cigar-Binder Tobacco (U.S. Types 54 and 55), effective November 20, 1969 (34 F.R. 17061); Official Standard Grades for Wisconsin Cigar-Binder Tobacco (U.S. Types 54 and 55), effective November 20, 1969 (34 F.R. 17061); Official Standard Grades for Georgia and Florida ShadeGrown Cigar-Wrapper Tobacco (U.S. Type 62), Effective April 1971. A USDA grade index value can be determined according to an industry accepted grade index. See e.g. Bowman et al. (1988) Tobacco Science, 32:39-40; Legacy Tobacco Document Library (Bates Document #523267826-523267833, July 1, 1988, Memorandum on the Proposed Burley Tobacco Grade Index); and Miller et al. (1990) Tobacco Intern., 192:55-57 (all foregoing references are incorporated herein in their entirety).

In one aspect, a USDA grade index is a 0-100 numerical representation of federal grade received and is a weighted average of all stalk positions. A higher grade index indicates higher quality. Alternatively, leaf grade may be determined via hyper-spectral imaging. See e.g. WO 2011/027315 (which is incorporated herein by reference).

In one embodiment, a tobacco plant of the present invention provides tobacco of commercially acceptable grade.

Suitably, the tobacco plant of the present invention provides cured tobacco of commercially acceptable grade.

In one embodiment, a tobacco plant of the present invention is capable of producing leaves having a USDA grade index value of at least about 70% of the USDA grade index value of leaves of a comparable plant when grown in similar growth conditions. Suitably, tobacco plants disclosed herein may be capable of producing leaves having a USDA grade index value of at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 98% of the USDA grade index value of a control plant when grown in similar growth conditions. Suitably, tobacco plants disclosed herein may be capable of producing leaves having a USDA grade index value of between 65% and 130%, between 70% and 130%, between 75% and 130%, between 80% and

130%, between 85% and 130%, between 90% and 130%, between 95% and 130%, between 100% and 130%, between 105% and 130%, between 110% and 130%, between 115% and 130%, or between 120% and 130% of the USDA grade index value of a comparable plant.

In one aspect, the tobacco plant of the present invention is capable of producing leaves having a USDA grade index value of at least 50. Suitably, tobacco plants disclosed herein may be capable of producing leaves having a USDA grade index value of 55 or more, 60 or more, 65 or more, 70 or more, 75 or more, 80 or more, 85 or more, 90 or more, and 95 or more.

Unless specified otherwise, used herein, tobacco yield refers to cured leaf yield which is calculated based on the weight of cured tobacco leaves per acre under standard field conditions following standard agronomic and curing practice.

In one aspect, a plant (e.g. a tobacco plant) of the present invention has a yield between 50% and 150%, between 55% and 145%, between 60% and 140%, between 65% and 135%, between 70% and 130%, between 75% and 125%, between 80% and 120%, between 85% and 115%, between 90% and 110%, between 95% and 105%, between 50% and 100%, between 55% and 100%, between 60% and 100%, between 65% and 100%, between 70% and 100%, between 75% and 100%, between 80% and 100%, between 85% and 100%, between 90% and 100%, between 95% and 100%, between 100% and 150%, between 105% and 150%, between 110% and 150%, between 115% and 150%, between 120% and 150%, between 125% and 150%, between 130% and 150%, between 135% and 150%, between 140% and 150%, or between 145% and 150% of the yield of a comparable plant when grown in similar field conditions.

In another aspect, the plant (e.g. a tobacco plant) yield of the present invention is approximately 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, or 3.0 times of the yield of a comparable plant when grown in similar field conditions.

In another aspect, the yield of a tobacco plant of the present invention is comparable to the yield of the flue cured comparable plant when grown in similar field conditions.

In one aspect, a tobacco plant of the present invention provides a yield selected from the group consisting of about between 1200 and 3500, between 1300 and 3400, between 1400 and 3300, between 1500 and 3200, between 1600 and 3100, between 1700 and 3000, between 1800 and 2900, between 1900 and 2800, between 2000 and 2700, between 2100 and 2600, between 2200 and 2500, and between 2300 and 2400 lbs/acre.

In another aspect, a tobacco plant of the present invention provides a yield selected from the group consisting of about between 1200 and 3500, between 1300 and 3500, between 1400 and 3500, between 1500 and 3500, between 1600 and 3500, between 1700 and 3500, between 1800 and 3500, between 1900 and 3500, between 2000 and 3500, between 2100 and 3500, between 2200 and 3500, between 2300 and 3500, between 2400 and 3500, between 2500 and

3500, between 2600 and 3500, between 2700 and 3500, between 2800 and 3500, between 2900 and 3500, between 3000 and 3500, and between 3100 and 3500 lbs/acre.

In a further aspect, a tobacco plant of the present invention provides a yield selected from the group consisting of about between 1200 and 3500, between 1200 and 3400, between 1200 and 3300, between 1200 and 3200, between 1200 and 3100, between 1200 and 3000, between 1200 and 2900, between 1200 and 2800, between 1200 and 2700, between 1200 and 2600, between 1200 and 2500, between 1200 and 2400, between 1200 and 2300, between 1200 and 2200, between 1200 and 2100, between 1200 and 2000, between 1200 and 1900, between 1200 and 1800, between 1200 and 1700, between 1200 and 1600, between 1200 and 1500, and between 1200 and 1400 lbs/acre.

Plant breeding

In one embodiment the present invention provides a method of producing a plant having a modified alkaloid content and/or modified content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA, comprising:

- a. crossing a donor plant having modified nicotine content and/or modified content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA and wherein the activity or expression of at least one gene encoding a SOUL haem-binding protein according to the present invention has been modulated in the donor plant in accordance with the present invention with a recipient tobacco plant that does not have modified nicotine content or modified content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA and possesses commercially desirable traits;
- b. isolating genetic material from a progeny of said donor plant crossed with said recipient plant; and
- c. performing molecular marker-assisted selection with a molecular marker comprising:
 - i. identifying an introgressed region comprising a mutation in a polynucleotide sequence encoding a protein defined in a.

Suitably, the activity or expression of a protein comprising an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22, or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24; or a protein encoded by a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8,

9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 is modulated in the donor plant when compared to a comparable plant.

The molecular marker assisted selection may comprise performing PCR to identify an introgressed nucleic acid sequence comprising a mutation which modulates the activity or expression of a protein comprising the amino acid sequence shown as SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or an amino acid sequence which has at least 80% identity thereto.

Plants

Suitable plants according to the invention include the *Solanaceae* family of plants which include, for example jimson weed, eggplant, mandrake, deadly nightshade (belladonna), capsicum (paprika, chilli pepper), potato and tobacco.

In one embodiment a suitable genus of *Solanaceae* is *Nicotiana*, e.g. *Nicotiana tabacum* or *Nicotiana rustica*.

A suitable species of *Nicotiana* may be *Nicotiana tabacum*. Species of *Nicotiana* may be referred to herein as a tobacco plant, or simply tobacco.

Tobacco plants

The present invention provides methods, uses directed to plants (e.g. tobacco plants) as well as a cell (e.g. a tobacco cell), a plant (e.g. a tobacco plant) and a plant propagation material.

The term "tobacco plant" as used herein refers to a plant in the genus *Nicotiana* that is used in the production of tobacco products. Non-limiting examples of suitable "tobacco" plants include *N. tabacum* and *N. rustica* (for example, *N. tabacum* L., LA B21, LN KY171, TI 1406, Basma, Galpao, Perique, Beinhart 1000-1, and Petico).

The tobacco material can be derived or obtained from varieties of *Nicotiana tabacum* types, commonly known as Burley varieties, flue or bright varieties and dark varieties. In some embodiments, the tobacco material is derived from a Burley, Virginia or a dark tobacco plant. The tobacco plant may be selected from Burley tobacco, rare tobacco, speciality tobacco, expanded tobacco or the like.

The use of tobacco cultivars and elite tobacco cultivars is also contemplated herein. The tobacco plant for use herein may therefore be a tobacco variety or elite tobacco cultivar. Particularly useful *Nicotiana tabacum* varieties include Flue-cured Virginia type, Burley type, and Oriental type.

In some embodiments, the tobacco plant may be, for example, selected from one or more of the following varieties: L. cultivar T.I. 1068, AA 37-1, B 13P, Xanthi (Mitchell-Mor), KT D#3 Hybrid 107, Bel-W3, 79-615, Samsun Holmes NN, F4 from cross BU21 x Hoja Parado, line 97, KTRDC#2 Hybrid 49, KTRDC#4 Hybrid 1 10, Burley 21, PM016, KTRDC#5 KY 160 SI,

KTRDC#7 FCA, KTRDC#6 TN 86 SI, PM021, K 149, K 326, K 346, K 358, K 394, K 399, K 730, KY 10, KY 14, KY 160, KY 17, KY 8959, KY 9, KY 907, MD 609, McNair 373, NC 2000, PG 01, PG 04, P01, P02, P03, RG 11, RG 17, RG 8, Speight G-28, TN 86, TN 90, VA 509, AS44, Banket A1, Basma Drama B84/31, Basma I Zichna ZP4/B, Basma Xanthi BX 2A, Batek, Besuki Jember, C104, Coker 319, Coker 347, Criollo Misionero, PM092, Delcrest, Djebel 81, DVH 405, Galpao Comum, HB04P, Hicks Broadleaf, Kabakulak Ellassona, PM102, Kutsage E1, KY 14 x L8, KY 171, LA BU 21, McNair 944, NC 2326, NC 71, NC 297, NC 3, PVH 03, PVH 09, PVH 19, PVH 21 10, Red Russian, Samsun, Saplak, Simmaba, Talgar 28, PM132, Wislica, Yayaldag, NC 4, TR Madole, Prilep HC-72, Prilep P23, Prilep PB 156/1, Prilep P12-2/1, Yaka JK-48, Yaka JB 125/3, TI-1068, KDH-960, TI-1070, TW136, PM204, PM205, Basma, TKF 4028, L8, TKF 2002, TN 90, GR141, Basma xanthi, GR149, GR153, and Petit Havana.

Non-limiting examples of varieties or cultivars are: BD 64, CC 101, CC 200, CC 27, CC 301, CC 400, CC 500, CC 600, CC 700, CC 800, CC 900, Coker 176, Coker 319, Coker 371 Gold, Coker 48, CD 263, DF91 1, DT 538 LC, Galpao tobacco, GL 26H, GL 350, GL 600, GL 737, GL 939, GL 973, HB 04P, HB 04P LC, HB3307PLC, Hybrid 403LC, Hybrid 404LC, Hybrid 501 LC, K 149, K 326, K 346, K 358, K394, K 399, K 730, KDH 959, KT 200, KT204LC, KY10, KY14, KY 160, KY 17, KY 171, KY 907, KY907LC, KTY14xL8 LC, Little Crittenden, McNair 373, McNair 944, msKY 14xL8, Narrow Leaf Madole, Narrow Leaf Madole LC, NBH 98, N-126, N-777LC, N-7371 LC, NC 100, NC 102, NC 2000, NC 291, NC 297, NC 299, NC 3, NC 4, NC 5, NC 6, NC7, NC 606, NC 71, NC 72, NC 810, NC BH 129, NC 2002, Neal Smith Madole, OXFORD 207, PD 7302 LC, PD 7309 LC, PD 7312 LC 'Periq'e' tobacco, PVH03, PVH09, PVH19, PVH50, PVH51, R 610, R 630, R 7-1 1, R 7-12, RG 17, RG 81, RG H51, RGH 4, RGH 51, RS 1410, Speight 168, Speight 172, Speight 179, Speight 210, Speight 220, Speight 225, Speight 227, Speight 234, Speight G-28, Speight G-70, Speight H-6, Speight H20, Speight NF3, TI 1406, TI 1269, TN 86, TN86LC, TN 90, TN 97, TN97LC, TN D94, TN D950, TR (Tom Rosson) Madole, VA 309, VA359, AA 37-1, B 13P, Xanthi (Mitchell-Mor), Bel-W3, 79-615, Samsun Holmes NN, KTRDC number 2 Hybrid 49, Burley 21, KY 8959, KY 9, MD 609, PG 01, PG 04, P01, P02, P03, RG 1 1, RG 8, VA 509, AS44, Banket A1, Basma Drama B84/31, Basma I Zichna ZP4/B, Basma Xanthi BX 2A, Batek, Besuki Jember, C104, Coker 347, Criollo Misionero, Delcrest, Djebel 81, DVH 405, Galpao Comum, HB04P, Hicks Broadleaf, Kabakulak Ellassona, Kutsage E1, LA BU 21, NC 2326, NC 297, PVH 21 10, Red Russian, Samsun, Saplak, Simmaba, Talgar 28, Wislica, Yayaldag, Prilep HC-72, Prilep P23, Prilep PB 156/1, Prilep P12-2/1, Yaka JK-48, Yaka JB 125/3, TI-1068, KDH-960, TI-1070, TW136, Basma, TKF 4028, L8, TKF 2002, GR141, Basma xanthi, GR149, GR153, Petit Havana. Low converter subvarieties of the above, even if not specifically identified herein, are also contemplated.

The tobacco plant may be a Burley, Flue-cured Virginia, or Oriental.

In one embodiment the plant propagation material may be obtainable from a plant (e.g. a tobacco plant) of the invention.

A “plant propagation material” as used herein refers to any plant matter taken from a plant from which further plants may be produced. Suitably, a plant propagation material may be selected
5 from a seed, plant calli and plant clumps. Suitably the plant propagation material may be a seed. Suitably, the plant propagation material may be plant calli. Suitably the plant propagation material may be plant clumps.

In one embodiment the cell (e.g. tobacco cell), tobacco plant and/or plant propagation material may be obtainable (e.g. obtained) by a method according to the invention.

10 Suitably a tobacco plant according to the present invention may have modulated (e.g. decreased) nicotine content when compared to an unmodified tobacco plant, wherein the tobacco plant has been modified to modulate (e.g. decrease) the activity or expression of at least on one gene encoding a SOUL haem-binding protein.

Suitably a tobacco plant according to the present invention may have modulated (e.g. reduced)
15 content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA when compared to an unmodified tobacco plant, wherein the tobacco plant has been modified to modulate (e.g. increase) the activity or expression of at least on one gene encoding a SOUL haem-binding protein.

In one embodiment the tobacco plant in accordance with the present invention comprises a
20 tobacco cell of the invention.

In another embodiment the plant propagation material may be obtainable (e.g. obtained) from a tobacco plant of the invention.

In one embodiment there is provided the use of a tobacco plant as described herein to breed a tobacco plant.

25 The present invention also provides in another embodiment the use of a tobacco plant of the foregoing embodiments for the production of a tobacco industry product.

In another embodiment there is provided the use of a tobacco plant of the invention to grow a crop.

In one embodiment there is provided the use of a cell as provided for in the foregoing
30 embodiments for production of a tobacco industry product.

In one embodiment the present invention provides a plant cell or cell culture (e.g. in *in vitro* culture).

The tobacco cell culture may be a cell suspension culture. These cells cultured in vitro may be incorporated into a tobacco industry product, e.g. as a substitute for conventional tobacco
35 particles, shreds, fine cut or long cut tobacco lamina, as an additive ingredient or as both a substitute and an additive. Suitably, the cell culture may produce nicotine.

In one embodiment there is provided the use of a cell culture, e.g. a harvested and/or processed cell culture according to the present invention for the production of a tobacco industry product.

The tobacco cells harvested from an *in vitro* culture may be dried, e.g. freeze-dried, for example to produce a powder.

5 In one embodiment the plant cell is a tobacco plant cell.

In one embodiment, the cell culture is a tobacco cell culture. The skilled person will be aware of known methods for establishing *in vitro* cultures of tobacco cells. By way of example only, the following method may be used: collecting seeds from a tobacco plant of interest and sterilising their exterior to eliminate unwanted organisms, planting said seeds to grow a tobacco plant of
10 interest, removing tissue from the tobacco plant (for example, from the tobacco stem) for use as an explant, establishing a callus culture from the tobacco explant, establishing a cell suspension culture from the callus culture, and harvesting culture material (e.g. including tobacco cells) to produce a tobacco cell culture.

The tobacco cells can be harvested by various methods, including filtration, e.g. vacuum
15 filtration. The sample may be washed in the filter by adding water and the remaining liquid removed with the filtration, e.g. vacuum filtration.

The harvested tobacco cell or cell culture may be further processed, e.g. dried, such as air-dried and/or freeze-dried. The harvested tobacco cell or cell culture or dried harvested tobacco cell or cell culture or an extract therefrom may be incorporated into tobacco industry products
20 according to the present invention.

In one embodiment, the present invention provides a plant (e.g. tobacco plant) or part thereof for use in molecular farming. Suitably, a plant or part thereof modified in accordance with the present invention may be used in the manufacture of proteins such as therapeutics e.g. antibiotics, virus like particles, nutraceuticals or small molecules.

25 In one embodiment, the present invention provides a method for the production of proteins (e.g. therapeutic proteins), the method comprising modifying a plant or part thereof capable of producing said protein (e.g. therapeutic protein) by modulating the activity or expression of at least one SOUL haem-binding protein gene encoding an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or
30 orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue of SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a nucleic acid sequence which has at
35 least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24; and

culturing the plant under conditions sufficient to allow the production of said protein (e.g. therapeutic protein).

Products

5 The present invention also provides for products obtainable or obtained from plants according to the present invention. Products are provided which are obtainable or obtained from a plant in which the activity or expression of gene encoding a SOUL haem-binding protein has been modulated.

10 In one embodiment, the product may comprise a construct of the invention which modulates the activity or expression of at least one gene encoding a SOUL haem-binding protein as defined herein. In one embodiment, the product may comprise a construct of the invention which modifies the nucleic acid sequence of at least one gene encoding a SOUL haem-binding protein as defined herein.

15 The present invention also provides for products obtainable or obtained from tobacco according to the present invention.

In one embodiment there is provided the use of a tobacco plant of the invention to produce a tobacco leaf.

Suitably the tobacco leaf may be subjected to downstream applications such as processing.

20 Thus in one embodiment the use of the foregoing embodiment may provide a processed tobacco leaf. Suitably the tobacco leaf may be subjected to curing, fermenting, pasteurising or combinations thereof. In another embodiment the tobacco leaf may be cut. In some embodiments the tobacco leaf may be cut before or after being subjected to curing, fermenting, pasteurising or combinations thereof.

25 In one embodiment the present invention provides a harvested leaf of a tobacco plant of the invention.

In a further embodiment the harvested leaf may be obtainable (e.g. obtained) from a tobacco plant propagated from a propagation material of the present invention.

In another embodiment there is provided a harvest leaf obtainable from a method or use of the present invention.

30 Suitably the harvested leaf may be a cut harvested leaf.

In some embodiments the harvested leaf may comprise viable tobacco cells. In other embodiments the harvested leaf may be subjected to further processing.

There is also provided a processed tobacco leaf.

35 The processed tobacco leaf may be obtainable from a tobacco plant of the invention. Suitably the processed tobacco leaf may be obtainable from a tobacco plant obtained in accordance with any of the methods and/or uses of the present invention.

In another embodiment the processed tobacco leaf may be obtainable from a tobacco plant propagated from a tobacco plant propagation material according to the present invention.

The processed tobacco leaf of the present invention may be obtainable by processing a harvested leaf of the invention.

- 5 The term “processed tobacco leaf” as used herein refers to a tobacco leaf that has undergone one or more processing steps to which tobacco is subjected to in the art. A “processed tobacco leaf” comprises no or substantially no viable cells.

The term “viable cells” refers to cells which are able to grow and/or are metabolically active. Thus, if a cell is said to not be viable, also referred to as “non-viable” then a cell does not display the characteristics of a viable cell.

- 10 The term “substantially no viable cells” means that less than about 5% of the total cells are viable. Preferably, less than about 3%, more preferably less than about 1%, even more preferably less than about 0.1% of the total cells are viable.

- In one embodiment the processed tobacco leaf may be processed by one or more of: curing, fermenting and/or pasteurising.

Suitably the processed tobacco leaf may be processed by curing.

Tobacco leaf may be cured by any method known in the art. In one embodiment tobacco leaf may be cured by one or more of the curing methods selected from the group consisting of: air curing, fire curing, flue curing and sun curing.

- 20 Suitably the tobacco leaf may be air cured.

Typically air curing is achieved by hanging tobacco leaf in well-ventilated barns and allowing to dry. This is usually carried out over a period of four to eight weeks. Air curing is especially suitable for burley tobacco.

- Suitably the tobacco leaf may be fire cured. Fire curing is typically achieved by hanging tobacco leaf in large barns where fires of hardwoods are kept on continuous or intermittent low smoulder and usually takes between three days and ten weeks, depending on the process and the tobacco.

- In another embodiment the tobacco leaf may be flue cured. Flue curing may comprise stringing tobacco leaves onto tobacco sticks and hanging them from tier-poles in curing barns. The barns usually have a flue which runs from externally fed fire boxes. Typically this results in tobacco that has been heat-cured without being exposed to smoke. Usually the temperature will be raised slowly over the course of the curing with the whole process taking approximately 1 week.

- Suitably the tobacco leaf may be sun cured. This method typically involves exposure of uncovered tobacco to the sun.

Suitably the processed tobacco leaf may be processed by fermenting.

Fermentation can be carried out in any manner known in the art. Typically during fermentation, the tobacco leaves are piled into stacks (a bulk) of cured tobacco covered in e.g. burlap to retain moisture. The combination of the remaining water inside the leaf and the weight of the tobacco generates a natural heat which ripens the tobacco. The temperature in the centre of the bulk is monitored daily. In some methods every week, the entire bulk is opened. The leaves are then removed to be shaken and moistened and the bulk is rotated so that the inside leaves go outside and the bottom leaves are placed on the top of the bulk. This ensures even fermentation throughout the bulk. The additional moisture on the leaves, plus the actual rotation of the leaves themselves, generates heat, releasing the tobacco's natural ammonia and reducing nicotine, while also deepening the colour and improving the tobacco's aroma. Typically the fermentation process continues for up to 6 months, depending on the variety of tobacco, stalk position on the leaf, thickness and intended use of leaf.

Suitably the processed tobacco leaf may be processed by pasteurising. Pasteurising may be particularly preferred when the tobacco leaf will be used to make a smokeless tobacco industry product, most preferably snus.

Tobacco leaf pasteurisation may be carried out by any method known in the art. For example pasteurisation may be carried out as detailed in J Foulds, L Ramstrom, M Burke, K Fagerstrom. Effect of smokeless tobacco (snus) on smoking and public health in Sweden Tobacco Control (2003) 12: 349–359, the teaching of which is incorporated herein by reference.

During the production of snus, pasteurisation is typically carried out by a process in which the tobacco is heat treated with steam for 24–36 hours (reaching temperatures of approximately 100°C). This results in an almost sterile product and without wishing to be bound by theory one of the consequences of this is believed to be a limitation of further TSNA formation.

In one embodiment the pasteurisation may be steam pasteurisation.

In some embodiments the processed tobacco leaf may be cut. The processed tobacco leaf may be cut before or after processing. Suitably, the processed tobacco leaf may be cut after processing.

In one embodiment, the use of the foregoing embodiment may provide reconstituted tobacco.

In one embodiment, there is provided reconstituted tobacco.

“Reconstituted” as used herein may also be referred to as recon, recycled or homogenized sheet tobacco and refers to tobacco material generated from remnants of tobacco leaf after processing. Reconstituted tobacco allows the production of a consistent, high quality blend and allows the adjustment of the ratio of individual components.

Reconstituted tobacco may be nano fibre recon (nanofibers can be extracted in solid or liquid form), paper making recon (which uses stems, scraps, and midribs, etc. as the raw material) or slurry type recon (which uses a mixture of fines and tobacco stems, ground to power, mixed

with water and vegetable binding agent; the soluble residue is formed to sheets by extracting the water).

Any method known in the art may be used for making reconstituted tobacco, for example see CORESTA Congress, Sapporo, 2012, Smoke Science/Product Technology Groups, SSPT 12 (incorporated herein by reference).

In some embodiments the tobacco plant, harvested leaf of a tobacco plant and/or processed tobacco leaf may be used to extract nicotine. The extraction of nicotine can be achieved using any method known in the art. For example a method for extracting nicotine from tobacco is taught in US 2,162,738 which is incorporated herein by reference.

In one aspect, the present invention provides cured tobacco material made from a tobacco plant or part thereof according to the invention.

In another aspect, the present invention provides a tobacco blend comprising tobacco material made from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. In one aspect, the present invention provides a tobacco blend comprising cured tobacco material according to the present invention.

Suitably, the tobacco blend according to the present invention may comprise approximately 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 10% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 20% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 30% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 40% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 50% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 60% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 70% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 80% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or

cell culture according to the present invention. Suitably, the tobacco blend may comprise approximately 90% tobacco from a tobacco plant or part thereof according to the present invention, or from a tobacco cell or cell culture according to the present invention.

In one aspect, a tobacco blend product of the present invention comprises at least about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 95 percent by dry weight of tobacco cured from a tobacco plant or part thereof according to the present invention, or a tobacco cell or cell culture according to the present invention.

Suitably, the cured tobacco material may be air cured. Suitably, the cured tobacco material may be flue cured. Suitably, the cured tobacco material may be sun cured. Suitably, the cured tobacco material may be fire cured.

A tobacco industry product or smoking article according to the present invention may comprise the tobacco material (e.g. cured tobacco material or reconstituted tobacco material) according to the present invention.

In another aspect the present invention provides a tobacco industry product.

In one embodiment the tobacco industry product according to the present invention may be a blended tobacco industry product. Suitably, the tobacco blend may comprise cured tobacco material according to the present invention.

In one embodiment the tobacco industry product may be prepared from a tobacco plant of the invention or a part thereof.

Suitably the tobacco plant or part thereof may be propagated from a tobacco plant propagation material according to the present invention.

The term “part thereof” as used herein in the context of a tobacco plant refers to a portion of the tobacco plant. Suitably, the “part thereof” may be a leaf, root or stem of a tobacco plant or the flowers. Suitably, the “part thereof” may be a leaf, root or stem of a tobacco plant.

Tobacco industry product

As used herein, the term “tobacco industry product” is intended to include combustible smoking articles such as cigarettes, cigarillos, cigars, tobacco for pipes or for roll-your-own cigarettes, (whether based on tobacco, tobacco derivatives, expanded tobacco, reconstituted tobacco, tobacco substitutes or other smokable material), non-combustible aerosol provision systems such as heating products that release compounds from substrate materials without burning such as electronic cigarettes, tobacco heating products, and hybrid systems to generate aerosol from a combination of substrate materials, for example hybrid systems containing a liquid or gel or solid substrate, as well as aerosolizable substrate materials used within these aerosol provision systems; and aerosol-free delivery articles such as lozenges, gums, patches, articles comprising breathable powders and smokeless tobacco industry

products such as snus and snuff, which aerosol-free delivery articles may or may not deliver nicotine.

In one embodiment the tobacco industry product may be prepared from (e.g. may comprise) a tobacco plant of the invention or a part thereof.

- 5 Suitably the tobacco plant or part thereof may be propagated from a tobacco plant propagation material according to the present invention.

The term “part thereof” as used herein in the context of a tobacco plant refers to a portion of the tobacco plant. Preferably the “part thereof” is a leaf of a tobacco plant.

- 10 In another embodiment the tobacco industry product may be prepared from a harvested leaf of the invention.

In a further embodiment the tobacco industry product may be prepared from a processed tobacco leaf of the invention.

Suitably the tobacco industry product may be prepared from a tobacco leaf processed by one or more of: curing, fermenting and/or pasteurising.

- 15 Suitably the tobacco industry product may comprise a cut tobacco leaf, optionally processed as per the foregoing embodiment.

In another embodiment, the tobacco industry product may be prepared from a tobacco cell or cell culture according to the present invention.

- 20 In another embodiment, the tobacco industry product may be prepared from (e.g. may comprise) a cured tobacco material according to the present invention.

In another embodiment, the tobacco industry product may be prepared from (e.g. may comprise) a tobacco blend according to the present invention.

In one embodiment the tobacco industry product may be a smoking article.

- 25 As used herein, the term “smoking article” can include smokeable products, such as rolling tobacco, cigarettes, cigars and cigarillos whether based on tobacco, tobacco derivatives, expanded tobacco, reconstituted tobacco or tobacco substitutes.

In another embodiment the tobacco industry product may be a smokeless tobacco industry product.

- 30 The term “smokeless tobacco industry product” as used herein refers to a tobacco industry product that is not intended to be smoked and/or subjected to combustion.

Smokeless tobacco industry products (including heat-not-burn materials) may contain tobacco in any form, including dried particles, shreds, granules, powders, or slurry, deposited on, mixed in, surrounded by, or combined with other ingredients in any format, such as flakes, films, tabs, foams, or beads.

- 35 In one embodiment a smokeless tobacco industry product may include snus, snuff, chewing tobacco or the like.

In one embodiment, the tobacco industry product is a combustible smoking article, selected from the group consisting of a cigarette, a cigarillo and a cigar.

In one embodiment, the tobacco industry product comprises one or more components of a combustible smoking article, such as a filter, a filter rod, a filter rod segments, tobacco, a tobacco rod, a tobacco rod segment, a spill, an additive release component such as a capsule, a thread, beads, a paper such as a plug wrap, a tipping paper or a cigarette paper.

In one embodiment, the tobacco industry product is a non-combustible aerosol provision system.

In one embodiment, the tobacco industry product comprises one or more components of a non-combustible aerosol provision system, such as a heater and an aerosolizable substrate.

In one embodiment, the aerosol provision system is an electronic cigarette also known as a vaping device.

In one embodiment the electronic cigarette comprises a heater, a power supply capable of supplying power to the heater, an aerosolizable substrate such as a liquid or gel, a housing and optionally a mouthpiece.

In one embodiment the aerosolizable substrate is contained in a substrate container. In one embodiment the substrate container is combined with or comprises the heater.

In one embodiment, the tobacco industry product is a heating product which releases one or more compounds by heating, but not burning, a substrate material. The substrate material is an aerosolizable material which may be for example tobacco or other non-tobacco products, which may or may not contain nicotine. In one embodiment, the heating product is a tobacco heating product.

In one embodiment, the heating product is an electronic device.

In one embodiment, the tobacco heating product comprises a heater, a power supply capable of supplying power to the heater, an aerosolizable substrate such as a solid or gel material.

In one embodiment the heating product is a non-electronic article.

In one embodiment the heating product comprises an aerosolizable substrate such as a solid or gel material and a heat source which is capable of supplying heat energy to the aerosolizable substrate without any electronic means, such as by burning a combustion material, such as charcoal.

In one embodiment the heating product also comprises a filter capable of filtering the aerosol generated by heating the aerosolizable substrate.

In some embodiments the aerosolizable substrate material may comprise a vapour or aerosol generating agent or a humectant, such as glycerol, propylene glycol, triacetin or diethylene glycol.

In one embodiment, the tobacco industry product is a hybrid system to generate aerosol by heating, but not burning, a combination of substrate materials. The substrate materials may comprise for example solid, liquid or gel which may or may not contain nicotine. In one embodiment, the hybrid system comprises a liquid or gel substrate and a solid substrate. The solid substrate may be for example tobacco or other non-tobacco products, which may or may not contain nicotine. In one embodiment, the hybrid system comprises a liquid or gel substrate and tobacco.

In a further embodiment the tobacco industry product may be a tobacco heating device or hybrid device or e-cigarette or the like.

Typically in tobacco heating devices or hybrid devices, an aerosol is generated by the transfer of heat from a heat source to a physically separate aerosol-forming substrate or material, which may be located within, around or downstream of the heat source. During smoking, volatile compounds are released from the aerosol-forming substrate by heat transfer from the heat source and entrained in air drawn through the smoking article. As the released compounds cool, they condense to form an aerosol that is inhaled by the user.

Aerosol-generating articles and devices for consuming or smoking tobacco heating devices are known in the art. They can include, for example, electrically heated aerosol-generating devices in which an aerosol is generated by the transfer of heat from one or more electrical heating elements of the aerosol-generating device to the aerosol-forming substrate of a tobacco heating device.

Suitably the tobacco heating device may be an aerosol-generating device.

Preferably the tobacco heating device may be a heat-not-burn device. Heat-not-burn devices are known in the art and release compounds by heating, but not burning, tobacco.

An example of a suitable, heat-not-burn device may be one taught in WO2013/034459 or GB2515502 which are incorporated herein by reference.

In one embodiment the aerosol-forming substrate of a tobacco heating device may be a tobacco industry product in accordance with the present invention.

In one embodiment the tobacco heating device may be a hybrid device.

Polynucleotides/polypeptides/constructs

In certain embodiments of the present invention, constructs which modulate activity or expression at least gene encoding a SOUL haem-binding protein may be transformed into plant cells, suitably under the direction of a promoter.

In certain embodiments of the present invention, constructs which decrease (i.e. inhibit) the activity or expression of at least one gene encoding a SOUL haem-binding protein may be transformed into plant cells under the direction of a promoter. For example, the genetic

construct may be a gene editing construct or may comprise an RNAi molecule, which may comprise a small interfering RNA (siRNA) molecule, or a short hairpin loop (shRNA) molecule.

In certain embodiments of the present invention, constructs which increase activity or expression of gene encoding a SOUL haem-binding protein may be transformed into plant cells, suitably under the direction of a promoter e.g. constructs which encode a gene encoding a SOUL haem-binding protein such as an endogenous SOUL haem-binding protein.

Constructs may be introduced into plants according to the present invention by means of suitable vector, e.g. plant transformation vectors. A plant transformation vector may comprise an expression cassette comprising 5'-3' in the direction of transcription, a promoter sequence, a construct sequence targeting gene encoding a SOUL haem-binding protein and, optionally a 3' untranslated, terminator sequence including a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase. The promoter sequence may be present in one or more copies, and such copies may be identical or variants of a promoter sequence as described above. The terminator sequence may be obtained from plant, bacterial or viral genes. Suitable terminator sequences are the pea *rbcS E9* terminator sequence, the *nos* terminator sequence derived from the *nopaline synthase* gene of *Agrobacterium tumefaciens* and the *35S terminator* sequence from cauliflower mosaic virus, for example. A person skilled in the art will be readily aware of other suitable terminator sequences.

The construct of the present invention may also comprise a gene expression enhancing mechanism to increase the strength of the promoter. An example of such an enhancer element is one derived from a portion of the promoter of the pea *plastocyanin* gene, and which is the subject of International Patent Application No. WO 97/20056 which is incorporated herein by reference. Suitable enhancer elements may be the *nos* enhancer element derived from the *nopaline synthase* gene of *Agrobacterium tumefaciens* and the *35S* enhancer element from cauliflower mosaic virus, for example.

These regulatory regions may be derived from the same gene as the promoter DNA sequence or may be derived from different genes, from *Nicotiana tabacum* or other organisms, for example from a plant of the family *Solanaceae*, or from the subfamily *Cestroideae*. All of the regulatory regions should be capable of operating in cells of the tissue to be transformed.

The promoter DNA sequence may be derived from the same gene as the gene of interest, e.g. the gene the promoter is going to direct, for instance a gene encoding a SOUL haem-binding protein according to the invention, a coding sequence used in the present invention or may be derived from a different gene, from *Nicotiana tabacum*, or another organism, for example from a plant of the family *Solanaceae*, or from the subfamily *Cestroideae*.

The expression cassette may be incorporated into a basic plant transformation vector, such as *pBIN 19 Plus*, *pBI 101*, *pKYLX71:35S2*, *pCAMBIA2300* or other suitable plant transformation

vectors known in the art. In addition to the expression cassette, the plant transformation vector will contain such sequences as are necessary for the transformation process. These may include the *Agrobacterium vir* genes, one or more T-DNA border sequences, and a selectable marker or other means of identifying transgenic plant cells.

- 5 The term "expression vector or plant transformation vector" means a construct capable of *in vivo* or *in vitro* expression. Preferably, the expression vector is incorporated in the genome of the organism. In one embodiment the vector of the present invention expresses a protein e.g. a SOUL haem-binding protein as described herein. The term "incorporated" preferably covers stable incorporation into the genome.
- 10 Techniques for transforming plants are well known within the art and include *Agrobacterium-mediated* transformation, for example. The basic principle in the construction of genetically modified plants is to insert genetic information in the plant genome so as to obtain a stable maintenance of the inserted genetic material. A review of the general techniques may be found in articles by Potrykus (*Annu Rev Plant Physiol Plant Mol Biol* [1991] 42:205-225) and Christon (AgroFood-Industry Hi-Tech March/April 1994 17-27), which are incorporated herein by reference.

Typically, in *Agrobacterium-mediated* transformation a binary vector carrying a foreign DNA of interest, i.e. a construct according to the present invention, is transferred from an appropriate *Agrobacterium* strain to a target plant by the co-cultivation of the *Agrobacterium* with explants from the target plant. Transformed plant tissue is then regenerated on selection media, which selection media comprises a selectable marker and plant growth hormones. An alternative is the floral dip method (Clough & Bent, 1998 *Plant J.* 1998 Dec;16(6):735-43, which is incorporated herein by reference) whereby floral buds of an intact plant are brought into contact with a suspension of the *Agrobacterium* strain containing the chimeric gene, and following seed set, transformed individuals are germinated and identified by growth on selective media. Direct infection of plant tissues by *Agrobacterium* is a simple technique which has been widely employed and which is described in Butcher et al. (1980) *Tissue Culture Methods for Plant Pathologists*, eds.: D. S. Ingrams and J.P. Helgeson, 203-208 which is incorporated herein by reference.

- 30 Further suitable transformation methods include direct gene transfer into protoplasts using polyethylene glycol or electroporation techniques, particle bombardment, micro-injection and the use of silicon carbide fibres for example. Transforming plants using ballistic transformation and production of fertile transgenic maize plants by silicon carbide whisker-mediated transformation is taught in Frame et al. (1994) *The Plant Journal* 6(6): 941-948, which is incorporated herein by reference, and viral transformation techniques is taught in, for example, Meyer et al. (1992) *Mol. Gen. Genet.* 231(3): 345-352, which is incorporated herein by reference. The use of cassava
- 35

mosaic virus as a vector system for plants is taught in Meyer et al. (1992) Gene 110: 213-217, which is incorporated herein by reference. Further teachings on plant transformation may be found in EP-A-0449375, incorporated herein by reference.

In a further aspect, the present invention relates to a vector system which carries a construct and introducing it into the genome of an organism, such as a plant, suitably a tobacco plant. The vector system may comprise one vector, but it may comprise two vectors. In the case of two vectors, the vector system is normally referred to as a binary vector system. Binary vector systems are described in further detail in Gynheung et al. (1980) Binary Vectors, Plant Molecular Biology Manual A3, 1-19, which is incorporated herein by reference.

One extensively employed system for transformation of plant cells uses the Ti plasmid from *Agrobacterium tumefaciens* or a Ri plasmid from *Agrobacterium rhizogenes* described by An et al. (1986) Plant Physiol. 81, 301-305 and Butcher et al. (1980) *Tissue Culture Methods for Plant Pathologists* eds.: D. S. Ingrams and J.P. Helgeson, 203-208 which are incorporated herein by reference. After each introduction method of the desired exogenous gene according to the present invention in the plants, the presence and/or insertion of further DNA sequences may be necessary. The use of T-DNA for the transformation of plant cells has been intensively studied and is described in EP-A-120516; Hoekema (1985) The Binary Plant Vector System, Offset-drukkerij Kanters B. B., Amsterdam Chapter V; Fraley et al. Crit. Rev. Plant Sci. 4:1-46; and An et al. (1985) EMBO J 4: 277-284, all incorporated herein by reference.

Plant cells transformed with construct(s) which modulate the activity or expression of at least one gene encoding a SOUL haem-binding protein may be grown and maintained in accordance with well-known tissue culturing methods such as by culturing the cells in a suitable culture medium supplied with the necessary growth factors such as amino acids, plant hormones, vitamins, etc.

The term "transgenic plant" in relation to the present invention includes any plant that comprises a construct which modulates the activity or expression of at least one gene encoding a SOUL haem-binding protein according to the invention. Accordingly a transgenic plant is a plant which has been transformed with a construct according to the invention. Preferably the transgenic plant exhibits modulated alkaloid content and/or modulated TSNA content (or precursor thereof) according to the present invention. The term "transgenic plant" does not cover native nucleotide coding sequences in their natural environment when they are under the control of their native promoter which is also in its natural environment.

In one aspect, a gene encoding a SOUL haem-binding protein, a construct, a plant transformation vector or a plant cell according to the present invention is in an isolated form.

The term "isolated" means that the sequence is at least substantially free from at least one other component with which the sequence is naturally associated in nature and as found in nature.

In one aspect, a gene encoding a SOUL haem-binding protein, a construct, plant transformation vector or a plant cell according to the invention is in a purified form. The term "purified" means in a relatively pure state, e.g. at least about 90% pure, or at least about 95% pure or at least about 98% pure.

5 The term "nucleotide sequence" as used herein refers to an oligonucleotide sequence or polynucleotide sequence, and variant, homologues, fragments and derivatives thereof (such as portions thereof). The nucleotide sequence may be of genomic or synthetic or recombinant origin, which may be double-stranded or single-stranded whether representing the sense or anti-sense strand.

10 The term "nucleotide sequence" in relation to the present invention includes genomic DNA, cDNA, synthetic DNA, and RNA. Preferably it means DNA, more preferably cDNA sequence coding for the present invention.

In a preferred embodiment, the nucleotide sequence when relating to and when encompassed by the *per se* scope of the present invention, i.e. the gene encoding a SOUL haem-binding protein, includes the native nucleotide sequence when in its natural environment and when it is
15 linked to its naturally associated sequence(s) that is/are also in its/their natural environment. For ease of reference, we shall call this preferred embodiment the "native nucleotide sequence". In this regard, the term "native nucleotide sequence" means an entire nucleotide sequence that is in its native environment and when operatively linked to an entire promoter with which it is
20 naturally associated, which promoter is also in its native environment.

The nucleotide sequence for use in the present invention may be present in a vector in which the nucleotide sequence is operably linked to regulatory sequences capable of providing for the expression of the nucleotide sequence by a suitable host organism. The constructs for use in the present invention may be transformed into a suitable host cell as described herein to provide
25 for expression of a polypeptide of the present invention. The choice of vector e.g. a plasmid, cosmid, or phage vector will often depend on the host cell into which it is to be introduced. Vectors may be used *in vitro*, for example for the production of RNA or used to transfect, transform, transduce or infect a host cell.

In some applications, the nucleotide sequence for use in the present invention is operably linked
30 to a regulatory sequence which is capable of providing for the expression of the nucleotide sequence, such as by the chosen host cell. By way of example, the present invention covers a vector comprising the nucleotide sequence of gene encoding a SOUL haem-binding protein as described herein operably linked to such a regulatory sequence, i.e. the vector is an expression vector.

35 The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence

"operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The term "regulatory sequences" includes promoters and enhancers and other expression regulation signals. The term "promoter" is used in the normal sense of the art, e.g. an RNA polymerase binding site. The nucleotide sequence within a construct which encodes gene

5 encoding a SOUL haem-binding protein may be operably linked to at least a promoter.

The term "construct" - which is synonymous with terms such as "cassette" or "vector" - includes a nucleotide sequence for use according to the present invention directly or indirectly attached to a promoter.

10 An example of an indirect attachment is the provision of a suitable spacer group such as an intron sequence, such as the Sh1-intron or the ADH intron, intermediate the promoter and the nucleotide sequence of the present invention. The same is true for the term "fused" in relation to the present invention which includes direct or indirect attachment. In some cases, the terms do not cover the natural combination of the nucleotide sequence coding for the protein ordinarily

15 associated with the wild type gene promoter and when they are both in their natural environment. The construct may even contain or express a marker, which allows for the selection of the genetic construct.

In some embodiments, a promoter may be operably linked to nucleotide sequence in a construct or vector which is used to modulate the concentration and/or total content of nicotine

20 in a cell or cell culture or tobacco plant or part thereof.

In some embodiments the promoter may be selected from the group consisting of: a constitutive promoter, a tissue-specific promoter, a developmentally-regulated promoter and an inducible promoter.

In one embodiment the promoter may be a constitutive promoter.

25 A constitutive promoter directs the expression of a gene throughout the various parts of a plant continuously during plant development, although the gene may not be expressed at the same level in all cell types. Examples of known constitutive promoters include those associated with the cauliflower mosaic virus 35S transcript (Odell JT, Nagy F, Chua NH. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature*. 313 810-2), the rice actin 1 gene (Zhang W, McElroy D, Wu R. (1991). Analysis of rice Act1 5' region activity in transgenic rice plants. *Plant Cell* 3 1155-65) and the maize ubiquitin 1 gene (Cornejo MJ, Luth D, Blankenship KM, Anderson OD, Blechl AE. (1993). Activity of a maize ubiquitin promoter in transgenic rice. *Plant Molec. Biol.* 23 567-81). Constitutive promoters such as the Carnation Etched Ring Virus (CERV) promoter (Hull R, Sadler J, Longstaff M (1986)

35 (CaMV/35S), figwort mosaic virus 35S promoter. The sequence of carnation etched ring virus

DNA: comparison with cauliflower mosaic virus and retroviruses. EMBO Journal, 5(2):3083-3090).

The constitutive promoter may be selected from a: a carnation etched ring virus (CERV) promoter, a cauliflower mosaic virus (CaMV 35S promoter), a promoter from the rice actin 1 gene or the maize ubiquitin 1 gene.

The promoter may be a tissue specific promoter. A tissue-specific promoter is one which directs the expression of a gene in one (or a few) parts of a plant, usually throughout the lifetime of those plant parts. The category of tissue-specific promoter commonly also includes promoters whose specificity is not absolute, i.e. they may also direct expression at a lower level in tissues other than the preferred tissue. Tissue specific promoters include the phaseolin-promoter, legumin b4-promoter, usp-promoter, sbp-promoter, ST-LS1 promoter, B33 (patatin class I promoter).

In another embodiment the promoter may be a developmentally-regulated promoter.

A developmentally-regulated promoter directs a change in the expression of a gene in one or more parts of a plant at a specific time during plant development. The gene may be expressed in that plant part at other times at a different (usually lower) level, and may also be expressed in other plant parts.

In one embodiment the promoter may be an inducible promoter.

An inducible promoter is capable of directing the expression of a gene in response to an inducer. In the absence of the inducer the gene will not be expressed. The inducer may act directly upon the promoter sequence, or may act by counteracting the effect of a repressor molecule. The inducer may be a chemical agent such as a metabolite, a protein, a growth regulator (such as auxin and salicylic acid which activate the OCS promoter), or a toxic element, a physiological stress such as heat, light (such as the soybean SSU promoter), wounding (e.g. the nos, nopaline synthase promoter), or osmotic pressure, or an indirect consequence of the action of a pathogen or pest. A developmentally-regulated promoter might be described as a specific type of inducible promoter responding to an endogenous inducer produced by the plant or to an environmental stimulus at a particular point in the life cycle of the plant. Examples of known inducible promoters include those associated with wound response, such as described by Warner SA, Scott R, Draper J. ((1993) Plant J. 3 191-201), temperature response as disclosed by Benfey & Chua (1989) (Benfey, P.N., and Chua, N-H. ((1989) Science 244 174-181), and chemically induced, as described by Gatz ((1995) Methods in Cell Biol. 50 411-424).

A nucleotide sequence encoding either a protein which has the specific properties as gene encoding a SOUL haem-binding protein as defined herein or a protein which is suitable for modification may be identified and/or isolated and/or purified from any cell or organism producing said protein. Various methods are well known within the art for the identification

and/or isolation and/or purification of nucleotide sequences. By way of example, PCR amplification techniques to prepare more of a sequence may be used once a suitable sequence has been identified and/or isolated and/or purified.

In a yet further alternative, the nucleotide sequence encoding the SOUL haem-binding protein may be prepared synthetically by established standard methods, e.g. the phosphoroamidite method described by Beucage et al. (1981) Tetrahedron Letters 22, 1859-1869 which is incorporated herein by reference, or the method described by Matthes et al. (1984) EMBO J. 3, 801-805 which is incorporated herein by reference. In the phosphoroamidite method, oligonucleotides are synthesised, e.g. in an automatic DNA synthesiser, purified, annealed, ligated and cloned in appropriate vectors.

As used herein, the term "amino acid sequence" is synonymous with the term "polypeptide" and/or the term "protein". In some instances, the term "amino acid sequence" is synonymous with the term "peptide". In some instances, the term "amino acid sequence" is synonymous with the term "enzyme".

The present invention also encompasses the use of sequences having a degree of sequence identity or sequence homology with amino acid sequence(s) of a polypeptide having the specific properties defined herein or of any nucleotide sequence i.e. a SOUL haem-binding protein gene encoding such a polypeptide (hereinafter referred to as a "homologous sequence(s)"). Here, the term "homologue" means an entity having a certain homology with the subject amino acid sequences and the subject nucleotide sequences. Here, the term "homology" can be equated with "identity".

The homologous amino acid sequence and/or nucleotide sequence and/or fragments should provide and/or encode a polypeptide which retains the functional activity and/or enhances the activity of the SOUL haem-binding protein gene. Typically, the homologous sequences will comprise the same active sites etc. as the subject amino acid sequence for instance or will encode the same active sites. Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity. Homologous sequences typically retain functional domains or motifs. Suitably, homologues of SOUL haem-binding proteins may contain a haem-binding site.

In one embodiment, a homologous sequence is taken to include an amino acid sequence or nucleotide sequence which has one, two or several additions, deletions and/or substitutions compared with the subject sequence.

Sequence identity

Sequence identity comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences. % homology or % identity may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence is directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an “ungapped” alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting “gaps” in the sequence alignment to try to maximise local homology.

However, these more complex methods assign “gap penalties” to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. “Affine gap costs” are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons.

Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the Vector NTI (Invitrogen Corp.). Examples of software that can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al. (1999) Short Protocols in Molecular Biology, 4th Ed - Chapter 18), BLAST 2 (see FEMS Microbiol Lett 1999 174(2): 247-50; FEMS Microbiol Lett 1999 177(1): 187-8 and tatiana@ncbi.nlm.nih.gov), FASTA (Altschul et al. 1990 J. Mol. Biol. 403-410) and AlignX for example. At least BLAST, BLAST 2 and FASTA are available for offline and online searching (see Ausubel et al. 1999, pages 7-58 to 7-60).

Although the final % homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity

score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. Vector NTI programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). For some applications, it is preferred to use the default values for the Vector NTI package.

Alternatively, percentage homologies may be calculated using the multiple alignment feature in Vector NTI (Invitrogen Corp.), based on an algorithm, analogous to CLUSTAL (Higgins DG & Sharp PM (1988), Gene 73(1), 237-244). Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

Should gap penalties be used when determining sequence identity, then preferably the following parameters are used for pairwise alignment:

FOR BLAST	
GAP OPEN	0
GAP EXTENSION	0

FOR CLUSTAL	DNA	PROTEIN	
WORD SIZE	2	1	K triple
GAP PENALTY	15	10	
GAP EXTENSION	6.66	0.1	

In one embodiment, CLUSTAL may be used with the gap penalty and gap extension set as defined above. In some embodiments the gap penalties used for BLAST or CLUSTAL alignment may be different to those detailed above. The skilled person will appreciate that the standard parameters for performing BLAST and CLUSTAL alignments may change periodically and will be able to select appropriate parameters based on the standard parameters detailed for BLAST or CLUSTAL alignment algorithms at the time.

Suitably, the degree of identity with regard to a nucleotide sequence is determined over at least 50 contiguous nucleotides, preferably over at least 60 contiguous nucleotides, preferably over at least 70 contiguous nucleotides, preferably over at least 80 contiguous nucleotides, preferably over at least 90 contiguous nucleotides, preferably over at least 100 contiguous nucleotides, preferably over at least 150 contiguous nucleotides, preferably over at least 200 contiguous nucleotides, preferably over at least 250 contiguous nucleotides, preferably over at least 300 contiguous nucleotides, preferably over at least 350 contiguous nucleotides, preferably over at least 400 contiguous nucleotides, preferably over at least 450 contiguous nucleotides, preferably over at least 500 contiguous nucleotides, preferably over at least 550 contiguous

nucleotides, preferably over at least 600 contiguous nucleotides, preferably over at least 650 contiguous nucleotides, or preferably over at least 700 contiguous nucleotides.

Suitably, the degree of identity with regard to a nucleotide, cDNA, cds or amino acid sequence may be determined over the whole sequence.

- 5 The sequences may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent substance. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the secondary binding activity of the substance is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine.

- 10 Conservative substitutions may be made, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar – uncharged	C S T M
		N Q
	Polar – charged	D E
		K R
AROMATIC		H F W Y

- The present invention also encompasses homologous substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue, with an alternative residue) that may occur i.e. like-for-like substitution such as basic for basic, acidic for acidic, polar for polar etc. Non-homologous substitution may also occur i.e. from one class of residue to another or alternatively involving the inclusion of unnatural amino acids such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine ornithine (hereinafter referred to as O), pyriylalanine, thienylalanine, naphthylalanine and phenylglycine.

- 25 Replacements may also be made by unnatural amino acids include; α * and α -disubstituted* amino acids, N-alkyl amino acids*, lactic acid*, halide derivatives of natural amino acids such as trifluorotyrosine*, p-Cl-phenylalanine*, p-Br-phenylalanine*, p-I-phenylalanine*, L-allyl-glycine*, β -alanine*, L- α -amino butyric acid*, L- γ -amino butyric acid*, L- α -amino isobutyric acid*, L- ϵ -amino caproic acid[#], 7-amino heptanoic acid*, L-methionine sulfone[#], L-norleucine*,

L-norvaline*, p-nitro-L-phenylalanine*, L-hydroxyproline[#], L-thioprolin*, methyl derivatives of phenylalanine (Phe) such as 4-methyl-Phe*, pentamethyl-Phe*, L-Phe (4-amino)[#], L-Tyr (methyl)*, L-Phe (4-isopropyl)*, L-Tic (1,2,3,4-tetrahydroisoquinoline-3-carboxyl acid)*, L-diaminopropionic acid[#] and L-Phe (4-benzyl)*. The notation * has been utilised for the purpose of the discussion above (relating to homologous or non-homologous substitution), to indicate the hydrophobic nature of the derivative whereas # has been utilised to indicate the hydrophilic nature of the derivative, #* indicates amphipathic characteristics.

Variant amino acid sequences may include suitable spacer groups that may be inserted between any two amino acid residues of the sequence including alkyl groups such as methyl, ethyl or propyl groups in addition to amino acid spacers such as glycine or β -alanine residues. A further form of variation, involves the presence of one or more amino acid residues in peptoid form, which will be well understood by those skilled in the art. For the avoidance of doubt, "the peptoid form" is used to refer to variant amino acid residues wherein the α -carbon substituent group is on the residue's nitrogen atom rather than the α -carbon. Processes for preparing peptides in the peptoid form are known in the art, for example Simon et al. (1992) PNAS 89(20), 9367-9371 and Horwell (1995) Trends Biotechnol. 13(4), 132-134.

The nucleotide sequences for use in the present invention may include within them synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones and/or the addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the nucleotide sequences described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or life span of nucleotide sequences of the present invention.

The present invention also encompasses sequences that are complementary to the nucleic acid sequences of the present invention or sequences that are capable of hybridising either to the sequences of the present invention or to sequences that are complementary thereto. The term "hybridisation" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" as well as the process of amplification as carried out in polymerase chain reaction (PCR) technologies.

The present invention also relates to nucleotide sequences that can hybridise to the nucleotide sequences of the present invention (including complementary sequences of those presented herein). Preferably, hybridisation is determined under stringency conditions (e.g. 50°C and 0.2xSSC {1xSSC = 0.15 M NaCl, 0.015 M Na₃citrate pH 7.0}). More preferably, hybridisation is determined under high stringency conditions (e.g. 65°C and 0.1xSSC {1xSSC = 0.15 M NaCl, 0.015 M Na₃citrate pH 7.0}).

A review of the general techniques used for transforming plants may be found in articles such as Potrykus et al. (1991) *Annu Rev Plant Physiol. Plant Mol. Biol.* 42:205-225 and Christou et al. (1994) *Agro-Food-Industry Hi-Tech* March/April 17-27, which are incorporated herein by reference. Further teachings on plant transformation may be found in EP-A-0449375, incorporated herein by reference.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Singleton, et al., *DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY*, 20 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 a general dictionary of many of the terms used in this disclosure.

This disclosure is not limited by the exemplary methods and materials disclosed herein, and any methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of this disclosure. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, any nucleic acid sequences are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

The headings provided herein are not limitations of the various aspects or embodiments of this disclosure 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.

Amino acids are referred to herein using the name of the amino acid, the three letter abbreviation or the single letter abbreviation. The term "protein", as used herein, includes proteins, polypeptides, and peptides. As used herein, the term "amino acid sequence" is synonymous with the term "polypeptide" and/or the term "protein". In some instances, the term "amino acid sequence" is synonymous with the term "peptide". In some instances, the term "amino acid sequence" is synonymous with the term "enzyme".

In the present disclosure and claims, the conventional one-letter and three-letter codes for amino acid residues may be used. The 3-letter code for amino acids as defined in conformity with the IUPACIUB Joint Commission on Biochemical Nomenclature (JCBN). It is also understood that a polypeptide may be coded for by more than one nucleotide sequence due to the degeneracy of the genetic code.

Other definitions of terms may appear throughout the specification. Before the exemplary embodiments are described in more detail, it is to understand that this disclosure is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only,

and is not intended to be limiting, since the scope of the present disclosure will be limited only by the appended claims.

Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within this disclosure. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within this disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in this disclosure.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an enzyme" or "a nitrate reductase" includes a plurality of such candidate agents and equivalents thereof known to those skilled in the art, and so forth.

Advantages

It has been surprisingly found that by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein as taught herein which acts as a positive regulator of nicotine in tobacco, the alkaloid content (e.g. nicotine and/or nornicotine and/or PON and/or anatabine and/or anabasine and/or myosmine content or total alkaloid content) and/or TSNA content of plants can be modulated. Thereby tobacco products with modulated alkaloid (e.g. nicotine and/or nornicotine and/or PON and/or anatabine and/or anabasine and/or myosmine content or total alkaloid content) and/or TSNA content and commercially desirable traits sought after by consumers of tobacco products can be produced.

The present inventors have surprisingly determined a method for modulating the alkaloid content (e.g. nicotine and/or nornicotine and/or PON and/or anatabine and/or anabasine and/or myosmine content or total alkaloid content), and/or TSNA content of a plant (e.g. tobacco plant) by modulating the activity or expression of a gene encoding a SOUL haem-binding protein. Alkaloid (e.g. nicotine and/or nornicotine and/or PON and/or anatabine and/or anabasine and/or myosmine content or total alkaloid content) or TSNA content of a plant (e.g. tobacco plant) may be decreased by decreasing or inhibiting the activity or expression of a gene encoding a SOUL haem-binding protein. Alkaloid (e.g. nornicotine or nicotine content) or TSNA content of a plant (e.g. tobacco plant) may be increased by increasing the activity or expression of gene encoding a SOUL haem-binding protein. Prior to the present invention it had not been

known that modulation of the activity or expression of a gene encoding a SOUL haem-binding protein as described herein could be used to modulate alkaloid (e.g. nicotine and/or nor nicotine and/or PON and/or anatabine and/or anabasine and/or myosmine content or total alkaloid content) and/or TSNA content of a plant (e.g. a tobacco plant).

- 5 The present inventors have determined that the inhibition of the activity or expression of a gene encoding a SOUL haem-binding protein can reduce the alkaloid content (e.g. nicotine and/or nor nicotine and/or PON and/or anatabine and/or anabasine and/or myosmine content or total alkaloid content or total alkaloid content) of the modified plant to a surprisingly low level.

The publications discussed herein are provided solely for their disclosure prior to the filing date
10 of the present application. Nothing herein is to be construed as an admission that such publications constitute prior art to the claims appended hereto.

EXAMPLES**Example 1 – Transient overexpression of a SOUL haem-binding protein increases alkaloid content in leaves****5 Methods and Materials****Cloning****SOUL haem-binding protein expression vector**

The SOUL haem-binding protein gene sequence (SEQ ID No. 3) was amplified from a Gateway™ compatible cDNA library using primers located outside restriction sites flanking the
 10 gene sequence. The gene sequence was then transferred to an expression vector, depicted in Figure 2.

The resulting plasmid was sequenced and transformed into *Agrobacterium tumefaciens* GV3101pMP90 by heat shock and transiently expressed in TN90 leaves.

15

Transient gene expression

Agrobacterium tumefaciens GV3101 strains carrying the construct of interest were grown overnight in Luria-Bertani (LB) medium supplemented with appropriate antibiotics. Cultures were spun down and re-suspended in buffer containing 10 mM MgCl₂, 10 mM MES pH 5.6 and
 20 100 µM acetosyringone to OD₆₀₀=0.6 and incubated for one hour at room temperature. Infiltration was performed with a needleless syringe into TN90 leaves. Samples are taken 5 days post-infiltration.

Tests were performed in three biological replicates.

25 Alkaloid measurement

Relative content of pyridine alkaloids was determined by reversed phase high performance liquid chromatography with tandem mass spectrometry (LC-MS/MS). Chromatographic separation was achieved using a Gemini-NX column (100 mm × 3.0 mm, particle size 3 µm, Phenomenex) and gradient chromatographic separation using 6.5 mM ammonium acetate
 30 buffer (aq) (pH10) and Methanol.

Mass Spectrometer operates in electrospray (ESI) positive mode using scheduled MRM data acquisition. Two MRM transitions were monitored for each analyte and one for the isotope labelled internal standard.

Analyte	Precursor Ion	Daughter Ion (quant/confirm)
Nicotine	163.1	130/106
Nicotine d4	167.1	134.1

76

Anabasine	163.1	80/120
Anatabine	161.1	144/80
Nornicotine	149.1	80/130
Nornicotine d4	153.1	84.1
PON	176.1	106.0/148
PON d4	183.1	110.0

Statistical Analysis

Statistical significances based on one-way ANOVA analyses was performed with Prism 5.01 software (GraphPad Software).

5

Results

Alkaloid content of 5-week-old TN90 leaves expressing the Nitab4.5_0013616g0010.2 construct is shown in Figure 3. Nicotine, Nornicotine and PON content is represented relative to control and comprises three biological replicates analysed by one-way ANOVA and Tukey's multiple-comparison post-test. Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.001.

10

Overexpression of Nitab4.5_0013616g0010.2 leads to a significant increase in nornicotine content in leaves.

15

Conclusions

Nitab4.5_0013616g0010.2 is a positive regulator of alkaloid content, in particular nornicotine and content in leaves and is a regulator of pyridine alkaloids in tobacco.

Example 2 – Virus-induced gene silencing (VIGS) of a SOUL haem-binding protein decreases alkaloid content in leaves

20

Virus-induced gene silencing (VIGS)

For virus induced gene silencing, a 300-nucleotide cDNA fragment (SEQ ID No. 25) was synthesized and cloned with In-Fusion cloning kit into pTV00. The plasmid was then transformed into *A. tumefaciens* GV3101.

25

The TRV vector comprising both TRV RNA1 (SEQ ID No. 26) and TRV RNA2 (SEQ ID No. 27) comprising the targeted nucleotide sequence were separately propagated in *A. tumefaciens*. These cultures were mixed (1:1) and syringe-infiltrated into 2-week-old TN90

30

plants. The silencing effect was assessed five weeks post-virus infection by assessing the expression level of the target gene.

Silencing

5 VIGS assays were performed as previously described (Ratcliff et al. (2001) The Plant Journal 25: 237-245, incorporated herein by reference). Briefly, independent cultures of *A. tumefaciens* GV3101 carrying TRV2 and TRV1 plasmids were propagated overnight in LB medium supplemented with appropriate antibiotics. Cultures were resuspended in VIGS buffer (10 mM morpholineethanesulfonic acid pH 5.6, 10 mM MgCL₂, and 100 µM acetocyringone) adjusting
10 optical density to OD₆₀₀=1, and incubated overnight at room temperature in the dark. These cultures were mixed (1:1) and syringe-infiltrated into 2-week-old TN90 plants. The silencing effect was assessed two weeks post-virus infection by assessing the expression level of the target gene. TRV-Luciferase was used as a negative control and TRV-PDS (reduced chlorophyll content of the silenced leaves) was used as a phenotypic silencing control.

15

Results

Nicotine, nornicotine, anabasine, PON and anatabine content of 5-week-old TN90 leaves expressing the indicated constructs which silence Nitab4.5_0013616g0010.2 are shown in Figure 4. Content is represented relative to control and comprises three biological replicates
20 analysed by t-test. Values are shown as means ± SEM. Asterisks indicate statistical significance of P value ≤ 0.001

Silencing Nitab4.5_0013616g0010.2 leads to a decrease in nicotine, nornicotine, anabasine, PON and anatabine content in leaves.

25

Conclusions

Nitab4.5_0013616g0010.2 is a positive regulator of alkaloid content, in particular nicotine, nornicotine, anabasine, PON and anatabine content in leaves and is a regulator of pyridine alkaloids in tobacco.

30

Example 3 – Homologue testing

The effects of the homologues of SEQ ID No. 1, namely SEQ ID Nos 4, 7, 10, 13, 16, 19 or 22 are tested in transient assays as described in Example 2.

35 The transient expression of SOUL haem-binding binding protein expression vectors for homologues SEQ ID Nos 4, 7, 10, 13, 16, 19 and 22 are shown in Figure 33.

Transient overexpression of all tested homologues led to a significant increase in nornicotine content.

Example 4 – Nitab4.5_0013616g0010.2 antisense leads to a significant reduction in nornicotine content

Nitab4.5_0013616g0010.2_AS expression vector

A Nitab4.5_0013616g0010.2_AS (antisense) construct was generated by two-step amplification and Gateway™ cloning, using the Nitab4.5_0013616g0010.2 expression vector as template, a first set of gene-specific primers and a second set of Gateway™ compatible primers. The amplification product was inserted into the Gateway™ pDONR™/Zeo vector (ThermoFisher Scientific). The sequence was then transferred to the expression vector.

The resulting plasmid was sequenced and transformed into *Agrobacterium tumefaciens* GV3101pMP90 by heat shock and transiently expressed in TN90 leaves.

Results

Nornicotine content of 5-week-old TN90 leaves expressing the indicated constructs which silence Nitab4.5_0013616g0010.2 is shown in Figure 32. Content is represented relative to control and comprises three biological replicates analysed by one-way ANOVA and Tukey's multiple-comparison post-test. Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.001.

Conclusion

Nitab4.5_0013616g0010.2 is a positive regulator of nornicotine content in leaves.

Example 5 – Stable T1 line overexpressing Nitab4.5_0013616g0010.2 greenhouse

We also validated the function of Nitab4.5_0013616g0010.2 with stable transformation assays. Twenty-four TN90 T1 plants overexpressing Nitab4.5_0013616g0010.2 were grown in the greenhouse. At the 12 leaf stage the nornicotine content was measured as described above.

Results

Nornicotine content of greenhouse-grown TN90 plants at 12 leaf stage overexpressing Nitab4.5_0013616g0010.2 is shown in Figure 34.

Example 6 – Field-grown T1 plants overexpressing Nitab4.5_0013616g0010.2

Twenty-four TN90 T1 plants overexpressing Nitab4.5_0013616g0010.2 were grown in the greenhouse. At the 12 leaf stage the nornicotine content was measured as described above.

5

Results

Nornicotine content of leaves of field-grown TN90 overexpressing Nitab4.5_0013616g0010.2 is shown in Figure 35A. Content is represented relative to control. Results were analysed by t-test.

10 Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.001.

NNN content of leaves of field-grown TN90 leaves overexpressing Nitab4.5_0013616g0010.2 is shown in Figure 35B. Content is represented relative to control. Results were analysed by t-test.

15 Values are shown as means $\pm$ SEM. Asterisks indicate statistical significance of P value $\leq$ 0.01.

Conclusion

Overexpression of Nitab4.5_0013616g0010.2 in T1 plants increases nornicotine content and NNN content.

20

Nitab4.5_0013616g0010.2 is a positive regulator of nornicotine content in leaves.

CLAIMS

1. A method of modulating (e.g. decreasing) the alkaloid content of a plant or a part thereof, or a cell or cell culture, the method comprising modifying said plant or cell culture by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein.
2. A method of modulating (e.g. decreasing) the content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA in a tobacco plant or part thereof, the method comprising modifying said plant or a cell or cell culture by modulating the activity or expression of at least one gene encoding a SOUL haem-binding protein.
3. Use of at least one gene encoding a SOUL haem-binding protein for modulating alkaloid content of a cell or plant or part thereof or a cell or cell culture.
4. A method for producing a plant or part thereof, a cell or cell culture, a plant propagation material, a leaf, a cut harvested leaf, a processed leaf or a cut and processed leaf which has modulated (e.g. decreased) alkaloid content, the method comprising modifying said plant or cell or cell culture to modulate the activity or expression of at least one gene encoding a SOUL haem-binding protein.
5. A method or use according to any one of the preceding claims, wherein the alkaloid content is modulated (e.g. decreased) in comparison to a plant or cell or cell culture which has not been modified to modulate the activity or expression of the at least one gene encoding a SOUL haem-binding protein.
6. A plant or part thereof or a cell or cell culture which has been modified to achieve a modulation (e.g. decrease) in alkaloid content in comparison to an unmodified plant or unmodified cell or cell culture, wherein the modification is the reduction of the activity or expression of at least one gene encoding a SOUL haem-binding protein.
7. A plant propagation material obtainable from a plant according to claim 6 or from a plant or cell or cell culture produced by the method of claim 4 or claim 5.
8. A method or use according to any one of claims 1-5, or a plant or part thereof or a cell or cell culture according to claim 6, or a plant propagation material according to claim 7, wherein the alkaloid content of the plant is decreased in comparison to a plant or cell or cell culture which

has not been modified to modulate the activity or expression of at least one gene encoding a SOUL haem-binding protein.

9. A method or use according to claim 8, a plant or part thereof or cell or cell culture according to claim 8, or a plant propagation material according to claim 8, wherein the activity or expression of at least one gene encoding a SOUL haem-binding protein is decreased in comparison to a plant or cell or cell culture which has not been modified to modulate the activity or expression of the at least one gene encoding a SOUL haem-binding protein.

10. A method or use according to any one of claims 1-5, or a plant or part thereof or a cell or cell culture according to claim 6, or a plant propagation material according to claim 7, wherein the alkaloid content of the plant or cell or cell culture is increased in comparison to a plant which has not been modified to modulate the activity or expression of the at least one gene encoding a SOUL haem-binding protein.

11. A method or use according to claim 10, a plant or part thereof or a cell or cell culture according to claim 10, or a plant propagation material according to claim 10, wherein the plant is modified to increase the activity or expression of at least one gene encoding a SOUL haem-binding protein and the plant or cell or cell culture exhibits increased alkaloid content in comparison to a plant or cell or cell culture which has not been modified to modulate the activity or expression of the at least one gene encoding a SOUL haem-binding protein.

12. A method or use according to any one of claims 1-5 or 8-11, a plant or part thereof or a cell or cell culture according to claim 6 or 8-11, or a plant propagation material according to claims 7-11, wherein the total alkaloid content of the plant or cell or cell culture is modulated (e.g. decreased).

13. A method or use according to any one of claims 1-5 or 8-12, a plant or part thereof or a cell or cell culture according to claim 6 or 8-12, or a plant propagation material according to claims 7-12, wherein the content of one or more alkaloids selected from nicotine, nornicotine, PON, anabasine, myosmine and anatabine is modulated (e.g. decreased), preferably the content of nicotine, nornicotine and/or PON is modulated (e.g. decreased).

14. A method or use according to any one of claims 1-5 or 8-13, a plant or part thereof or a cell or cell culture according to claim 6 or 8-13, or a plant propagation material according to claims 7-13, wherein the plant or plant cell is from the *Solanaceae* family.

15. A method or use according to any one of claims 1-5 or 8-14, a plant or part thereof or a cell or cell culture according to claim 6 or 8-14, or a plant propagation material according to claims 7-14, wherein the plant or plant cell is from the *Nicotiana* genus.

5

16. A method or use according to claim 15, a tobacco plant or part thereof or a tobacco cell or cell culture according to claim 15, or a plant propagation material according to claim 15 wherein the nicotine content is modulated.

10 17. A method or use according to claim 16, a tobacco plant or part thereof or a tobacco cell or cell culture according to claim 16, or a plant propagation material according to claim 16 wherein the nicotine content is decreased.

15 18. A method or use according to any one of claims 1-5 or 8-17, a plant or part thereof or a cell or cell culture according to claims 6 or 8-17, or a plant propagation material according to claims 7-17 wherein the at least one gene encoding a SOUL haem-binding protein encodes a polypeptide which comprises an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or

20 the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

25 19. A method or use according to any one of claims 1-5 or 8-18, a plant or part thereof or cell or cell culture according to claim 6 or 8-18, or a plant propagation material according to claims 7-18 wherein an additional gene encoding a SOUL haem-binding protein is also modulated, wherein the additional gene encodes a polypeptide which comprises an amino acid sequence as set out in SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional
30 fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or

the additional gene comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9,
35 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

20. Use of a plant or part thereof or cell or cell culture according to any one of claims 6 or 8-19, or of a plant produced by the method of any one of claims 4, 5 or 8-19 to breed a plant.

21. Use of a plant or part thereof or a cell or cell culture according to any one of claims 6 or 8-19, or of a plant produced by the method of any one of claims 4, 5 or 8-19 for production of a product.

22. Use of a plant or part thereof according to any one of claims 6 or 8-19, or of a plant produced by the method of any one of claims 4, 5 or 8-20 to grow a crop.

10

23. Use of a plant or part thereof according to any one of claims 6 or 8-19, or of a plant produced by the method of any one of claims 4, 5 or 8-19 to produce a leaf.

24. A harvested leaf of a plant according to any one of claims 6 or 8-19, or obtainable from a plant propagated from a propagation material according to any one of claims 6-19, or obtainable from a plant obtained by a use according to any one of claims 2 or 20-21, or obtainable from a plant produced by the method of any one of claims 4, 5 or 8-19.

15

25. A harvested leaf of a plant according to claim 24, wherein the harvested leaf of a plant is a cut harvested leaf.

20

26. A processed leaf, preferably a processed tobacco leaf, preferably a non-viable processed tobacco leaf:

obtainable from a plant obtainable from a use according to any one of claims 2 or 20-22;
obtainable by processing a plant according to any one of claims 5 or 7-19;
obtainable from a plant propagated from a plant propagation material according to any one of claims 6-19; or
obtainable by processing a harvested leaf of a plant according to claim 24 or 25; or
obtainable from a plant produced by the method of any one of claims 4, 5 or 8-19.

25

30

27. A processed leaf according to claim 26, wherein the leaf is processed by curing, fermenting, pasteurising or a combination thereof.

28. A processed leaf according to claim 26 or 27, wherein the processed leaf is a cut processed leaf.

35

29. Cured tobacco material made from a plant or a part thereof according to any one of claims 15-19 or an extract thereof.

30. A tobacco blend comprising said cured tobacco material of claim 29.

5

31. A tobacco industry product prepared from:

a tobacco plant according to any one of claims 15-19, or a part thereof or a tobacco cell or cell culture according to any one of claims 15-19;

10

a tobacco plant or part thereof propagated from a tobacco plant propagation material according to claim 15;

a harvested leaf of a plant according to claim 23 or 24, wherein the plant is tobacco;

a processed leaf according to any one of claims 25-27, wherein the plant is tobacco;

or

a plant produced by the method of claim 15.

15

32. A tobacco industry product according to claim 31, wherein the tobacco product is a combustible smoking article.

20

33. A tobacco industry product according to claim 31, wherein the tobacco product is a smokeless tobacco product.

34. A tobacco product according to claim 31, wherein the tobacco product is a non-combustible aerosol provision system such as a tobacco heating device or an aerosol-generating device.

25

35. Use of a tobacco cell according to claim 15, for modulating alkaloid content in cell or cell cultures.

30

36. A combustible smoking article, non-combustible aerosol provisioning system, smokeless tobacco product or tobacco heating device comprising a plant or a part thereof according to any one of claims 6-19 or an extract (e.g. a tobacco extract) thereof or a tobacco cell or cell culture according to any one of claims 15-19; or a cured tobacco material according to claim 29; or a tobacco blend according to claim 31.

35

37. Use of a nucleotide sequence of at least one gene encoding a SOUL haem-binding protein to select a plant having modulated (e.g. decreased) alkaloid content and/or modulated (e.g. decreased) content of tobacco specific nitrosamine (TSNA) or a precursor of a TSNA, preferably

wherein the sequence of the SOUL haem-binding protein is selected from SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 1, 4, 7, 10, 13, 16, 19 or 22; or wherein the at least one gene encoding a SOUL haem-binding protein comprises a nucleotide sequence as set out in SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24.

38. A mutant of a plant carrying a heritable mutation in a nucleotide sequence of at least one gene encoding a SOUL haem-binding protein, preferably wherein the gene is selected from SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24; wherein said heritable mutation modulates (e.g. decreases) the activity or expression of the at least one gene encoding a SOUL haem-binding protein and wherein the mutant plant has modulated (e.g. decreased) alkaloid content and/or modulated content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA relative to a comparable plant which does not carry said heritable mutation.

39. Progeny or seed of a mutant plant which carries the heritable mutation according to claim 38.

40. A harvested leaf, a processed leaf or cured tobacco material produced from a plant comprising a modification in a nucleotide sequence of at least one gene encoding a SOUL haem-binding protein, wherein the at least one gene is selected from SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24 or a functional variant or functional fragment or orthologue thereof, or a sequence which has at least 80% identity to SEQ ID No. 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, 23 or 24; or a functional variant or functional fragment or orthologue thereof; wherein said modification modulates (e.g. decreases) the activity or expression of the at least one gene encoding a SOUL haem-binding protein and wherein said plant has modulated (e.g. decreased) alkaloid content and/or modulated content of a tobacco specific nitrosamine (TSNA) or a precursor of a TSNA relative to a comparable plant which does not carry said modification in the at least one gene encoding a SOUL haem-binding protein.

41. A method, a leaf, a plant, a plant propagation material, a harvested leaf, a processed tobacco, a tobacco product, a use or a combination thereof as described herein with reference to the description and drawings.

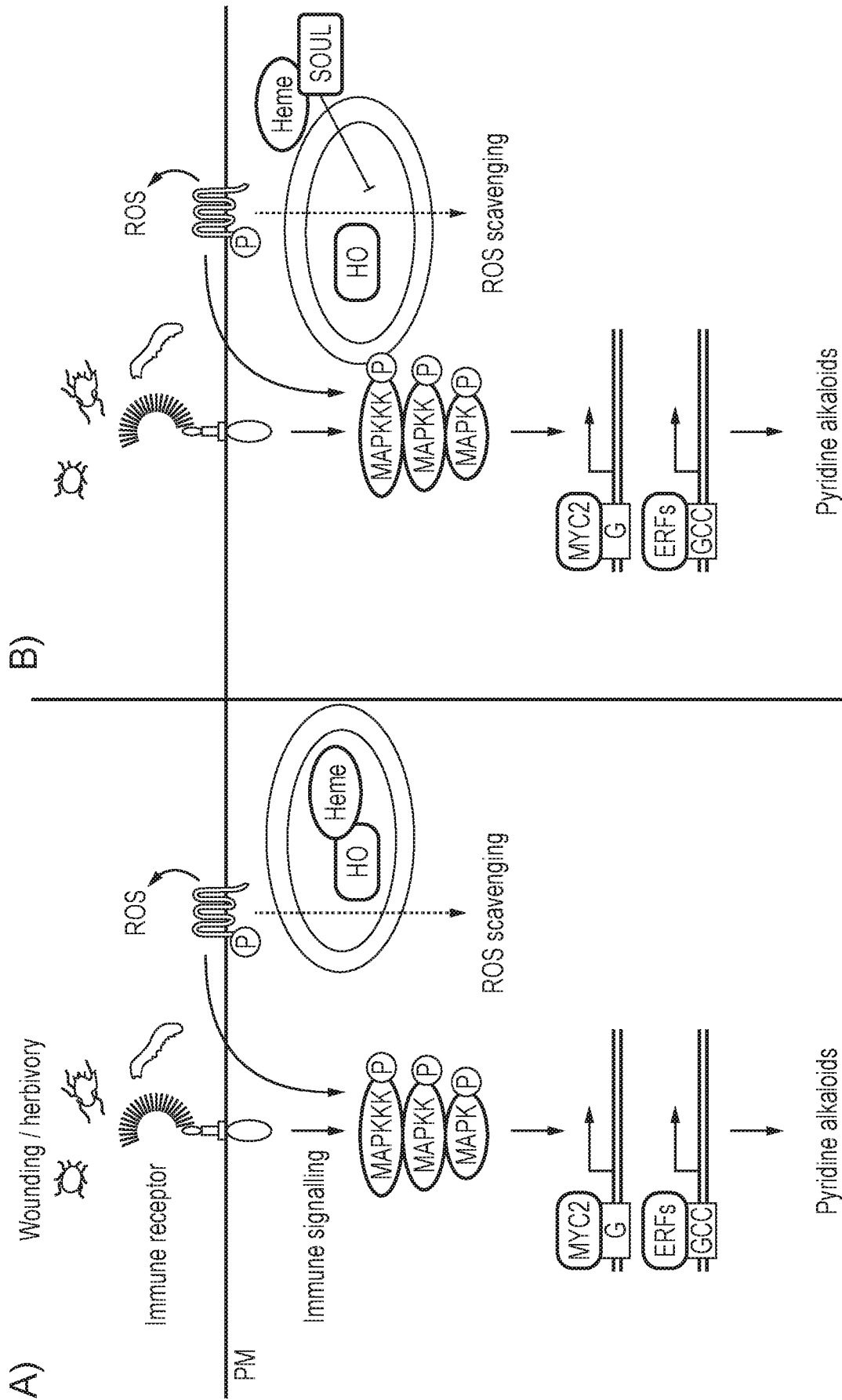
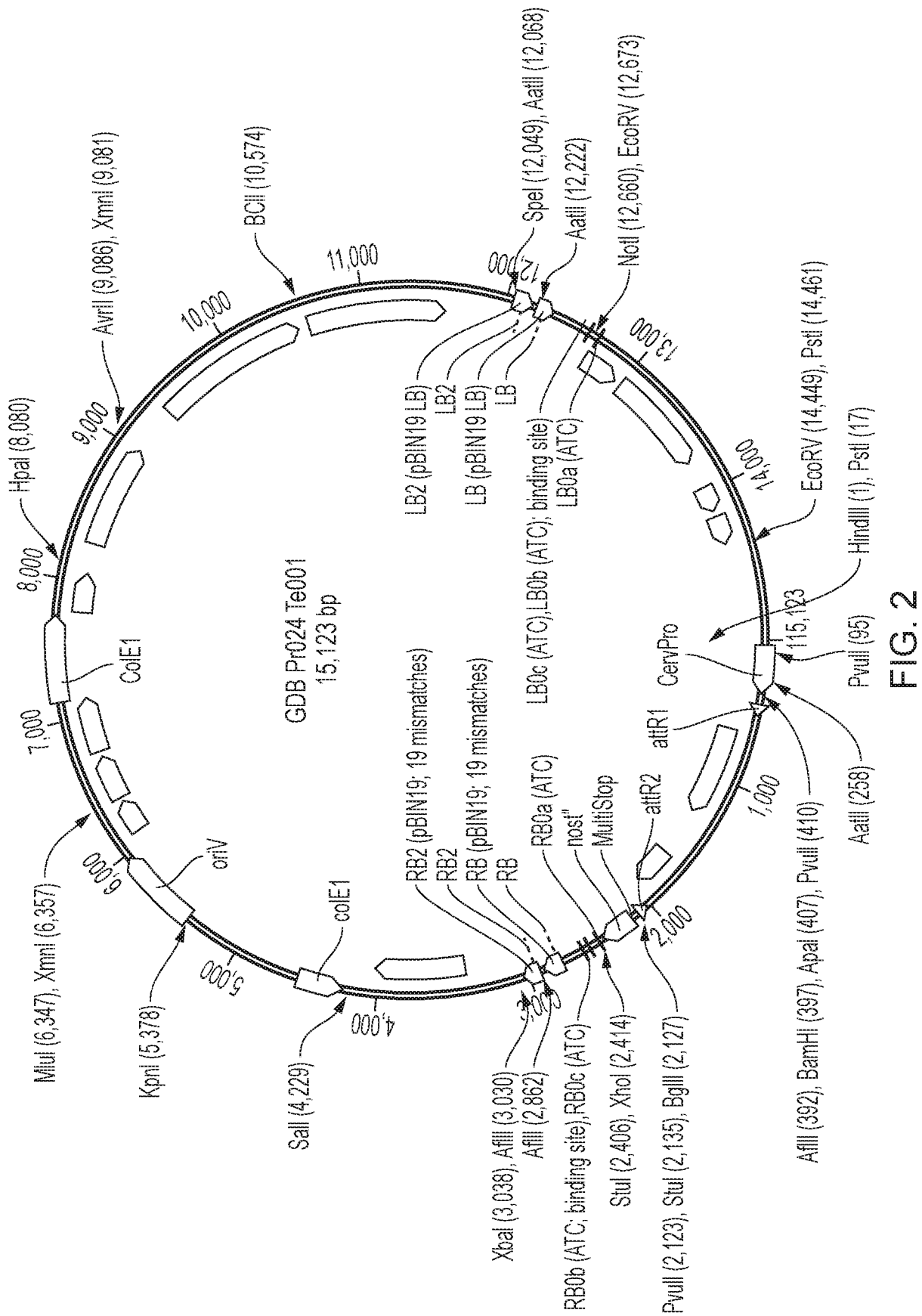


FIG. 1



3/31

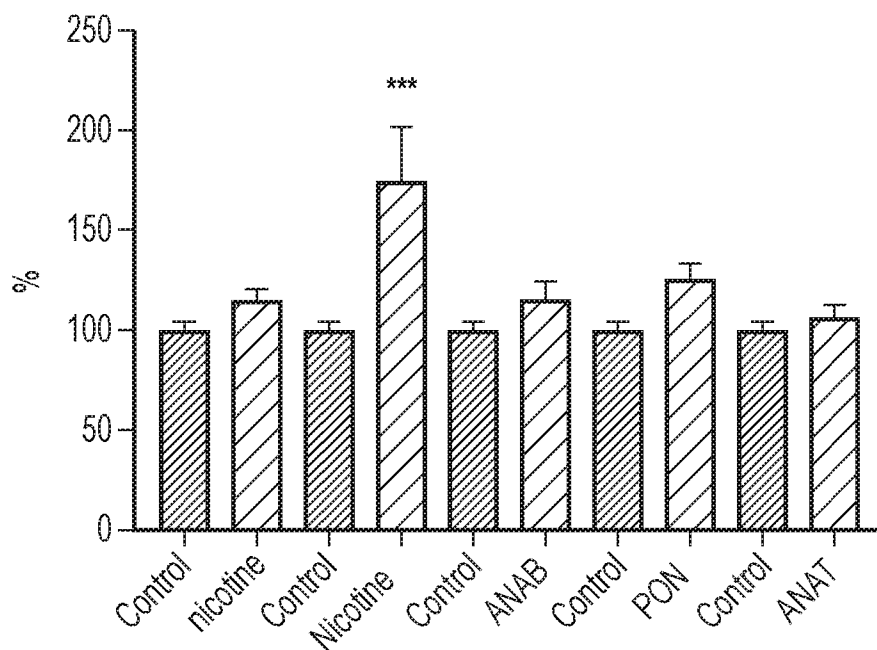


FIG. 3

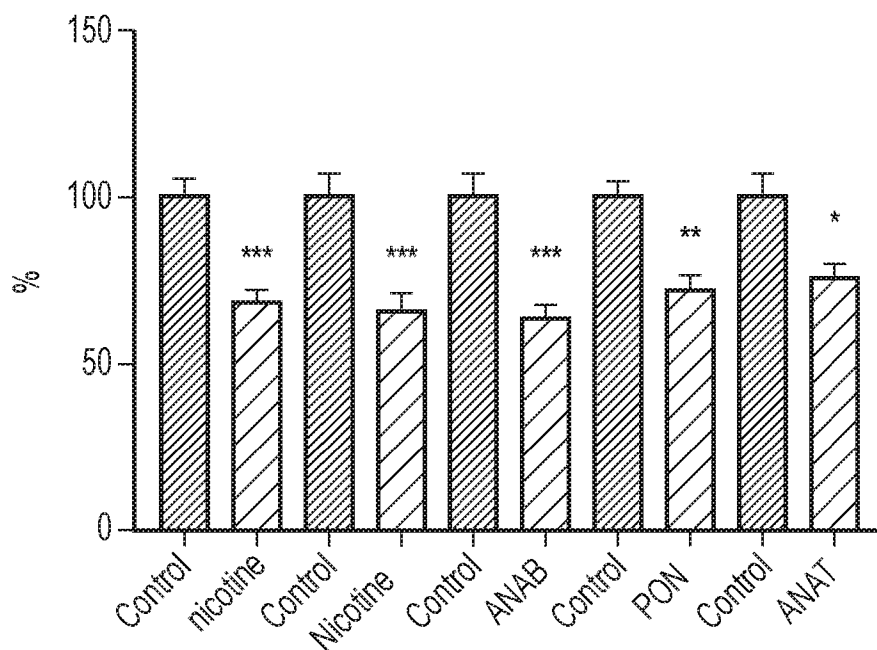


FIG. 4

4/31

Nitab4.5_0013616g0010.2 protein sequence (SEQ ID No. 1):

MLGKYGFDNFNGASQSFNTLAEYLF GKNTKKESMAMTTPVITRRTQSDGEKMEMTTTSVITKR
VEDQGKWRMSFVMP SKYGSDLPLPKDSSVTI KEVPRKTVAVVAFSGFVTDEEVKARESRL
CAALKGD AEFVRVKDGASIEVAQYNPPFTLPFTRRNEISLEVERKQE

FIG. 5

Nitab4.5_0013616g0010.2 genomic sequence (SEQ ID No. 2):

TAATATTGTGACTTTAAACAGCCTTACTTTGCGGCGGAGGCTACAATGCTTGGGAAATA
TGGGTTTGATTTTAATGGTGCATCTCAATCCTTCAACACATTGGCAGAGTACTTGTGG
TAAGGTAGTTAAAGGTACTTATAGAGAGTCAAACCTTAGATCGCCTCCAACCACATTGTAA
TGGTGTTCCTTCTTTGGTTATGTTCTCTTTCTTGATGATTTCCAGAACACAAAGAAGGAAA
GTATGGCGATGACAACACCCGTAATCACTCGTAGAACTCAATCTGATGGGGAGAAGAT
GGAAATGACTACTTCAGTGATACTAAAAGGGTACAGATCGTCTTCTGGTCATTCCCTTT
TCCCTTTTAGTTCTTGTGCTGTGTATTTTACTACGTCAATCGTGACATATGGTTTGGTTA
CTAATGCGCATGCATTTTTGTATGACATGGTGTACTCAAAATATTAAGTCTAATCTATT
GAAGTTGAGTCACCACATAACTTGGGAAATGGTAAATGAAAACTACTCTAAATGACC
ACAGGTCGTCAAATGGAAATACAGACTTCACTGATTCTATTGAAAAGAAACATCATCTGG
TTACCTATCACAAGGAAAACATTTAAGCTCTTGATCTGGAATTGATAAAAGAGGACATGT
GCTTTAGATTGAGATTTGAGGATCTTCAATTGTATGCTAGGCCTATACGCCGGGGGAT
GGGGGTACATATAAAACGTAAACATTCTGGATGCTTATAGTTTAAACCTGTAATACTCTC
AGATAGTATTGCAAATCCATATCTGGGAGAGTGGAACCTGTTAGATTCCAGGTTTGATTA
TGTTGCATGGATCCTTCAAACAAGTGTGACATGGGTGTGGGTACTTCTTGCAGGTTCTT
CAAAGTATACTAATAATTAGGAAAAAATGCACAAAACCTCCTATCAGATACTTACACACA
CACAAACACACACACATGGTATATACCTATAGGTTTGTCCATATAATGCGATATCTAATC
TTAGTCTAGAACAAGTGTGCAATGGAATAACAGTCACAAAGAACGATCTTCTTAATGATA
AAGGTTATGGTACGAATAAAATATACACCCCTATATAAACGATAACACATTTTCTTGTAA
TCTTAAGTTAATTTCTGAGAATAACATTTTGTAAACATAAAAAACCAATTTTATGTGAAAT
GGGCCCTCCATATAGATTGGCATCCTATAAGAATTCTTTACTCCCAGAAACCTTGAAAC
CGGATTTAGTCCATTCTAGAAAAGAAAAACTCATTTTAGCTTGATGTATGATCCTATGTT
ATAGGGCTTAAACCCAGAAGAAAAAAGATAAACATTGTGTATATAGTTTTTGTATTTG
GAATGTCGTGCGCACATATTTTCAACCAAAAAATATACCCACTGGAATTTTATTAAGAATT
TCATTAGGTGGAAGATCAAGGAAAGTGGAGGATGTCCTTTGTCATGCCCTCAAAGTATG
GTTCCGACTTGCCACTACCAAAGGATTCCCTCCGTAACCTATCAAGGAGGTGCCTAGGAAA
ACTGTGCGCGTTGTTGCTTTTTTCAAGTTTGTGATTTTCAAGTTTATGATGTCAATATAAA
GTTCTATTTTATATGATGATGATATGAGTTGAGCTTATCGAAAGAAGTGAGGATTCATAG
CGCCGACCCCAACTTTGTGGGACTGAGGCATAGTTGTTGCTGTTAATACAAAGTTCTAG
ACAGATTTGCATTAAAGTTTATTTGAAATTCAATATCCTAGTAATTGGAAGCCGAGCTCT
GTCTCATTGGTCTATAGAAGTAAATCTTGTTTAAAGTTTACTTAGTGCTTTATTGATAGCTT
CTCTTGTTTTGATATGAGAGAACTTGCATCTAGTTGTATTTTCTGTGCCCTTCCCTTGAC
GTGCTATATGCCTGTTGTGGTAGATAGCAATTTCCATTCAGAGTTCAAAGCAAAAACAT
TTAGGAATTCTATTTTCAAGTTTATATCCTAACTTAGCTAGCTTTCTCCCATTTGATTCA
CCAAAGGTTTTGTGACTGATGAAGAAGTTAAAGCCCGAGAATCAAGACTATGTGCTGCG
TTAAAGGGAGATGCAGAGTTTCGGGTAAAAGATGGTGCCTCGATAGAAGTTGCACAGG
TATATTCGTAGAACCATTTCATAGTACTTGTAAATCTCCTGGGGAACAGATTTTGGCCT
AGTGTAAACCCCAAGTGACATACTAGTTGTGTTTTTTTATACTTCTGCTTGCAGTATAAT
CCACCATTCACTCTTCCATTCACACGTCGGAATGAGATTAGCCTGGAAGTTGAAAGGAA
ACAGGAATAGCTAGTCAGGGTGCAACATCTTGCAGATAAATCCGTGCACTGATATATAG
AAACCAGTGAAGCAAAACATATATGGCGTAAATATATTGTGTTGATATACTCCTTGATA
TATATATATGTATATATATATATTTGGTGATTGCTTGACATTTTGGTAACAAGGTTATGTA
CATGCACAGAAATGTAAGGTCACTTATGCTTCAACTCTGAATATAATAACACTGTCTTTA
ATACTCTTCTCGATCCTC

FIG. 6

5/31

Nitab4.5_0013616g0010.2 coding sequence (SEQ ID No. 3):

ATGCTTGGGAAATATGGGTTTGATTTTAATGGTGCATCTCAATCCTTCAACACATTGGCA
GAGTACTTGTGGTAAGAACAACAAAGAAGGAAAGTATGGCGATGACAACACCCGTAAT
CACTCGTAGAATCAATCTGATGGGGAGAAGATGGAAATGACTACTTCAGTGATAACTAA
AAGGGTGGAAAGATCAAGGAAAGTGGAGGATGTCCTTTGTCATGCCCTCAAAGTATGGT
TCGGACTTGCCACTACCAAAGGATTCTCCGTAAGTATCAAGGAGGTGCCTAGGAAAAAC
TGTCGCCGTTGTTGCTTTTTCAGGTTTTGTGACTGATGAAGAAGTTAAAGCCCCGAGAAT
CAAGACTATGTGCTGCGTTAAAGGGAGATGCAGAGTTTCGGGTAAAAGATGGTGCCTC
GATAGAAGTTGCACAGTATAATCCACCATTCACTCTTCCATTACACGTCGGAATGAGA
TTAGCCTGGAAGTTGAAAGGAAACAGGAATAG

FIG. 7

Nitab4.5_0000652g0130.2 protein sequence (SEQ ID No. 4):

MGMILGKICVETPKYELIQSTADYEIRKYPASVIAQVTYDPTQFKGNKDGGFMLLANYIGALG
NPQNAKPETIAMTAPVITKSSEKIAMTAPVVTKNGDGESNMVTMQFILPAKYTKAEEAPKPL
DERVMIKEEGERKFGVVQFSGTASDKAVKEKVENLRKCLERDGYKIIIGDYELARYNPPWTIP
PFKTNEVMIPVE

FIG. 8

Nitab4.5_0000652g0130.2 genomic sequence (SEQ ID No. 5):

CTGATTTTCGTAATGCAGATTGGTTTAACTTTTACAGAGCAAGTCAATTTTAAACGGCGAT
ACTAAAATTACTGTGTGCACAAATAGTGCACAACTATTCTGTATAGGTTTATCTCCAC
CAGCTTATGGTGTAAACGTGGCAAAACAACTCCACAGCATACTTCCCTGATCTGCCGC
CACTCATTTTGCAGCAAACAAACCCCTTTCCCAAGTCTTCTGATCGTCTCTTTCCGAAT
TCCACATGAACTTCTCTTTTCTTCCCTTATATGTAACAATTACAGCCCCCATTTACT
ATCAAAACATTCTCAAATTCATTTTCAATAACTAATTTCCCATTCATCACAAAAATGGG
CATGATTTTGGGTAAAGATCTGTGTGGAAACACCAAATACGAGTTGATTCAATCTACAG
CTGACTACGAAATCCGCAAATACCCAGCATCTGTTATAGCACAAGTCACATATGATCCA
ACCCAGTTCAAAGGAAACAAAGACGGTGGGTTTATGCTATTAGCCAATTACATCGGCGC
ACTGGGCAATCCTCAAACGCTAAGCCTGAAACAATCGCCATGACAGCTCCTGTAATCA
CCAAATCGTCCGAAAAAATCGCGATGACTGCACCAGTAGTGACTAAGAATGGTGACGG
AGAGAGCAATATGGTGACGATGCAGTTTATTTTACCTGCAAAGTATACAAAAGCTGAAG
AGGCACCTAAGCCGTTGGATGAGAGAGTGATGATTAAAGAAGAAGGTGAAAGGAAGTT
TGGTGTGGTGCAGTTTATGTTGGGACTGCAAGTGATAAAGCGGTAAAAGAGAAAGTGGAG
AATTTAAGGAAATGTTTAGAGAGAGATGGGTATAAGATAATTGGTGATTATGAGTTGGCT
AGGTATAATCCTCCTTGGACAATTCCTCCATTTAAGACCAATGAAGTTATGATTCCAGTT
GAGTGAGCTGAAAATGGTGTAGTAAAGAAAAACATGAGACTGCTGTGAAGCTAGCTCGT
GACATGAACCTACTTTATCTGGTGTATTATTGAATTGTGATTTCTTGTTTATATGTTTA
CGCTGTTGACAATGTCTTTCATACACTAGTGTCAATTATTCATAGTTATAATTAGCACAT
ATTTTGTCTTATTTTATACTTTGCCATAGTAAAGCAGTAAAACTGATGCAGTTTTGCAAG
CATCTAATGTG

FIG. 9

Nitab4.5_0000652g0130.2 coding sequence (SEQ ID No. 6):

ATGGGCATGATTTTGGGTAAAGATCTGTGTGGAAACACCAAATACGAGTTGATTCAATC
TACAGCTGACTACGAAATCCGCAAATACCCAGCATCTGTTATAGCACAAGTCACATATG
ATCCAACCCAGTTCAAAGGAAACAAAGACGGTGGGTTTATGCTATTAGCCAATTACATC
GGCGCACTGGGCAATCCTCAAACGCTAAGCCTGAAACAATCGCCATGACAGCTCCTG
TAATACCAAATCGTCCGAAAAAATCGCGATGACTGCACCAGTAGTGACTAAGAATGGT
GACGGAGAGAGCAATATGGTGACGATGCAGTTTATTTTACCTGCAAAGTATACAAAAGC

FIG. 10

6/31

TGAAGAGGCACCTAAGCCGTTGGATGAGAGAGTGATGATTAAAGAAGAAGGTGAAAGG
AAGTTTGGTGTGGTGCAGTTTAGTGGGACTGCAAGTGATAAAGCGGTAAAAGAGAAAAG
TGGAGAATTTAAGGAAATGTTTAGAGAGAGATGGGTATAAGATAATTGGTGATTATGAG
TTGGCTAGGTATAATCCTCCTTGGACAATTCCTCCATTTAAGACCAATGAAGTTATGATT
CCAGTTGAGTGA

FIG. 10 (Continued)

Nitab4.5_0001140g0220.2 protein sequence (SEQ ID No. 7):

MQGVMNFFTRLSSFFLILVSKLSNGHWNNLKLDFYPPTCNRIECPNYDLIQSGKDYAIRRYN
SSMWMSTAPIDDINLYSATRTGFLRLFDYIQGKNSYQEKIEMTAPVITQVKPSDGPFCASSF
VVSFYVPKKNQPNPPPAKGLHVQKWSNTYVAVRQFGGFVADVDVAKEAAALSASIADTKW
AAAVEKSHAADNTTMYTVAGYNPFEFKDRVNEIWFTFDLDKASAI

FIG. 11

Nitab4.5_0001140g0220.2 genomic sequence (SEQ ID No. 8):

TACCAATTCATTCTGGTCAGTTCATCTCTTCCCTATAATTTCTGCTGCAATTAAACCACC
AAAAAGAAACAAAAGATGCAGGGCGTTATGAATTTCTTCACGAGGCTATCTTTCTTTCTC
ATCCTTGTTTCTAAATTAAGCAATGGGCATTGGACAAACAATTTGAAATTGGATTTCTAT
CCTCCAACCTGTAACCGAATTGAGTGCCCAAATTATGACTTGATTCAATCTGGTAAAGA
CTATGAAATTCGTCGTTACAATTCATCCATGTGGATGTCTACTGCACCCATTGATGATAT
TAATCTTTATTCCGCCACCAGAACTGGTTTCCTCAGGTATCTATGGTTTTTTTTCTTCTTT
TTGGTCATGAATTCCTTTCGGTTGTGTTGCATATTTGGATACGTTAGCTACTGTGCAATA
CCAATTAAAGCATGTCTGTTTTCTGTCCAAGATTCCTTAAGAAAAACAATAAGGGTTACT
TACATTTCAATCAATTGAAATTAATCGACTATTAAGAAGAAGAAGAAAAAAGATA
GTTATCTGTAAACGATTTGACTTATATATGCTGATAATGTAAAGTATTCAATTCACCTTTT
TCAAGCACGTAACCTGCGTGGTTTTCAAGATAGTAATCTCACTTTTTATGGATGAATAC
TAATCTCACTTTTTATGGATATGATTATTGAAAGCGTAAAAGTTCTTTTGCAGTGTTAGTC
AGTAGAATATTTTCGGGTGAGCTTGGTTTTTATACTCAAAATGTTAGTGTATCTATTCGC
GAAAACATAAAAGGGGTCGTTGTGAACCTTGCCTTGGCAGGCTATTCGATTACATTCAA
GGGAAGAACAGTTACCAGGAGAAAATAGAGATGACAGCTCCAGTTATCACTCAAGTAAA
ACCAAGTGATGGACCATTTTGTGCATCTTCATTTGTTGTGAGCTTCTATGTACCAAAGAA
GAACCAACCAAATCCTCCTCCAGCTAAAGGCCTTCACGTCCAAAAATGGAGCAATACTT
ATGTGGCCGTCAGGCAATTCGGCGGATTTGTAGCTGATGTTGATGTTGCAAAAAGAGCT
GCTGCTTTGAGTGCTAGTATTGCTGACACTAAATGGGCAGCTGCTGTTGAAAAAGCCA
TGCTGCAGATAACACTACCATGTATACAGTGGCGGGATACAACCTCTCCATTTGAGTTCA
AGGACAGAGTTAATGAGATTTGGTTTACTTTTGATTTGGACAAAGCATCTGCCATTTGAG
AGGCTCTTATTCTGGACTCTTTAACATTTGAGGAGACATGAACTATCTTATTTTCTTGG
GTTTGTATTATTAAACAGCAGTTTTGTATGACCACAACGCTACATTATATATAGCAAATAA
GATAGGTTGTATAATAACAATCTGATTCTGAGGATTTAGTGGTTAAGATACCTCTTTCAAG
ACCTTTCAAATAATCCGATATATATGAAAAGGGCTAGACCATATTGAGTTTATTGTAT
GTACGCAGTTTTACCTTCCATTT

FIG. 12

Nitab4.5_0001140g0220.2 coding sequence (SEQ ID No. 9):

ATGCAGGGCGTTATGAATTTCTTCACGAGGCTATCTTTCTTTCTCATCCTTGTTTCTAAA
TTAAGCAATGGGCATTGGACAAACAATTTGAAATTGGATTTCTATCCTCCAACCTGTAAC
CGAATTGAGTGCCCAAATTATGACTTGATTCAATCTGGTAAAGACTATGAAATTCGTCGT
TACAATTCATCCATGTGGATGTCTACTGCACCCATTGATGATATTAATCTTTATTCCGCC
ACCAGAACTGGTTTTCCTCAGGCTATTCGATTACATTCAAGGGAAGAACAGTTACCAGGA
GAAAATAGAGATGACAGCTCCAGTTATCACTCAAGTAAAACCAAGTGATGGACCATTTT
GTGCATCTTCATTTGTTGTGAGCTTCTATGTACCAAAGAAGAACCAACCAAATCCTCCTC

FIG. 13

7/31

CAGCTAAAGGCCTTCACGTCCAAAAATGGAGCAATACTTATGTGGCCGTCAGGCAATTC
 GGCGGATTTGTAGCTGATGTTGATGTTGCAAAAAGAGCTGCTGCTTTGAGTGCTAGTAT
 TGCTGACACTAAATGGGCAGCTGCTGTTGAAAAAGCCATGCTGCAGATAACACTACCA
 TGTATACAGTGGCGGGATACAACTCTCCATTTGAGTTCAAGGACAGAGTTAATGAGATT
 TGGTTTACTTTTGATTTGGACAAAGCATCTGCCATTTGA

FIG. 13 (Continued)

Nitab4.5_0003235g0060.2 protein sequence (SEQ ID No. 10):

MATSQLSDLIFRPSLHRRTNFRQCHPTSVFLTPPKNIKTCLKYDRKIKWLIKFSLVDKQSPT
 KKPTVDMNQLVEFLYEDLPHLFDDQGIDRKAYDDYVKFRDPITKHDSIDGYLFNIAMLKQLF
 KPDLQLHWAKQTGPYEITTRWTMVMKIFILLPWKPELVFTGTSVMGVNPETNKFNSHVDYW
 DSIKNNDYFSVEGLLEVIKQLRIYKTPDLETPSYQILRRTATYEVRYDPFIVVETEGDKLAGN
 RGFNDVAGYIFGKNSATEKIPMTTPVFTQAFDAEKSQVSIQIVLP SDKSLSSLPAPNQEGISL
 RKTEGGIAAALKFSGKPTDDVVREKEKQLRSSLIKDGLKPQSDCMLARYNDPGRTWKFIMR
 NEVLIWLEDFKLD

FIG. 14

Nitab4.5_0003235g0060.2 genomic sequence (SEQ ID No. 11):

AATGGCCACCTCACAACTTTCCGACCTCATTTTCCGGCCATCACTTCACAGGCGAACCA
 ATTTCCGGCAATGCCACCCTACTTCAGTCTTTCTCACTCCTCCAAAAAATATCAAAACAA
 AAACCTCTTAAATATGACAGAAAAATTAAGTGGTTAATTAAGTTTAGTTAGTTGATAAACA
 GAGCCCAACCAAAAAACCAACAGTTGATATGAACCAATTAGTGAATTTTTATATGAGGA
 TTTACCTCATCTTTTTGATGATCAAGGTATTGATCGTAAGGCATATGATGATTATGTGAA
 GTTTAGGGATCCAATTACAAAACATGATTCAATTGATGGCTATTTGTTTAATATTGCCAT
 GTTGAAACAGTTGTTTAAGCCTGATTTTCAGCTGCATTGGGCTAAACAGGTTAGTATTTT
 TACCTTCTTGTTAATTGGTTAGTTAACTTCCTTTGTTGGGAGCATGGCGTAATTAGTA
 GTTGCCAACGGGTTCAAGCCATGAAAACCGACTCTTGCATAAATGCGAGGTAGGTCCA
 AATGATCGAATTGTAGTTCTTAACAATAGTAGGTGAGATGATCGAATGGAAATGACAAT
 TTTTACATGGCATATTTTCATACATTCTGGTCTTCAGTCTCTTACATAAATTAAGAACATT
 GAAATAATATGTCAAGGAGTCGTTTGAACATTATTTTTTTTCTTCTATCCAAAAGACGTA
 TTTTCAAGCCATGTTTGGTTATCAAATTGGACTGAACTTTTGTGTTTTTTTTTTTTTGA
 ATGAGTTTTAAGTTTGAAGTAAAAATTCACCTTTCTCTTACAAAATGTAATTTTTTTCAA
 GTGAAATACATGTCCAAACATAACTTTAAATTTCAAATATTATTTTTCAACTTAACTCCGA
 TCCAAACACTACTTTGTTTTCTTTAGAAATCACATTTTTATGTCCTAAGCGTCTACTAAA
 TGTCTCATCCGAAGCTCATGAGACTATCAACAACGTGACCATCACCTTCATCGATATAT
 TTTACAGTGAATTTCTACACTATAAATGAAATTTAACAGGCTATAAGATGTCAAACCTCGT
 TATGTAGATCAATTGTTGTAAGTCAATATAACCTAATTGTTCAACTATCTATACGTTATCA
 GTTTACATAATCTTTACCTGTAGTTTAACTTCTATTAAGTGTATATAGTTGTAA
 CGTATAACTGTCTAACTTCATGTTAATGATAGCATTTACTTGTAGATGATATTATTCATT
 TATCTTGTTCGTTAGTGATTAAATTTTGATCACTCTTCTAAGCTTTCTACCGTTAATGCTG
 GTCGATAAAAAATCCTTCAAAAAGTTCACTAGTACGTTAAGCACTTAAGCCTAGACAGTTG
 ACATTAATAGACTACTAATACTCACATGATAGACAAAATTTTTATAATTTGCTGATTAT
 ATATAATAATCATGTGATGACCAATTATATATGGAATTATCAATACAATCACAATTCACAT
 CATTATGTGAGGCTAATTCCACTACTAGAAAATGTCAAAAAAGCATCCAGAGCATACC
 GACAGAAATTGATTGGAAATAATACCGACTGACATCGGTCAGAAACGAAGTAAATTGGT
 CGAAAATATCAAATAACAAATTGTAAATAAAAATACCGACCGCAATTAGTAGAAAAATGT
 GATCCCACAAAAAAGAAAGTTTACTTTATCTGACACTGTAAATTTTCGACCGATTACG
 GTCGGAATGATCAATTTTTTACCAAAGCCCGATCAGACTTGCATACTATCAAATATAG
 GTTTTAGAAGTTTATTTTTCTTTTATGATTTGTATCGAATTAAGTTAGATTTGTGTGCAATT
 AGATTTGCCCGGAGATAATGAAATTAATAATCAAAAAGAAAGTTTCAGAAATAGAAAAAC
 TTAACCAACCTCGATATTTAAATTATACTTATACTTAAATTAAGAATTTAAAAAGAAAA

FIG. 15

8/31

AAAAACAAATGAGGGACCTGCAATAAATCACAACCAATCCTTTCTATGGTCAATATTTA
TTACACGTAAGCCAAAGACATCCTTTTACAACATGCAAACATAAAACATCACTACTAAAA
ACAAATGGCCACCTCACAACCTTTCCGACCTCATTTTCCGGCCATCACTTCACCGGCGCA
CCACTTTCCGGCAATGCCACCCTACTTCAGTCTTTCTCACTCCTCCAAAAAATATCAAAA
CAAAAACCTCTTAAATATGACAGAAAAATCAAGTGTTTAATTAAGTTTATGTTTATGATAA
ACAGAGCCCCAACCAAAAAACCAACAGTTGATATGAACCAATTAGTGGAATTTTTATATGA
GGATTTGCCTCATCTTTTTGATGATCAAGGTATTGATCGTAAGGCATATGATGATTATGT
GAAGTATAGGGATCCAATTACAAAACATGATTCAATTGATGGCTATTTGTTAATATTGC
AATGTTGAAACAGTTGTTTAGGCCTGATTTTCAGCTGCATTGGGCTAAACAGGTTAGCA
TTTCTGCAGTTTCTTGTAATTGGTTGTGTGGCTATGAGGTGCCTAAAGTAAAAATAGCA
CGGGCTAGCCAATTTTCGGATTGGTAATTGAAAAATAGTCAACGTTTGCAAAGTCATTG
AAAAATAGCCACTATTTTGCTGCAACATGGAAAGTTACAACATAATACTGGAGATTAG
TGCACATGTGTATGAACTTTCAGCATAATACTGGAGATTGGAGGAGCACATGTGTAT
GAACTTTCAGCACCATGATATTATGCTGGAACCTCCAGTATATTATGCTGGAAGTTCATAG
GTAAAAAATTCGAACTCCAGTATATTATGCTGGAATATTTTTCGGATTTTGAATAGGGTTT
TCGTTTCAAGTTTATCTTTACATGAAAAGTGGCTAAGTTTTGATTACTTTTGAAGTGTGGC
TATTTTTTAATTACCACTTGTAATCTGACTATTTTTGAATTTAACCCGTGCCTAAAAGAT
GGAGTAGGATTTAAGTTATATACGCCGTCATGTAAAGGTTATTATTATACTCAGTGGATA
TAACTTGTGGGTTGATTACAATATTTAGCAGATTACTAGTTAATACTTACTAGAGATTAA
CATGTAATTACTAGTTAATAAGTGACCTCATTGTTTAAATATTTTTATACCATTAATGGAT
AGAATTTCAACTCTAAAGGATAACTATACCTAATATAGGTGACTATTAGGGTGCCTAAGA
AGCCACCACACTAATTGTCCTTAGAAAGTTGATTAGTAATCTGGATAACAAGGCAGATT
ACCTATTACGACAAGTTAAATATACCGATAGTGTGTGTATATAAGTTAAATAAATTGGTT
GTTTAAACTCTCTGTTTTTTTATCGCTTTTCGTTCTTCTTATTATGTTTTCGGTTGGTTTAA
GGTGTTCACAAATATTCAAGAACATAGTATCACCCCTTATACATAGTACATTTACTGCTC
TTTGAATATCAGTCCACTGTTTGTTAAATGCAGAGGCACAGATGAACTGATGAACTGCTT
CTGATTGTTTAAAGTGTGTCTATTTTCAATTAAGTGAAGAATTTTACATGTTGAGTATTATGTAG
TTTGGTATTGGGGCAGAAAGTTGCATCTAATCTGTTTGTGTTTGTGCTACTAACTTAATTG
CGTATTAGTGTCTCATGGTTTTGCATATCCATTTAAGGATACCGATACATGTAGTGTACT
ATGTAGGTTCTCTTTTTCAAATTTAGTTTGGAACAGCCTCTCTATCCTCACAAGATAG
GGGTAAGGTCTGCGTATATGCTACCCTCCCAGATCCACGGGTGTGGGATAATACTGG
GTATGTTATTGTTGTTGTTGTAGTTTGGTATTGGAGCAGAAGTTTTATTTAATCATTAAAT
CAGTTTGCTTTGAGCTACTAACTTGTATTTTTAAGGAATGTGATGCATAGATTATGCATT
TGAATCTTTTGAGTGATTGAGTTTGTGTAGACAGGACCCTATGAAATAACTACAAGGTG
GACAATGGTAATGAAGTTCATTCTTCTTCCATGGAAACCTGAATTAGTCTTCACCGGCAC
TTCTGTTATGGGCGTCAATCCTGAAACAAACAAATTTAACAGCCATGTGGTTTGATTCTT
CACTCTTCTCGACGCGAATTCAAATTTTTAGATTGCTTTGCGAATTAGTCACCGACTCT
TTGTTGAAATCTATAGGACTACTGGGATTCAATTAAGAATAACGATTATTTTTCTGTAGAA
GGTCTGCTGGAGGTCATAAAACAGGTGCGCGTTTCAAACTGTTTTAGTTTATGCTTTTAA
TTAAATCTTGGACTGTTGGGATAAACCAGGCCCCAATGATACTCTTAACATGGTGTGATAT
TGTCGCTTGGGGCCAAAGCCCGCACGGTTTTCACTAAAAGGGCTCGTACCATTAAGAG
ATCCTTACACCTTATATATAAGCTCCCAATATTTTCAGCTACCAATGTCGCCTTATTATGA
TTTGAGAATACAAATCTTGGGCAGTTAAGGATTTACAAGACTCCGGATTTGGAAACACC
TAGTTATCAGATATTAAGAAGAACAGCAACTTACGAGGTTTGTAGTATTTATGTTATAAC
ATCGTATTTTCAAGCTAAGAACTTCTATATTAGGCGTACATTTTTATTTGGTTTGCCTAC
TTTTAGCTGAACTGATGCACTAGGACGTTTAGATTATGATGTGGAATGGTCACAACCTT
CATTGGTGCAATACAGGTCAGGAAATATGATCCGTTTATAGTTGTAGAAACAGAAGGTG
ACAACTCGCAGGAAATAGAGGTTTCAATGATGTTGCGGGGTGAGATAAGTTGACTATT
TTTTTTAAAGAAAAAACTTCTGTGATGTTTTCTGATTTAGATTGGTAGACTGTTAACTC
GTCTTAATATAGGTACATATTTGGCAAAAATTCAGCGACAGAGAAGATACCGATGACTA
CTCCTGTCTTCACTCAGGCATTCGATGCTGAAAAATCTAAAGTGTCATCCAGATAGTTC
TTCCATCGGATAAATCTTTGAGCAGGTAACACTTAACACGTTGTATTAATAATATACAAG
AAAACATCTCACACTGAGTGAACACAGTATGGAATCTTTAAATATTCTTAGCACTGATCT
GATTATACTTTTTGACAATATGTTTTACATTTTAACAATTTGTTGAACTTAGTGGCTTT
TCATATATATAAGTCATCATAAGCATCGCCCGCCTATGTGTAAGAGTTGGACTTTAGTCC

FIG. 15 (Continued)

TTTTCTTTGTTCAAATGAGACCTATGGAAATATTGCCTTATTCTTAACAGCTTACCAGCT
CCCAATCAAGAAGGTATAAGCTTAAGGAAAACGGAAGGAGGCATCGCTGCTGCATTAA
AATTTAGTGGA AAAACCAACTGATGATGTTGTTGCGAGAAAAGGAGAAAACAACTCAGATCA
AGTCTTATTAAGGACGGTCTCAAACCTCAATCGGATTGCATGCTTGCTCGCTACAATGA
TCCTGGTCTGAACGTGGAAGTTTATAATGGTATATTTCTATGTA CTGATAATGTTCCTTTC
ATTGCATTCCTGCTTAAGTTCCTCTTTTGCTTTAGACTTTAGAAAAAGGAATCTTCTTTTG
CTCTTTCTATTTCTGTTTTCAATTATCAGTGCCTTCATTACCAAGGGCGGAGCTAGTATA
CTATGTACGGTTTCGGCAGAACCCACTAACTTTAGCTGAAACCATGTATTTTGTTAAAAA
ATAAAACCACTCAATGTGTATAAAATAATTTATTCAGAACCCAGTAAGTCAGCCTTTTCTAT
AATAAGACCCCATAACTTAAATCTTAGCTCGGCCTCTGTTCACTACTGAATATTTTGA
TATTTCAA CTAGTTTGAGATTGAGGCATAGGTGATTGATTGATCAATAAAATTAGTTTGA
CATTATGTTTG GTATATTATCTACTGCTAATATTTTCTCATCTGATGGAAGGTTTTTGTTG
GTGTTTGCATTATATTTTATGATCCTTTGGGTGACAAGAGTGGGTTGCTCTAGTGGTGA
GCACCTCCACTTCCAACCAAGAGGTTGTGAGTTCGAGTCACCCCAAGAACAAGGTGG
GGAGTTCTTGTCGGGAGGGAGCCGAGGGTCTATCGAAAACAGCCTTTCTACCCAGGG
TAGGGGTAAAGTCTGCGTACACACTACCTTCCCATACCCCACTAGTTCCCATACCCC
ACTAGTGGGATTATACTGGGTTGTTGTTGTTGTAGTGTTCTTTATCACTTCTTAAAGTCA
TTGTGTGTCTAATCTATTTGTTTTTCTGTTTGGATTTTTGCAGAGGAATGAAGTTC TTATA
TGGCTTGAGGATTTCAAGCTGGATTAAGCTCCCATCTCAA AATTGCTGAGAGTTTCAT
GAGTGAAAAACAGGAGAGATTCA CAAGCAAAGACAGTTTAGAAAAATGTTCTTATACTG
AACAGATCCAAAAGCTATGAAATTAGGTTTACTGTATTTTGTATCTCAA AATATTTAGAGA
TTTTGACAAATATATGGCTTG TACATAGCACAGAATGCTTGTGCTGAACACTACTAACT
GATTTCTCTTG TGC ACTGCACCAAATATTGTTAGAGCTGTAGTTCTATG

Nitab4.5_0003235g0060.2 coding sequence (SEQ ID No. 12):

FIG. 16

Nitab4.5 0004868g0020.2 protein sequence (SEQ ID No. 13):

FIG. 17

10/31

DERVVVKEEGERKFGVVQFSGTASDKVVKEKVENLRKCLERDGYKIIGDFELARYNPPWTI
PPFKTNEVMIPVE

FIG. 17 (Continued)

Nitab4.5_0004868g0020.2 genomic sequence (SEQ ID No. 14):

CTCCACCAGCTTATGGTGCAAACGTGGCAAACAAACTCCACAGCAGTTCCCTCATCTA
CGGCCATTTCATTTTGCAGCAATCAAAAACCCCTTTCCCAAGTTCTCTCTTCTCTTTT
CGATTTCCACAAAGAACTTCTCTTCTCTCCCTTATATGTAACAAATTCACAGCCTCC
ATTTACTATCAAAACATTCTCAAATTCATTCCCAATAACTAATTTTTCGTTCCATCAAAAT
GGGCATGATTTTGGGGAAAATCACTGTGGAAACACCAAAATACGAATTGATTCAATCTA
CAGCTGACTACGAAATCCGCAAATACCCAGCATCTGTTATAGCACAAGTCACATATGAT
CCAACCCAGTTCAAAGGAAACAAAGACGGTGGGTTTATGCTATTAGCCAATTACATCGG
CGCACTTGGCAATCCTCAAAACGCTAAGCCTGAAAAAATCGCCATGACAGCTCCTGTAA
TCACCAAATCGTCCGAAAAAATCGCGATGACTGCACCAGTGGTGACTAAGAGTGGTGA
TGGAGAGAACAATATGGTGACGATGCAGTTTATTTTACCAGCAAAGTATACAAAAGCTG
AAGAGGCACCTAAGCCGTTGGATGAAAGGGTGGTGGTTAAAGAAGAAGGGGAAAGAAA
GTTTGGTGTGGTGCAGTTTAGTGGGACTGCAAGTGATAAAGTGGTTAAAGAGAAAGTG
GAGAATTTAAGGAAATGTTTGGAGAGAGATGGGTATAAGATAATTGGTGATTTTGAGTT
GGCTCGGTATAATCCTCCTTGGACAATTCCTCCATTTAAGACCAATGAAGTTATGATTCC
AGTTGAGTGAGCTGAAAATGGTGTATTAAAGAAAAATATGAGACGGCTGTGAAGCTAGC
TTGTGAACCTACTTTATCTGGTGTATTATTAGTATTGAATTGTGATTTCTTGTTTATATGTT
TACGTTGTTGACAATGTCTTTGCATACCCTAGCTAGTGTCAATTATTCATAGTTAATTAG
CACATATTTTAGCTCAATTTTATACTTTGCCATAGTAAAGTAGTAAACCAATGTGGTTTT
GCAAGCACCTAATGTGGTAAGTCCTTACTCTCTGTTTTTCATTTACACTCTTCTCTCTG
TCCATTTCTGCCAGTATAAGTGCTTCTTCTGTCAATACCCGGTGAATCATGTAAGGGCC
CTGTTAGTTGGGTAAAAATTTTCTTTTGCTTCATCTTTATGTGATAAGATTCGCTTTAGC
ACCAATTGCCCCAGCATGAATTTCTTGATCTGACCTTTTTGTTGAAAAATCTTGCCATT
CTGTTCTAGTAGAGTTGACCGTGACATACTACATTCATTCTCTTTTCATCAATGAGAGCT
AGTTGTTCTACTGGCTTCGTACCAATTCCTCGTCGCTGAGCTCGGTTTTCTTGATGATT
CTTAAAGAAGGAATTTCTACTTCGGCGGAAATAACGACTTCAGTACCATAAACTAGTAG
GTAGGGAGTTGCCCCATTTGATGTACGAAGTGTGGTACGATACCCCGAGTAAGAAAAT
GGTAGATTATGATGCCATTATTTGTAATTGTTACCATTTCCCTCAATATCTTTTTGATGT
TCTTGTTGGCGGCTTCACGGCTTCGTTCAATTTGCGACCTGTACGCTATAGAGTTCCTA
TGCTTGATCTTGAATGTTTCACACATGGCTTCATCAAATTGCTATTGAGA

FIG. 18

Nitab4.5_0004868g0020.2 coding sequence (SEQ ID No. 15):

ATGGGCATGATTTTGGGGAAAATCACTGTGGAAACACCAAAATACGAATTGATTCAATC
TACAGCTGACTACGAAATCCGCAAATACCCAGCATCTGTTATAGCACAAGTCACATATG
ATCCAACCCAGTTCAAAGGAAACAAAGACGGTGGGTTTATGCTATTAGCCAATTACATC
GGCGCACTTGGCAATCCTCAAAACGCTAAGCCTGAAAAAATCGCCATGACAGCTCCTG
TAATACCAAATCGTCCGAAAAAATCGCGATGACTGCACCAGTGGTGACTAAGAGTGGT
GATGGAGAGAACAATATGGTGACGATGCAGTTTATTTTACCAGCAAAGTATACAAAAGC
TGAAGAGGCACCTAAGCCGTTGGATGAAAGGGTGGTGGTTAAAGAAGAAGGGGAAAGA
AAGTTTGGTGTGGTGCAGTTTAGTGGGACTGCAAGTGATAAAGTGGTTAAAGAGAAAGT
GGAGAATTTAAGGAAATGTTTGGAGAGAGATGGGTATAAGATAATTGGTGATTTTGAGT
TGGCTCGGTATAATCCTCCTTGGACAATTCCTCCATTTAAGACCAATGAAGTTATGATTC
CAGTTGAGTGA

FIG. 19

11/31

Nitab4.5_0006614g0010.2 protein sequence (SEQ ID No. 16):

MQAVMNFFMRLAFLILVSKLSNGYSTNHLQPQKLGFPPTCNRIECPNYDLIQSGKDYEIR
LYNSSMWMSTAPIDDINLYSATRTGFLRLFDYIQGKNSYQEKIEMTAPVITQVKPSDGPFC
SSFVVSFYVPKKNQPNPPAKGLHVQIWSNTYVAVRQFGGFVADVDVAKEAAALSASIADT
KWAAAVEKSHAADNTTMYTVAGYNPFEFKDRVNEIWFTFDLDKASAI

FIG. 20

Nitab4.5_0006614g0010.2 genomic sequence (SEQ ID No. 17):

AAAACCTCTGGACACCCCTCATTGATTTTCTTTTTCTTTTATGCATGAGGAAATTAGTTC
ATTCGGGTCACTTCATTCTCTCTCCGTATAATTTCTAGGAACTAAAACACTAAAAAGAAA
CAAAAGATGCAGGCTGTTATGAATTTCTTCATGAGGCTAGCTTTCTTTCTGATCCTCGTT
TCTAAATTGAGCAATGGGTATTCGACAAACCACTTGCAGCCACAGAAGTTGGGTTTCTA
TCCTCCAACCTGTAACCGAATTGAGTGCCCAAATTATGACTTAATTCAATCTGGAAAAGA
CTATGAAATTCGTCTTTATAATTCATCCATGTGGATGTCTACTGCACCCATTGATGATATT
AACCTTTATTCCGCCACCAGAACTGGTTTCCTCAGGTTAATATTATATTCCCTCTTTCTC
CCACTAGTTCCCTTTTACATCAGTTATCTGTATCTATTGTTTTTTTTTTTTCAATTCCTTTC
TATTTGTTGCATATTTGATAAGTTAGCTACACTGTGCAACACCAACTAAAGCATGTGTGT
TTGCTTTCCAAGATTCTCAAAAAAAAAAAAAAAAAAACAAGGGACTTAATTAAGGTTT
ACTTACATCTCAATCAATTGAAAATAATCGACTATTAAGGAAAAAAAAAATTATGTCTT
GCCAGGTTACATTGAAGCAGGTTGCTGCCAGTTACCAATTTACAATTATATTGCTAGTTA
TCTGTAAGTGAATTTGATTTATATGTACTGAAAATGCATAGTATTTAATCTAGATTGTTAAA
GCAAATACCTGCCTTATTTTACAAGTTACTAATCTCCCTTTTTACGGTTAGTTATCGTGTT
ATTATTGAGTTGACTTGATAGTGTAAGAGTGATCTTTTCCACTGTCAGTATTTAGAAGTTA
TTTTATGTGAACCTAATCTTTATACTTAAGATGTTAGCCTATCCATTTGCAAAACATAAAA
AACAGTTGTTGTAACTTTGCCTTGGCAGGCTATTTCGATTACATTCAAGGGAAGAACAG
TTACCAGGAGAAAATAGAGATGACAGCTCCAGTTATCACTCAAGTAAACCAAGTGATG
GACCATTTTGTGCATCTTCATTTGTTGTGAGCTTCTACGTACCAAAGAAGAACCAACCAA
ATCCTCCTCCAGCTAAAGGCCTTCACGTCCAAATATGGAGCAATACTTATGTGCGCCGTC
AGGCAATTCGGCGGATTTGTAGCTGATGTTGATGTTGCAAAAGAAGCTGCTGCTTTGAG
TGCTAGTATTGCTGACACTAAATGGGCAGCAGCCGTTGAAAAAGCCATGCTGCAGATA
ACACTACGATGTATACAGTGGCGGGATACAACCTCTCCTTTTTGAGTTCAAGGACAGAGTT
AATGAGATTTGGTTTACATTTGATTTGGACAAAGCATCTGCCATTTGATTCTGGACTCTT
TAACATTTCAAGGAGACATGAACTATCTTATTTTCTTGTTTTGTATTACTATTAACAACA
GTTTTGTATGACCACAATGCTACATTATATAGATAGAAAATAAGATAGTTGTACAATAC
AATTTGATTCTGAGGATTTAGTGGTTAAGATACCTCTTCAAGACCTTTCAAATAATCCA
ATATATATGGAAGGGCTTAGACCATATTGAGTTTATTGTATGCAGTT

FIG. 21

Nitab4.5_0006614g0010.2 coding sequence (SEQ ID No. 18):

ATGCAGGCTGTTATGAATTTCTTCATGAGGCTAGCTTTCTTTCTGATCCTCGTTTCTAAA
TTGAGCAATGGGTATTCGACAAACCACTTGCAGCCACAGAAGTTGGGTTTCTATCCTCC
AACCTGTAACCGAATTGAGTGCCCAAATTATGACTTAATTCAATCTGGAAAAGACTATGA
AATTCGTCTTTATAATTCATCCATGTGGATGTCTACTGCACCCATTGATGATATTAACCTT
TATTCGCCACCAGAACTGGTTTCCTCAGGCTATTTCGATTACATTCAAGGGAAGAACAG
TTACCAGGAGAAAATAGAGATGACAGCTCCAGTTATCACTCAAGTAAACCAAGTGATG
GACCATTTTGTGCATCTTCATTTGTTGTGAGCTTCTACGTACCAAAGAAGAACCAACCAA
ATCCTCCTCCAGCTAAAGGCCTTCACGTCCAAATATGGAGCAATACTTATGTGCGCCGTC
AGGCAATTCGGCGGATTTGTAGCTGATGTTGATGTTGCAAAAGAAGCTGCTGCTTTGAG
TGCTAGTATTGCTGACACTAAATGGGCAGCAGCCGTTGAAAAAGCCATGCTGCAGATA

FIG. 22

12/31

ACACTACGATGTATACAGTGGCGGGATACAACCTCTCCTTTTGAGTTCAAGGACAGAGTT
AATGAGATTTGGTTTACATTTGATTTGGACAAAGCATCTGCCATTTGA

FIG. 22 (Continued)

Nitab4.5_0006991g0050.2 protein sequence (SEQ ID No. 19):

MATSQLSGHIFRSSLHRRITIFRQCHPTSVFLTPPKNIKTPLKYDRKFKWLIKFSLVDKQTPT
KKPTVDMKQLVEFLYEDLPHLFDDQGIDRKAYDDYVKFRDPITKHDSIDGYLFNIAMLKQLF
RPDFQLHWAKQTGPYEITTRWTMVMKFILLPWKPELVFTGTSVMGVNPETNKFNSHVDYW
DSIKNNEYFSLEGLLEVIKQLRIYKTPDLETPSYQILRRTATYEVRYKDPFIVVETEGDKLAGN
RGFNDVAGYIFGKNAATEKIPMTTPVFTQAFDAEKSQVSIQIVLPDLSLNSLPAPNQEGISIR
KTEGGIAAALKFSGKPTDDIVREKEKQLRSSLIKDGLKPQSGCMLARYNDPGRTWKFIMRN
EVLIWLEDFKLD

FIG. 23

Nitab4.5_0006991g0050.2 genomic sequence (SEQ ID No. 20):

ACTAAAACCAAATGGCCACCTCACAACCTTTCCGGCCACATTTTCCGGTCATCACTTCAC
CGGCGCACCATTTTCCGGCAATGCCACCCAACTTCAGTCTTCCTTACTCCCCCAAAAAA
TATCAAAACAAAACCTCTTAAATATGACAGAAAATTCAAGTGGTTAATAAAGTTTAGTTTA
GTTGATAAACAGACCCCAACAAAAAACCAACAGTTGATATGAAACAATTAGTTGAATTT
TTATATGAGGATTTACCTCATTTATTTGATGATCAAGGTATTGATCGAAAGGCATATGAT
GATTATGTGAAGTTTAGAGATCCAATTACAAAACATGATTCAATTGATGGCTATTTGTTTA
ATATTGCCATGTTGAAACAGTTGTTTAGGCCTGATTTTCAGCTGCATTGGGCTAAACAG
GTTAGCATTTCTACAGTTTCTTGTTAATTGGTTAGTTAAACTGCTGTTTGTTAACTTTCT
TACTTGGATTTAAGTTTACGGGAGCCTTGCGCTAACTGGTAAAGTTGATGTCATGTCAC
CAAAAGGTCATGAGTTCGAGCAGTGGAAACAACCTCTTGCAGAAATGCAGTATAAAGCT
GCATACAATAGACCTTTGTGGTCCAGCCCTTTTCCGGACCCCGCGCATAGCGGGAGCT
TATGCACTGGGCTGTCTTTTTTCTCGGATTTTAAGTTATGTTCAATGATGGTGTGAAA
TTATTTTACTCTAATTAATGTGTGTCGTAGAGATAATGTAATCACCTCATAATTGACCTGA
CTGTGTAAATAATTTTATATCGTCCGTGCATAAAACGAAAGGTCCACAATCACTTAGAGT
ACCATGTTACAAATAAATGTTTATCATCAAGAATCACTTAGAATTACCTTGCGATATGATT
GTGATTTGCACTGTCAGTGCCTAAACTTAAGCGTCACGCCTATAATACAATTATTAAT
GCTTATTATAGGGAATCACTTGGAATTGCCTTACGATATGATTGTGATTTGCATTGTTA
GTGCCTGAACTTAAAGGTCTGCGAACATATTACCTTTAGCCGAGGGTCTATCGAAATA
ACTCTCTACTTGCACAAGGTAGGGGTAAGGTCTGCGTATATACTAATCTCCCTAGACCC
CGCTTGTGGGATTACACTGGGTTTATTGTTGTTGTTGTGCCTAAAGCTTAAGCACCATAT
CTATTATACAGTTACAATTGAATCACTTGGACTACCTTAATTGTGTATTAGTGTCTCATGG
TTTTGTAAATCTATTTATGGATACCGATACATGTAGTGTACTATGTAGGTTCTCTCTTTCA
AAATTTAGTTTGGTGTGAGAAGTCTTATTCAATCATTTAATCTATTTGCTTTGGCCGAGG
GTCCATCGGAAACGGTCTCTCTACCTTCTCAAGGTAAAGGTAAAGGTCTGCGTACACACT
ACCCTTCCGAGACCCCACTCGTGGTATTACACTGTGGTATTACACTGAATTTATTATTGT
TGTTGTAGTCAGTTTGCTTTGTGCTATTAACTTGTATTTCTAAGGAATGTGATACATAAA
TTATGCCTTGAAATTTTGTAGTGATTGAGTTTGTGTAGACAGGGCCCTATGAAATAACTA
CAAGATGGACTATGGTGATGAAGTTCATTCTTCTTCCATGGAAACCTGAATTAGTCTTCA
CTGGCACTTCTGTTATGGGCGTCAATCCTGAAACAAACAAGTTTAAACAGCCATGTGGTT
TGATTCTTCACTCTTCCCTCGACGCAAATTCAAATTTATTAGATTGCTTTGCAGATTAGTCA
CCGACTCTTGTTGAAATCTATAGGACTACTGGGATTCAATTAAGAATAACGAATATTTTT
CCCTGGAAGGTCTGCTCGAGGTCATAAAACAGGTGCGCGCTCGAACTGTTTTAGTTTAT
GTTTCTTGAAGTGTGGACTAAACCAGCGCAAAGATACTCTTATCATGGTGTGATATTGT
CTGCTTTGGGCCAAATTCGCGCGGTTTTCCCTATGAGGCCTCACACCATTTGGGAGATCT
CTACACCTTATATATAGGCTCTCAATCTTATCAACTACCAATGTGGCTTTATTATAATTTG
AGAACACAAATTTTGGGCAGTTAAGGATTTACAAGACTCCGGACCTGGAAACGCCTAGT
TATCAGATACTAAGAAGAACAGCAACTTACGAGGTTTGTCAATTTATGTTTACAGCGTT

FIG. 24

13/31

GTACTTCAAGCTAAGAACTTCTATATTAGGCGTACATTTTTATTGGGTTTGCGCTACTTTT
AGCTGAACTGATGCACTAGGACGTTTAGATTTATGATGTGGAATGGTCACAATTATCATT
TATGTAATACAGGTCAGGAAATATGATCCGTTTATAGTTGTTGAAACAGAAGGTGACAAA
CTCGCAGGAAATAGAGGTTTCAATGATGTTGCAGGGTGAGATAAGTTGACCCTTTTTCT
TTAAAATCTTCTGTGGTGTCTTCTGATTTAGATTGGCGGACCATTAACTCGTCTTAATATA
GGTACATATTTGGGAAAAATGCTGCAACAGAGAAGATACCGATGACTACTCCTGTCTTC
ACTCAGGCATTTGATGCTGAAAAATCTAAAGTGTCAATCCAGATAGTTCTTCCATCGGAT
AAATCTTTGAACAGGTAACACGTTTTATTAATTAATAGACACGAAAACATCTCACACCGA
GCGATCATATTGACTTATTACAATATGAGTTCAAATTTTTACATTGCTCGAAACTTAGTA
ATTCATATATATAATTCATAAGCATCGTTCGCCTCTGTGTAAGAGTTGGATTTTAGTTATT
TTCTTTGATCAAATGAGAATTATGGAAATTAAGCTTTATTCTTAACAGCTTACCAGCTCC
CAATCAAGAAGGTATTAGCATAAGGAAGACAGAAGGAGGAATCGCGGCTGCACTAAAA
TTTAGTGGAAGCCAACAGATGATATTGTTGCGGAAAAGGAGAAACAACCTTAGATCCAG
TCTTATTAAGGACGGTCTCAAACCTCAATCGGGTTGTATGCTTGCTCGCTACAATGATC
CTGGTCTGAACGTGGAAGTTTATAATGGTATACTTCTATGTACTGAATATCGTTCTGTTTCA
TTGCATTCTTGTTTAAGTTCTTATTTTGTTTTAGTATGAAGTGTCTCTTTAGAATGAGAAC
CTTGGCGTAACTGGTAAAGTTGCTGCTATGTGACCAGGAGGTCACGGGTTTCGAGCCGT
GGAAACACCCTCTTGCAGAAATGTTGGGTAAAGGCTGCGTACAATTGACCCTTGTGGTC
CGGCCCTTCCGCGGACCCCTCGTATAGCAGGAACCTTAGTGACCGGGCTTCCCTTTTA
GTCTGAAGTGTCTCTTTATCACTTTGTCTATCTGGTGTCTAATCTATTTGTTTTCTGTTTG
GATTTTTGCAGAGGAATGAAGTTCTTATATGGCTTGAGGATTTCAAGCTGGATTAACCTCC
CATTCCAAAAATTGCTGAGATTTTCATGTATGTAAAAACAGGAGAGATTCACAAGCAAAGA
CAGTTTAGAAAAATTCTTATACTGAAATGAAATTAGTTCTCAAATGTTTAGAGATTTTGA
CAAATATATGACTTGTGCATAGCCACAAATGCTTGTGCTGAACTCACTTTACACTAAACA
AACTTCTATTGTGCATTACAAAGTTTCGATGTGTTGATGGTGTAATAATATTTGTCATTG
TAAGGTAATTAATGGTAATTAAC

FIG. 24 (Continued)

Nitab4.5_0006991g0050.2 coding sequence (SEQ ID No. 21):

ATGGCCACCTCACAACTTTCCGGCCACATTTTCCGGTCATCACTTCACCGGCGCACCAT
TTTCCGGCAATGCCACCCAACCTTCAGTCTTCTTACTCCCCCAAAAAATATCAAAACAAA
ACCTCTTAAATATGACAGAAAATTCAAGTGGTTAATAAAGTTTAGTTTAGTTGATAAACAG
ACCCCAACAAAAAACCAACAGTTGATATGAAACAATTAGTTGAATTTTTATATGAGGAT
TTACCTCATTTATTTGATGATCAAGGTATTGATCGAAAGGCATATGATGATTATGTGAAG
TTTAGAGATCCAATTACAAAACATGATTCAATTGATGGCTATTTGTTTAATATTGCCATGT
TGAAACAGTTGTTTAGGCCTGATTTTCAGCTGCATTGGGCTAAACAGACAGGGCCCTAT
GAAATAACTACAAGATGGACTATGGTGATGAAGTTCATTCTTCTTCCATGGAAACCTGAA
TTAGTCTTCACTGGCACTTCTGTTATGGGCGTCAATCCTGAAACAAACAAGTTTAACAGC
CATGTGGACTACTGGGATTCAATTAAGAATAACGAATATTTTCCCTGGAAGGTCTGCTC
GAGGTCATAAAACAGTTAAGGATTTACAAGACTCCGGACCTGGAACGCCTAGTTATCA
GATACTAAGAAGAACAGCAACTTACGAGGTCAGGAAATATGATCCGTTTATAGTTGTTG
AAACAGAAGGTGACAACTCGCAGGAAATAGAGGTTTCAATGATGTTGCAGGGTACATA
TTTGGGAAAAATGCTGCAACAGAGAAGATACCGATGACTACTCCTGTCTTCACTCAGGC
ATTTGATGCTGAAAAATCTAAAGTGTCAATCCAGATAGTTCTTCCATCGGATAAATCTTT
GAACAGCTTACCAGCTCCCAATCAAGAAGGTATTAGCATAAGGAAGACAGAAGGAGGA
ATCGCGGCTGCACTAAAATTTAGTGGAAGCCAACAGATGATATTGTTGCGGAAAAGGA
GAAACAACTTAGATCCAGTCTTATTAAGGACGGTCTCAAACCTCAATCGGGTTGTATGC
TTGCTCGCTACAATGATCCTGGTCTGAACGTGGAAGTTTATAATGAGGAATGAAGTTCTT
ATATGGCTTGAGGATTTCAAGCTGGATTAA

FIG. 25

14/31

Nitab4.5_0009023g0010.2 protein sequence (SEQ ID No. 22):

MLLCSPSSISAHNLSRNRTRPSPINSMAVDRSSSRVATTASQRRNGTSALEARISLVIALAS
 QTSSLSQKLLTELGETAKYVLPKRIFESRTLEEALMSVPDLETVKFNVLKRSQDQYEIREVEP
 YFVAEATMPGKYGFDFNGASQSFNTLAEYLF GKNTKKESMAMTTPVITRRTQSDGERMEM
 TTPVITKRVEDQGKWRMSFVMP SKYGS DLPK DSSVTI KEVPRKTVAVVAFSGFVTDEEV
 KARESRLRAALKGDAEFRVKDGASIEVAQYNPPFTLPFTRRNEISLEVEREQE

FIG. 26

Nitab4.5_0009023g0010.2 genomic sequence (SEQ ID No. 23):

ATGTGACATACACAGTATGTGATAAGGAATTAACCAAGTAATTCGGGGGCGACAAATAC
 AAATATAGATGAACCACAATCCCCCTTCTTTGAAGCACAGAATTTCAAAAGGAGTGAAACA
 AACACTGTAACGTAGTTAAAGATGCTTCTGTGCTCACCTTCTTCCATTTCTGCTCACAAT
 CTTAGCAGAAACAGAACCCAGACCTTCGCCTATCAATTCTATGGCAGTTGACAGAAAGCAG
 CTCCCGCGTTGCCACAACCTGCTTCGCAGCGGAGAAACGGCACGTCAGCTCTCGAAGCT
 CGCATCTCCCTCGTCATCGCTCTCGCCTCCCAAACCTCTTCTCTCTCTCAGAAACGTAA
 TTTCTTCACTCCATTTCTTAATTCCTCATTTTAATCGGATTTTAAGCATTTCTCTCCCAATT
 CTTTTTTTGATTTGTTTAAAACATTTTCGTCCAGTGCTTCCAGGACTCTTTTAAGTCAGATT
 AGAGAACTTCTTAGAAAAGAATTAATTTAGTTTAAGAGACTAATTATTTACCAGTGTTGGA
 ATGAACTTGTGCTTGACGGCTAAAGCGGAACGGATATAAATGATTCTATGTAGTCCTCT
 GGCTAGTTTAGGAATGAGAACCTAGTCAATTGACTAATTGAGTAATTGCTACTTATATTG
 TTAAATTTTTTATTGTTTCGCAGTTCTGACGGAATTGGCTGGTGAAACCGCGAAATACGTG
 TTACCGAAGAGGATATTCGAAAGTCGGACTTTAGAGGAAGCTTTGATGTCTGGTATTCT
 TTCTCAATCTTTTCTATGATTGTTATATCCTCTATTTTAATAATAATGAGAATTTAGATGGT
 TAAAAATAATGACTAGCTCCCATCGCTGTTATCTTCTTGAAGTGCCGGATCTGGAGACA
 GTGAAATTCAATGTTTTGAAACGCAGTGATCAGTATGAAATAAGGGAAGTTGAGGTATTT
 TCTCTATGATTCTATTTATAAATTACTAATGTTGTCATTCTTTTTTTTTTTTTTCCCAGTAGA
 GTCTAGATTTGTATGCATCCATTCTGCAACTTTATCGGGGGAGTGTTCCAGACGAGAG
 AGGAGCAATATTTGACGGGAAAAAAAAGGTTTAAACCTTCATTAAAAAATGTTGAAGAT
 AGAGGGCAAACCTGTTTTGAGAACACCTTTAGAATACTACTTACATAGATACAAATATTA
 GGCAAAGATACGTTTATTATATTTGTTACACCAGCTAAATATTCTCAAACAAAGATTAAT
 CCTTTAGTTAGAATTTCTTGATATATGGTAGAATGCTCCGTCTCAGAAAGTAATCAAGGG
 TACTAATACAAGAAAGCATTAGTTGGTAAAAGTTACTAGGCCATGTAATTAGATACAAGA
 GACATTTGTGAAGATTTTGTGTAACCTGATTCTACATAATGGGCCTTAGTTCATATTTGT
 GTTTTCCCTTATTATGGTCATTAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA
 CAATTCTGAAAGGTACAATACATATGGAAGTGGAAGTGGCACAACCTGGAGACTGATAGT
 TATAGTGGGGAGTCAGCATGTCTATCAATAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT
 AGAGATTTGGTATTCAAAAGTATTTTGGAAAAGAATGAAATTGACAGAAATAGACTAGTG
 ATCACAATCAGGAGTTTCTCAATTTTGGAGAACATGTTATTGACTCGGTGTTAAACAA
 TGATGAGAAAGCAGAGGTGCCAGTTTTTTGAAATAAACATATGATGTTCTAACTTTTGGT
 TTGATAACCTAGAATTAAGCATATCATAAGCTCTTCCATGAACATTAAAGCCCTGGATC
 CTGATACTTGAAAGTCATTCATTCACAAATTGTTTTTTGGCAGATCTACCTTTGTATTAAA
 ATTGAATTCTGGACCATCAGTCACCCTGTTATTTATGGTGCACATATTTCTTCTTATGTGT
 ATATGCTCTTATTCTTTGCTTTTTTCGGAACAGCCTCTACCCCCGGGTAGGGGTAAGGT
 CTGCATACACACTACCCTCCCCAGACCCCATACACTGGTTGTTGTTGTTGTTGTTGTTGTTGTT
 CTCTTAAATTCTTGTCTCAGTTATTCCTAACCTCTCTTAATAGCATTTCTCTTGATTATA
 GATATGCTGTATTATTAGAACCTTAATCACCGATATCTCTTTATGTTGTCATGTTGATAGT
 TATTCTTTCCCTTCTTGTCTCTTAACAATAACTCTGAGTTTTTTATTTTTTATTTTGATAA
 AGTTCATTATAAACTCTGAGTTAATTTATCTAAAGTAATATTGCGACTTTGAACAGCCTTA
 CTTTGTTGCTGAGGCTACAATGCCTGGGAAGTATGGGTTTGAAGTTAATGGTGCATCTC
 AATCCTTCAACACATTGGCTGAGTACTTGTGTTGGTAAGGTAGTTAAAGGTACTTATAGAG
 AGTCAAACCTTAGATCGCCTCCAACCACATTGTAATGGTGGTCTTCTTTGGCTCTGTTCTA

FIG. 27

SUBSTITUTE SHEET (RULE 26)

16/31

ATTTCCATTTCAGAGTTCAAAAGCAAAAACATTTAGGAATTCTATTTTTGAAGTTTTACATC
CTAACTTAGCTAGTTTTCTCCCATTTGATTTCAACAAAGGTTTTGTGACTGATGAAGAAG
TTAAAGCCCCGAGAATCAAGACTCCGTGCTGCGCTAAAGGGAGATGCAGAGTTTCGGGT
AAAAGATGGTGCCTCAATAGAAGTTGCACAGGTATATTTCGTAGAACCATTTCAATAGTAC
TTGTAATTCTCCTGGGGAACATATTTTGGCCTAGTGTAAACCCCAAATGACATACTAGTT
GTGTGTTTTTTATACTTCTGCTTGCAGTATAATCCACCATTTACTCTTCCGTTTCACACGTC
GGAATGAGATTAGCCTGGAAGTTGAAAGGGAACAGGAATAGCTAGTCAGTGTGCAGTA
TATTGCAGATAAATCCGTGCACTGATGAAACCAGAGAAACAAAACATATATGGTGTAAAT
AAATTATATATATATATGTATATATATTTGGTGTATTGCTTGACATTTTGGTAACAAGGTTA
TGACATGCACAGAAATGTAAGGTCATTTATGCTTCAACTCTATAATAACACGTTGTCTT
TAATACTC

FIG. 27 (Continued)

Nitab4.5_0009023g0010.2 coding sequence (SEQ ID No. 24):

ATGCTTCTGTGCTCACCTTCTTCCATTTCTGCTCACAATCTTAGCAGAAACAGAACCAGA
CCTTCGCCTATCAATTCTATGGCAGTTGACAGAAGCAGCTCCCGCGTTGCCACAACCTGC
TTCGCAGCGGAGAAACGGCACGTCAGCTCTCGAAGCTCGCATCTCCCTCGTCATCGCT
CTCGCCTCCCAAACCTCTTCTCTCTCAGAACTTCTGACGGAATTGGCTGGTGAAAC
CGCGAAATACGTGTTACCGAAGAGGATATTCGAAAGTCGGACTTTAGAGGAAGCTTTGA
TGTCTGTGCCGGATCTGGAGACAGTGAAATTCAATGTTTTGAAACGCAGTGATCAGTAT
GAAATAAGGGAAGTTGAGCCTTACTTTGTTGCTGAGGCTACAATGCCTGGGAAGTATGG
GTTTGACTTTAATGGTGCATCTCAATCCTTCAACACATTGGCTGAGTACTTGTGTTGTA
GAACACGAAGAAGGAGAGCATGGCGATGACAACACCAGTTATCACTCGTAGAACTCAA
TCTGATGGGGAGAGGATGGAAATGACTACTCCAGTGATAACTAAACGGGTGGAAGATC
AAGGAAAGTGAGGATGTCTTTGTCTATGCCCTCAAAGTATGGTTCAGACTTGCCACTA
CCAAAGGATTCTCCGTAACCTATCAAGGAGGTGCCTAGGAAAACCTGTCCCGTTGTTG
CTTTTTCAGGTTTTGTGACTGATGAAGAAGTTAAAGCCCAGAAATCAAGACTCCGTGCT
GCGCTAAAGGGAGATGCAGAGTTTCGGGTAAAGATGGTGCCTCAATAGAAGTTGCAC
AGTATAATCCACCATTTACTCTTCCGTTTCACACGTCGGAATGAGATTAGCCTGGAAGTT
GAAAGGGAACAGGAATAG

FIG. 28

300-nucleotide cDNA fragment (SEQ ID No. 25):

GTAAGAACACAAAGAAGGAAAGTATGGCGATGACAACACCCGTAATCACTCGTAGAACT
CAATCTGATGGGGAGAAAGATGGAAATGACTACTTCAGTGATAACTAAAAGGGTGGAAGA
TCAAGGAAAGTGAGGATGTCCTTTGTCATGCCCTCAAAGTATGGTTCGGACTTGCCAC
TACCAAAGGATTCTCCGTAACCTATCAAGGAGGTGCCTAGGAAAACCTGTCCCGTTGTT
GCTTTTTTCAGGTTTTGTGACTGATGAAGAAGTTAAAGCCCAGAAATCAAGACTATGTGC
TGCGT

FIG. 29

TRV1 (SEQ ID No. 26):

GTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATTTTCGGGAAACCT
CCTCGGATTCCATTGCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAG
GTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCATTCAAGATGCCTCT
GCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGA AAAAGAA
GACGTCCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGACATCTCCACTGACGTAAG
GGATGACGCACAATCCCACTATCCTTCGCAAGACCCCTTCTTCTATATAAGGAAGTTCATT
TCATTTGGAGAGGACAGCCCAAGCTTTCTAGAGGATCCATAAAACATTTCAATCCTTTGA
ACGCGGTAGAACGTGCTAATTGGATTTTGGTGAGAACGCGGTAGAACGTACTTATCACC
TACAGTTTTATTTTGTCTTTTCTTTTGGTTTAATCTATCCAGCTTAGTACCGAGTGGGGGA
AAGTGACTGGTGTGCCTAAACCTTTTCTTTGATACTTTGTAAAAATACATACAGATACA

FIG. 30

17/31

ATGGCGAACGGTAACTTCAAGTTGTCTCAATTGCTCAATGTGGACGAGATGTCTGCTGA
GCAGAGGAGTCATTTCTTTGACTTGATGCTGACTAAACCTGATTGTGAGATCGGGCAAA
TGATGCAAAGAGTTGTTGTTGATAAAGTCGATGACATGATTAGAGAAAGAAAGACTAAA
GATCCAGTGATTGTTTCATGAAGTTCTTTCTCAGAAGGAACAGAACAAAGTTGATGGAAATT
TATCCTGAATTCAATATCGTGTTTAAAGACGACAAAAACATGGTTCATGGGTTTGCGGCT
GCTGAGCGAAAACTACAAGCTTTATTGCTTTTAGATAGAGTTCCTGCTCTGCAAGAGGT
GGATGACATCGGTGGTCAATGGTCGTTTTGGGTAAGTACTAGAGGTGAGAAAAGGATTCAAT
CCTGTTGTCCAAATCTAGATATTCGGGATGATCAGAGAGAAATTTCTCGACAGATATTTCT
TACTGCTATTGGTGATCAAGCTAGAAGTGGTAAGAGACAGATGTCCGAGAATGAGCTG
TGGATGTATGACCAATTTTCGTGAAAATATTGCTGCGCCTAACGCGGTTAGGTGCAATAA
TACATATCAGGGTTGTACATGTAGGGGTTTTCTGATGGTAAGAAGAAAGGCGCGCAGT
ATGCGATAGCTCTTCACAGCCTGTATGACTTCAAGTTGAAAGACTTGATGGCTACTATG
GTTGAGAAGAAAACATAAGTGGTTCATGCTGCTATGCTTTTTGCTCCTGAAAGTATGTTA
GTGGACGAAGGTCCATTACCTTCTGTTGACGGTTACTACATGAAGAAGAACGGGAAGAT
CTATTTGCGTTTTTGAGAAAGATCCTTCTTTTCTTACATTCATGACTGGGAAGAGTACAA
GAAGTATCTACTGGGGAAGCCAGTGAGTTACCAAGGGAATGTGTTCTACTTCGAACCGT
GGCAGGTGAGAGGAGACACAATGCTTTTTTCGATCTACAGGATAGCTGGAGTTCCGAG
GAGGTCTCTATCATCGCAAGAGTACTACCGAAGAATATATATCAGTAGATGGGAAAGCA
TGGTTGTTGTCCCAATTTTCGATCTGGTCAATCAACGCGAGAGTTGGTCAAGAAAGAC
CTGTTTGTAGAGAAACAATTCATGGACAAGTGGTTGGATTACATAGCTAGGTTATCTGAC
CAGCAGCTGACCATAAGCAATGTTAAATCATACTTGAGTTCAAATAATTGGGTCTTATTC
ATAAACGGGGCGGCGCGTGAAGAACAAGCAAAGTGATAGATTCTCGAGATTTACAGTTGTT
GGCTCAAACCTTTGCTAGTGAAGGAACAAGTGGCGAGACCTGTCATGAGGGAGTTGCGT
GAAGCAATTCTGACTGAGACGAAACCTATCACGTCATTGACTGATGTGCTGGGTTTAAT
ATCAAGAAAACGTGGAAGCAGTTTGCTAACAAAGATCGCAGTCGGCGGATTCTGTTGGC
ATGGTTGGTACTCTAATTGGATTCTATCCAAAGAAGGTACTAACCTGGGCGAAGGACAC
ACCAAATGGTCCAGAACTATGTTACGAGAACTCGCACAAAACCAAGGTGATAGTATTTCT
TGAGTGTGTTGTATGCCATTGGAGGAATCACGCTTATGCGTCGAGACATCCGAGATGG
ACTGGTGAAAAAATATGTGATATGTTTGATATCAAACGGGGGGGCCATGTCTTAGACG
TTGAGAATCCGTGCCGCTATTATGAAATCAACGATTTCTTTAGCAGTCTGTATTCGGCAT
CTGAGTCCGGTGAGACCGTTTTACCAGATTTATCCGAGGTAAAAGCCAAGTCTGATAAG
CTATTGCAGCAGAAGAAAGAAATCGCTGACGAGTTTCTAAGTGCAAAATTCTCTAACTAT
TCTGGCAGTTCCGGTGAGAACTTCTCCACCATCGGTGGTGGTTCATCTCGAAGCGGAC
TGGGTCTGTTGTTGGAAGACAGTAACGTGCTGACCCAAGCTAGAGTTGGAGTTTCAAG
AAAGGTAGACGATGAGGAGATCATGGAGCAGTTTCTGAGTGGTCTTATTGACACTGAAG
CAGAAATTGACGAGGTTGTTTCAGCCTTTTCAGCTGAATGTGAAAGAGGGGAAACAAGC
GGTACAAAGGTGTTGTGTAAACCTTTAACGCCACCAGGATTTGAGAACGTGTTGCCAGC
TGTCAAACCTTTGGTCAGCAAAGGAAAAACGGTCAAACGTGTGATTACTTCCAAGTGA
TGGGAGGTGAGAGATTACCAAAAAGGCCGGTTGTGAGTGGAGACGATTCTGTGGACGC
TAGAAGAGAGTTTCTGTACTACTTAGATGCGGAGAGAGTCGCTCAAATGATGAAATTA
TGTCTCTGTATCGTGACTATTTCGAGAGGAGTTATTTCGAACTGGAGGTCAGAATTACCCG
CACGGACTGGGAGTGTGGGATGTGGAGATGAAGAAGTGGTGCATACGTCCAGTGGTC
ACTGAACATGCTTATGTGTTCCAACCAGACAAACGTATGGATGATTGGTCCGGGATACTT
AGAAGTGGCTGTTTGGGAACGAGGTATGTTGGTCAACGACTTCGCGGTGCAAGGATG
AGTGATTATGTCATAGTTTGCATCAGACGTATCTTTGCAATAACAGGTAATAATCCTCT
CTCTTGATATTTTTAAATTATAGAATTAATTAGTTTACTTTATTCTTTACTATATGATTTAA
TAGTTTAATCTTGTTTTTGAGTAACTATTCGATTTTGATATTTGTATTCGTCTACAAAG
TTGGAAATACTGATGATATTTCTTTGAACGTGATACCTACCAATACTAATCTTACGGAA
TCTTTTAATAGAGCACTAATCAACATGGAACATAAGACCAATTCTTAAGTGTCTCTGTTG
TACAGTTCATTTTAGTAGTGCGTTTTAAGTATTATTATCTCCCTTCATGCGGGGCAATTAT
GTAGATTAAAATCGAAATTATATAAAATTTACATAAGTCTAAGTCTAGGGTCTCCAGCTA
ATTGTTATTTTTTAAACGATGTTGACTAAAGCAATAACGACGTTGACTTGTGTTAAACAG
GTTGATCTTGGACAATTTAAGTGCCCTGGATCTAGGACCAGTTAACTGTTCTTTTGAATT
AGTTGACGGTGTACCTGGTTGTGGTAAGTCGACAATGATTGTCAACTCAGCTAATCCTT
GTGTCGATGTGGTTCTCTACTGGGAGAGCAGCAACCGACGACTTGATCGAGAGATT

FIG. 30 (Continued)

18/31

CGCGAGCAAAGGTTTTCCATGCAAATTGAAAAGGAGAGTGAAGACGGTTGATTCTTTTT
TGATGCATTGTGTGCGATGGTTCTTTAACCGGAGACGTGTTGCATTTGACGAAGCTCTC
ATGGCCCATGCTGGTATGGTGTACTTTTTGCGCTCAGATAGCTGGTGCTAAACGATGTAT
CTGTCAAGGAGATCAGAATCAAATTTCTTTCAAGCCTAGGGTATCTCAAGTTGATTTGAG
GTTTTCTAGTCTGGTCGAAAGTTTGACATTGTTACAGAAAAAGAGAACTTACAGAAG
TCCAGCAGATGTGGCTGCCGTATTGAACAAGTACTATACTGGAGATGTCAGAACACATA
ACGCGACTGCTAATTCGATGACGGTGAGGAAGATTGTGTCTAAAGAACAGGTTTCTTTG
AAGCCTGGTGCTCAGTACATAACTTTCTTCAGTCTGAGAAGAAGGAGTTGGTAAATTT
GTTGGCATTGAGGAAAGTGGCAGCTAAAGTGAGTACAGTACACGAGTCGCAAGGAGAG
ACATTCAAAGATGTAGTCCTAGTCAGGACGAAACCTACGGATGACTCAATCGCTAGAGG
TCGGGAGTACTTAATCGTGGCATTGTGCGCTCACACACAATCACTTGTGTATGAAACTG
TGAAAGAGGACGATGTAAGCAAAGAGATCAGGGAAAGTGCCGCGCTTACGAAGGCGG
CTTTGGCAAGATTTTTTGTACTGAGACCGTCTTATGACGGTTTCGGTCTAGGTTTGATG
TCTTTAGACATCATGAAGGGCCTTGCGCCGTTCCAGATTCAGGTACGATTACGGACTTG
GAGATGTGGTACGACGCTTTGTTCCGGGAAATTCGTTAAGAGACTCAAGCCTAGACG
GGTATTTGGTGGCAACGACTGATTGCAATTTGCGATTAGACAATGTTACGATCAAAAGT
GGAACTGGAAAGACAAGTTTGCTGAAAAAGAAACGTTTCTGAAACCGGTTATTCGTAC
TGCTATGCCTGACAAAAGGAAGACTACTCAGTTGGAGAGTTTGTAGCATTGCAGAAAA
GGAACCAAGCGGCACCCGATCTACAAGAAAATGTGCACGCGACAGTTCTAATCGAAGA
GACGATGAAGAAGCTGAAATCTGTTGTCTACGATGTGGGAAAAATTCGGGCTGATCCTA
TTGTCAATAGAGCTCAAATGGAGAGATGGTGGAGAAATCAAAGCACAGCGGTACAGGC
TAAGGTAGTAGCAGATGTGAGAGAGTTACATGAAATAGACTATTCGTCTTACATGTATAT
GATCAAATCTGACGTGAAACCTAAGACTGATTTAACACCGCAATTTGAATACTCAGCTCT
ACAGACTGTTGTGTATCACGAGAAGTTGATCAACTCGTTGTTCCGGTCCAATTTTCAAAGA
AATTAATGAACGCAAGTTGGATGCTATGCAACCACATTTTGTGTTCAACACGAGAATGA
CATCGAGTGATTTAAACGATCGAGTGAAGTTCTTAAATACGGAAGCGGCTTACGACTTT
GTTGAGATAGACATGTCTAAATTCGACAAGTCGGCAAATCGCTTCCATTTACAACCTGCA
GCTGGAGATTTACAGGTTATTTGGGCTGGATGAGTGGGCGGCCTTCTTTGGGAGGTG
TCGCACACTCAAACCTACTGTGAGAGATATTCAAATGGTATGATGGCGCATATTTGGTA
CCAACAAAAGAGTGGAGATGCTGATACTTATAATGCAAATTCAGATAGAACACTGTGTG
CACTCTTGTCTGAATTACCATTGGAGAAAGCAGTCATGGTTACATATGGAGGAGATGAC
TCACTGATTGCGTTTTCTAGAGGAACGCGAGTTTGTGATCCGTGTCCAAAGTTGGCTAC
TAAGTGGAATTTGAGTGAAGATTTTAAAGTACGATGTCCCAATGTTTTGTGGGAAGTT
CTTGCTTAAGACGTCATCGTGTTACGAGTTCGTGCCAGATCCGGTAAAAGTTCTGACGA
AGTTGGGGAAGAGATATAAAGGATGTGCAACATTTAGCCGAGATCTACATCTCGCTG
AATGATTCCAATAGAGCTCTTGGGAACCTACATGGTGGTATCCAACTGTCCGAGTCTGT
TTCAGACCGGATTTGTACAAAGGTGATTCTGTTTCATGCGCTTTGTGCGCTATGGAAGC
ATATTAAGAGTTTTACAGCTCTGTGTACATTATTCGAGACGAAAACGATAAGGAATTGA
ACCCGGCTAAGGTTGATTGGAAGAAGGCACAGAGAGCTGTGTCAAACCTTTTACGACTG
GTAATATGGAAGACAAGTCATTGGTCACCTTGAAGAAGAAGACTTTTCAAGTCTCAAAA
TTCTCAAATCTAGGGGCCATTGAATTGTTTGTGGACGGTAGGAGGAAGAGACCGAAGT
ATTTTCACAGAAGAAGAGAACTGTCTAAATCATGTTGGTGGGAAGAAGAGTGAACAC
AAGTTAGACGTTTTTGACCAAAGGGATTACAAAATGATTAAATCTTACGCGTTTCTAAAG
GTAGTAGGTGTACAACTAGTTGTAACATCACATCTACCTGCAGATACGCCTGGGTTTCA
TCAAATCGATCTGTTGGATTGAGACTTACTGAGAAAAAGAAAGAGAGGAAAGACTATTC
AGAGATTCAAAGCTCGAGCTTGCGATAACTGTTTCAAGTTGCGCAGTACAAGGTTGAATAC
AGTATTTCCACACAGGAGAACGTACTTGATGTCTGGAAGGTGGGTTGTATTTCTGAGGG
CGTTCCGGTCTGTGACGGTACATACCCTTTTCAAGTATCGAAGTGTGCTAATATGGGTTG
CTACTGATTCGACTAGGCGCCTCAATGTGGAAGAACTGAACAGTTTCGGATTACATTGAA
GGCGATTTTACCGATCAAGAGGTTTTTCGGTGAGTTCATGTCTTTGAAACAAGTGGAGAT
GAAGACGATTGAGGCGAAGTACGATGGTCCTTACAGACCAGCTACTACTAGACCTAAG
TCATTATTGTCAAGTGAAGATGTTAAGAGAGCGTCTAATAAGAAAAACTCGTCTTAATGC
ATAAAGAAATTTATTGTCAATATGACGTGTGTACTCAAGGGTTGTGTGAATGAAGTCACT
GTTCTTGGTCACGAGACGTGTAGTATCGGTCATGCTAACAAATTGCGAAAGCAAGTTGC
TGACATGGTTGGTGTACACGTAAGGTGTGCGGAAATAATTGTGGATGGTTTGTCTGTG

FIG. 30 (Continued)

19/31

TTGTTATCAATGATTTTACTTTTGATGTGTATAATTGTTGTGGCCGTAGTCACCTTGAAAA
GTGTCGTAAACGTGTTGAAACAAGAAATCGAGAAATTTGAAACAAATTCGACGAAATC
AAGCTGAAAACATGTCTGCGACAGCTAAAAAGTCTCATAATTCGAAGACCTCTAAGAAG
AAATTCAAAGAGGACAGAGAATTTGGGACACCAAAAAAGATTTTTAAGAGATGATGTTCTT
TTCGGGATTGATCGTTTGTGCTTTTGTATTTTATTTTATATTGTTATCTGTTTCTGTGTA
TAGACTGTTTGAGATTGGCGCTTGGCCGACTCATTGTCTTACCATAGGGGAACGGACTT
TGTTTGTGTTGTTATTTTATTTGATTTTATTAAATCTCAATGATCTGAAAAGGCCTCG
AGGCTAAGAGATTATTGGGGGGTGAGTAAGTACTTTTAAAGTGATGATGGTTACAAAGG
CAAAAGGGGTAAACCCCTCGCCTACGTAAGCGTTATTACGCCCGGATCCCCCGGGGA
GCTCGAATTCGCTGAAATCACCAGTCTCTCTCTACAAATCTATCTCTCTATTTTTTCCA
TAAATAATGTGTGAGTAGTTTCCCGATAAGGGAAATTAGGGTTCTTATAGGGTTTCGCTC
ATGTGTTGAGCATATAAGAAACCCTTAGTATGTATTTGTATTTGTAAATACTTCTATTAT
CAATAAAATTTCTAATTCCTAAAACCAAATCCAGTACTAAAATCCAGATCTCCTAAAGTC
CCTATAGATCTTTGTCGTGAATATAAACCAGACACGAGACGACTAAACCTGGAGCCCGAG
ACGCCGTTTGAAGCTAGAAGTACCGCTTAGGCAGGAGGCCGTTAGGGGAAAAGATGCTA
AGGCAGGGTTGGTTACGTTGACTCCCCCGTAGGTTTGGTTTAAATATGATGAAGTGGAC
GGAAGGAAGGAGGAAGACAAGGAAGGATAAGGTTGCAGGCCCTGTGCAAGGTAAGAA
GATGGAAATTTGATAGAGGTACGCTACTATACTTATACTATACGCTAAGGGAATGCTTGT
ATTTATACCCTATACCCCTAATAACCCCTTATCAATTTAAGAAATAATCCGCATAAGCC
CCCGCTTAAAAATTGGTATCAGAGCCATGAATAGGTCTATGACCAAACTCAAGAGGAT
AAAACCTCACCAAAATACGAAAGAGTTCTTAACTCTAAAGATAAAAGATCTTTCAAGATC
AAAAGTATGTTCCCTCACACCGGAGCATGCGATATCCTCGACCTGCAGGCATGCAAGCT
TGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAAATCCAC
ACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTA
ACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCCGTTTGCCTATTGGGCCAA
AGACAAAAGGGCGACATTCAACCGATTGAGGGAGGGGAAGGTAAATATTGACGGAAATT
ATTCATTAAAGGTGAATTATCACCGTCACCGACTTGAGCCATTTGGGAATTAGAGCCAG
CAAAATCACCAAGTAGCACCATTACCATAGCAAGGCCGGAACGTCACCAATGAAACCA
TCGATAGCAGCACCGTAATCAGTAGCGACAGAATCAAGTTTGCCTTTAGCGTCAGACTG
TAGCGCGTTTTTCATCGGCATTTTCGGTCATAGCCCCCTTATTAGCGTTTGCCTCTTTTC
ATAATCAAAATCACCGGAACCAAGAGCCACCACCGGAACCGCCTCCCTCAGAGCCGCCA
CCCTCAGAACCAGCCACCCTCAGAGCCACCACCCTCAGAGCCGCCACCAGAACCACCA
CCAGAGCCGCCGCCAGCATTGACAGGAGGCCCGATCTAGTAACATAGATGACACCGC
GCGCGATAATTTATCCTAGTTTGCAGCGCTATATTTTGTCTTCTATCGCGTATTAAATGTAT
AATTGCGGGACTCTAATCATAAAAACCCATCTCATAAATAACGTCATGCATTACATGTTA
ATTATTACATGCTTAACGTAATTCAACAGAAATTATATGATAATCATCGCAAGACCGGCA
ACAGGATTCAATCTTAAGAACTTTATTGCCAAATGTTTGAACGATCGGGGATCATCCG
GGTCTGTGGCGGGAACCTCCACGAAAATATCCGAACGCAGCAAGATATCGCGGTGCATC
TCGGTCTTGCCTGGGCAGTCGCCGCCGACGCCGTTGATGTGGACGCCGGGCCCGATC
ATATTGTCGCTCAGGATCGTGCGCTTGTGCTTGTGCGCCGTTGCTGTGCTAATGATATC
GGCACCTTCGACCGCCTGTTCCGCGAGAGATCCCGTGGGCGAAGAACTCCAGCATGAG
ATCCCCGCGCTGGAGGATCATCCAGCCGGCGTCCCGGAAAACGATTCCGAAGCCCAA
CCTTTCATAGAAGGCGGCGGTGGAATCGAAATCTCGTGATGGCAGGTTGGGCGTCGCT
TGGTCGGTCATTTTGAACCCAGAGTCCCGCTCAGAAGAACTCGTCAAGAAGGCGATA
GAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTC
AGCCCATTCGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGA
TAGCGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTT
CCACCATGATATTCCGGCAAGCAGGCATCGCCATGGGTACGACGAGATCATCGCCGTC
GGGCATGCGCGCCTTGAGCCTGGCGAACAGTTCCGGCTGGCGCGAGCCCCTGATGCTC
TTCGTCCAGATCATCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCG
ATGCGATGTTTTCGCTTGGTGGTCAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCC
GCCGCATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAG
GAGATCCTGCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACA
ACGTCGAGCACAGCTGCGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCCGCGCT

FIG. 30 (Continued)

20/31

GCCTCGTCCTGCAGTTCATTACAGGGCACCGGACAGGTCCGGTCTTGACAAAAAGAACCG
GGCGCCCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTT
GTGCCCAGTCATAGCCGAATAGCCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCA
ATCCATCTTGTTCAATCATGCGAAACGATCCAGATCCGGTGCAGATTATTTGGATTGAG
AGTGAATATGAGACTCTAATTGGATACCGAGGGGAATTTATGGAACGTCAGTGGAGCAT
TTTTGACAAGAAATATTTGCTAGCTGATAGTGACCTTAGGCGACTTTTGAACGCGCAATA
ATGGTTTCTGACGTATGTGCTTAGCTCATTAACTCCAGAAACCCGCGGCTGAGTGGCT
CCTTCAACGTTGCGGTTCTGTCAGTTCCAAACGTAAACGGCTTGTCCCGCGTCATCGG
CGGGGGTGCATAACGTGACTCCCTTAATTCTCCGCTCATGATCAGATTGTCTGTTTCCCGC
CTTCAGTTTAACTATCAGTGTTTGACAGGATATATTGGCGGGTAAACCTAAGAGAAAA
GAGCGTTTATTAGAATAATCGGATATTTAAAAGGGCGTGAAAAGGTTTATCCGTTCTGTC
CATTTGTATGTGCATGCCAACCACAGGGTTCCCCAGATCTGGCGCCCGGCCAGCGAGAC
GAGCAAGATTGGCCGCCGCCGAAACGATCCGACAGCGCGCCAGCACAGGTGCGCA
GGCAAATTGCACCAACGCATACAGCGCCAGCAGAATGCCATAGTGGGCGGTGACGTC
GTTTCGAGTGAACCAGATCGCGCAGGAGGCCCGGCAGCACCGGCATAATCAGGCCGAT
GCCGACAGCGTCGAGCGCGACAGTGCTCAGAATTACGATCAGGGGTATGTTGGGTTTC
ACGTCTGGCCTCCGGACCAGCCTCCGCTGGTCCGATTGAACGCGCGGATTCTTTATCA
CTGATAAGTTGGTGGACATATTATGTTTATCAGTGATAAAGTGTCAAGCATGACAAAGTT
GCAGCCGAATACAGTGATCCGTGCCGCCCTGGACCTGTTGAACGAGGTCCGCCGTAGA
CGGTCTGACGACACGCAAACTGGCGGAACGGTTGGGGGTTTCAGCAGCCGGCGCTTTA
CTGGCACTTCAGGAACAAGCGGGCGCTGCTCGACGCACTGGCCGAAGCCATGCTGGC
GGAGAATCATACGCATTTCGGTGCCGAGAGCCGACGACGACTGGCGCTCATTTCTGATC
GGGAATGCCCGCAGCTTCAGGCAGGCGCTGCTCGCCTACCGCGATGGCGCGCGCATC
CATGCCGGCACGCGACCGGGCGCACCGCAGATGGAAACGGCCGACGCGCAGCTTCG
CTTCTCTGCGAGGCGGGTTTTTCGGCCGGGACGCCGTCAATGCGCTGATGACAATC
AGCTACTTCACTGTTGGGGCCGTGCTTGAGGAGCAGGCCGGCGACAGCGATGCCGGC
GAGCGCGGCGGCACCGTTGAACAGGCTCCGCTCTCGCCGCTGTTGCGGGCCGCGATA
GACGCCCTTCGACGAAGCCGGTCCGGACGCAGCGTTTCGAGCAGGGACTCGCGGTGATT
GTCGATGGATTGGCGAAAAGGAGGCTCGTTGTCAGGAACGTTGAAGGACCGAGAAAG
GGTGACGATTGATCAGGACCGCTGCCGGAGCGCAACCCACTCACTACAGCAGAGCCA
TGTAGACAACATCCCCCTCCCCCTTTCCACCGCGTCAGACGCCCGTAGCAGCCCGCTAC
GGGCTTTTTTCATGCCCTGCCCTAGCGTCCAAGCCTCACGGCCGCGCTCGGCCTCTCTG
GCGGCCTTCTGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCTGTT
GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCG
TAAAAAGGCCGCGTTGCTGGCGTTTTTTCATAGGCTCCGCCCCCTGACGAGCATCAC
AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCCGACCCTGCCGCTTACCGG
ATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGCGCTTTTCCGCTGCATAACCCCTG
CTTCGGGGTCATTATAGCGATTTTTTTCGGTATATCCATCCTTTTTTCGCACGATATACAGG
ATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGC
AGGATAGGTGAAGTAGGCCACCCGCGAGCGGGTGTTCTTCTTCACTGTCCCTTATT
CGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCG
GCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCA
GCCACCTATCAAGGTGTAAGTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGC
GGCGGCGGCCGGCATGAGCCTGTGCGCCTACCTGCTGGCCGTGCGCCAGGGCTACA
AAATCACGGGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCG
ACCTGGGCGCCTGGGCGGCCTGCTGAACTCTGGCTCACCGACGACCCGCGCACGG
CGCGGTTCCGGTATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGG
ACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTT
TTAGCCGCTAAAACGGCCGGGGGGTGC GCGTGATTGCCAAGCACGTCCCCATGCGC
TCCATCAAGAAGAGCGACTTCGCGGAGCTGGTGAAGTACATCACCGACGAGCAAGGCA
AGACCGAGCGCCTTTGCGACGCTCACCGGGCTGGTTGCCCTCGCCGCTGGGCTGGCG
GCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTGCAAGCCGTGTGCGAGACAC
CGCGGCCCGCGCGCTTGTGGATACCTCGCGGAAACTTGGCCCTCACTGACAGATGA

FIG. 30 (Continued)

21/31

GGGGCGGACGTTGACACTTGAGGGGGCCGACTCACCCGGCGCGGGCGTTGACAGATGA
GGGGCAGGCTCGATTTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCG
AAAACGCCTGATTTTACGCGAGTTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGT
GCCCTGCGGTATTGACACTTGAGGGGGCGCGACTACTGACAGATGAGGGGGCGCGATCC
TTGACACTTGAGGGGGCAGAGTGCTGACAGATGAGGGGGCGCACCTATTGACATTTGAGG
GGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTTCCGCCCGTTTTTCGGCCA
CCGCTAACCTGTCTTTTAACTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAAACAG
GGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCTTCTCG
AACCTCCCGGCCCGCTAACGCGGGCCTCCCATCCCCCAGGGGGCTGCGCCCCCTCGG
CCGCGAACGGCCTCACCCCAAAAATGGCAGCGCTGGCAGTCCTTGCCATTGCCGGGA
TCGGGGCAGTAACGGGATGGGCGATCAGCCCGAGCGCGACGCCCGGAAGCATTGAC
GTGCCGACAGGTGCTGGCATCGACATTCAGCGACCAGGTGCCGGGCAGTGAGGGCGG
CGGCCTGGGTGGCGGCCTGCCCTTCACTTCGGCCGTCGGGGCATTACGGACTTCAT
GGCGGGGGCCGGCAATTTTTACCTTGGGCATTCTTGGCATAGTGGTCGCGGGTGCCGT
GCTCGTGTTCCGGGGGTGCGATAAACCCAGCGAACCATTTGAGGTGATAGGTAAGATTA
TACCGAGGTATGAAAACGAGAATTGGACCTTTACAGAATTACTCTATGAAGCGCCATAT
TAAAAAAGCTACCAAGACGAAGAGGATGAAGAGGATGAGGAGGCAGATTGCCTTGAAT
ATATTGACAATACTGATAAGATAATATATCTTTTATATAGAAGATATCGCCGTATGTAAGG
ATTTGAGGGGGCAAGGCATAGGCAGCGCGCTTATCAATATATCTATAGAATGGGCAAA
GCATAAAAACTTGATGGACTAATGCTTGAAACCCAGGACAATAACCTTATAGCTTGTA
ATTCTATCATAATTGGGTAATGACTCCAACCTTATTGATAGTGTTTTATGTTGAGATAATGC
CCGATGACTTTGTGATGCAGCTCCACCGATTTTGAGAACGACAGCGACTTCCGTCCCA
GCCGTGCCAGGTGCTGCCTCAGATTCAGGTTATGCCGCTCAATTCGCTGCGTATATCG
CTTGCTGATTACGTGCAGCTTTCCCTTCAGGCGGGATTTCATACAGCGGCCAGCCATCC
GTCATCCATATCACACGTCAAAGGGTGACAGCAGGCTCATAAGACGCCCCAGCGTCG
CCATAGTGCGTTACCCGAATACGTGCGCAACAACCGTCTTCCGGAGACTGTCATACGC
GTAAAACAGCCAGCGCTGGCGCGATTTAGCCCCGACATAGCCCCACTGTTTCGTCCATT
TCCGCGCAGACGATGACGTCACTGCCCGGCTGTATGCGCGAGGTTACCGACTGCGGC
CTGAGTTTTTTAAGTGACGTAAAATCGTGTTGAGGCCAACGCCCATATGCGGGCTGTT
GCCCGGCATCCAACGCCATTTCATGGCCATATCAATGATTTTCTGGTGCGTACCGGGTT
GAGAAGCGGTGTAAGTGAAGTGCAGTTGCCATGTTTTACGGCAGTGAGAGCAGAGATA
GCGCTGATGTCCGGCGGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAG
GAGGGACAGCTGATAGACACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACAC
TAAATCAGTAAGTTGGCAGCATCACCCATAATTGTGGTTTTCAAATCGGCTCCGTCGAT
ACTATGTTATACGCCAACTTTGAAAACAACCTTTGAAAAAGCTGTTTTCTGGTATTTAAGG
TTTTAGAATGCAAGGAACAGTGAATTGGAGTTCGTCTTGTTATAATTAGCTTCTTGGGGT
ATCTTTAAATACTGTAGAAAAGAGGAAGGAAATAATAAATGGCTAAAATGAGAATATCAC
CGGAATTGAAAAAACTGATCGAAAAATACCGCTGCGTAAAAGATACGGAAGGAATGTCT
CCTGCTAAGGTATATAAGCTGGTGGGAGAAAATGAAAACCTATATTTAAAAATGACGGA
CAGCCGGTATAAAGGGACCACCTATGATGTGGAACGGGAAAAGGACATGATGCTATGG
CTGGAAGGAAAGCTGCCTGTTCCAAAGGTCCTGCACCTTTGAACGGCATGATGGCTGGA
GCAATCTGCTCATGAGTGAGGCCGATGGCGTCTTTGCTCGGAAGAGTATGAAGATGA
ACAAAGCCCTGAAAAGATTATCGAGCTGTATGCGGAGTGCATCAGGCTCTTTCCTCCA
TCGACATATCGGATTGTCCCTATACGAATAGCTTAGACAGCCGCTTAGCCGAATTGGAT
TACTTACTGAATAACGATCTGGCCGATGTGGATTGCGAAAACCTGGGAAGAAGACACTCC
ATTTAAAGATCCGCGCGAGCTGTATGATTTTTTAAAGACGGAAAAGCCCCGAAGAGGAAC
TTGTCTTTTCCACGCGCACCTGGGAGACAGCAACATCTTTGTGAAAGATGGCAAAGTA
AGTGGCTTTATTGATCTTGGGAGAAGCGGCAGGGCGGACAAGTGGTATGACATTGCCT
TCTGCGTCCGGTCGATCAGGGAGGATATCGGGGAAGAACAGTATGTGCGAGCTATTTTT
TGACTTACTGGGGATCAAGCCTGATTGGGAGAAAATAAAAATATTATTTTTACTGGATGA
ATTGTTTTAGTACCTAGATGTGGCGCAACGATGCCGGCGACAAGCAGGAGCGCACCGA
CTTCTTCCGCATCAAGTGTTTTGGCTCTCAGGCCGAGGCCACGGCAAGTATTTGGGC
AAGGGGTGCTGGTATTCGTGCAGGGCAAGATTCCGAATACCAAGTACGAGAAGGACG
GCCAGACGGTCTACGGGACCGACTTCATTGCCGATAAGGTGGATTATCTGGACACCAA
GGCACCAGGCGGGTCAAATCAGGAATAAGGGCACATTGCCCGGCGTGAGTCGGGGC

FIG. 30 (Continued)

22/31

AATCCCGCAAGGAGGGTGAATGAATCGGACGTTTGACCGGAAGGCATACAGGCAAGAA
CTGATCGACGCGGGGTTTTCCGCCGAGGATGCCGAAACCATCGCAAGCCGACCGTC
ATGCGTGCGCCCCGCGAAACCTTCCAGTCCGTCGGCTCGATGGTCCAGCAAGCTACG
GCCAAGATCGAGCGCGACAGCGTGCAACTGGCTCCCCCTGCCCTGCCCGCGCCATCG
GCCGCCGTGGAGCGTTCGCGTCGTCTCGAACAGGAGGCGGCAGGTTTGGCGAAGTCG
ATGACCATCGACACGCGAGGAACTATGACGACCAAGAAGCGAAAAACCGCCGGCGAG
GACCTGGCAAAACAGGTCAGCGAGGCCAAGCAGGCCGCGTTGCTGAAACACACGAAG
CAGCAGATCAAGGAAATGCAGCTTTCCTTGTTTCGATATTGCGCCGTGGCCGGACACGA
TGCGAGCGATGCCAAACGACACGGCCCCGCTCTGCCCTGTTACCCACGCGCAACAAGA
AAATCCCGCGCGAGGCGCTGCAAAACAAGGTCATTTTCCACGTCAACAAGGACGTGAA
GATCACCTACACCGGCGTCGAGCTGCGGGCCGACGATGACGAACTGGTGTGGCAGCA
GGTGTGGAGTACGCGAAGCGCACCCCTATCGGCGAGCCGATCACCTTCACGTTCTAC
GAGCTTTGCCAGGACCTGGGCTGGTCGATCAATGGCCGGTATTACACGAAGGCCGAG
GAATGCCTGTCGCGCCTACAGGCGACGGCGATGGGCTTCACGTCCGACCGCGTTGGG
CACCTGGAATCGGTGTCGCTGCTGCACCGCTTCCGCGTCCTGGACCGTGGCAAGAAAA
CGTCCCGTTGCCAGGTCCTGATCGACGAGGAAATCGTCGTGCTGTTTGTGGCGACCA
CTACACGAAATTCATATGGGAGAAGTACCGCAAGCTGTCGCCGACGGCCCGACGGATG
TTCGACTATTTAGCTCGCACCCGGGAGCCGTACCCGCTCAAGCTGGAAACCTTCCGCC
TCATGTGCGGATCGGATTCCACCCGCGTGAAGAAGTGGCGCGAGCAGGTGCGCGAAG
CCTGCGAAGAGTTGCGAGGCAGCGGCCTGGTGGAAACACGCCTGGGTCAATGATGACC
TGGTGCATTGCAAACGCTAGGGCCTTGTGGGGTCAGTTCCGGCTGGGGGTTTCAGCAG
CCAGCGCTTTACTGGCATTTCAGGAACAAGCGGGCACTGCTCGACGCACTTGCTTCGC
TCAGTATCGCTCGGGACGCACGGCGCGCTCTACGAACTGCCGATAAACAGAGGATTAA
AATTGACAATTGTGATTAAGGCTCAGATTCGACGGCTTGGAGCGGCCGACGTGCAGGA
TTTCCGCGAGATCCGATTGTGCGCCCTGAAGAAAGCTCCAGAGATGTTTCGGGTCCGTT
TACGAGCACGAGGAGAAAAAGCCCATGGAGGCGTTCGCTGAACGGTTGCGAGATGCC
GTGGCATTGCGCGCCTACATCGACGGCGAGATCATTGGGCTGTGCGTCTTCAAACAGG
AGGACGGCCCCAAGGACGCTCACAAGGCGCATCTGTCCGGCGTTTTCGTGGAGCCCC
AACAGCGAGGCCGAGGGGTGCGCCGTATGCTGCTGCGGGCGTTGCCGGCGGGTTTAT
TGCTCGTGATGATCGTCCGACAGATTCCAACGGGAATCTGGTGGATGCGCATCTTCAT
CCTCGGCGCACTTAATATTTCCGCTATTCTGGAGCTTGTTGTTTATTTCCGGTCTACCGCCT
GCCGGGCGGGGTGCGGGCGACGGTAGGCGCTGTGCAGCCGCTGATGGTTCGTGTTCA
TCTCTGCCGCTCTGCTAGGTAGCCCGATACGATTGATGGCGGTCTTGGGGGCTATTTG
CGAACTGCGGGCGTGGCGCTGTTGGTGTGACACCAAACGCAGCGCTAGATCCTGT
CGGCGTGCAGCGGGCCTGGCGGGGGCGGTTTCCATGGCGTTCGGAACCGTGCTGA
CCCGCAAGTGGCAACCTCCCGTGCTCTGCTCACCTTTACCGCCTGGCAACTGGCGGC
CGGAGGACTTCTGCTCGTTCCAGTAGCTTTAGTGTGTTGATCCGCCAATCCCGATGCCTA
CAGGAACCAATGTTCTCGGCCTGGCGTGGCTCGGCCTGATCGGAGCGGGTTTAACCTA
CTTCCTTTGTTCCGGGGGATCTCGCGACTCGAACCTACAGTTGTTTCCTTACTGGGCT
TTCTCAGCCCCAGATCTGGGGTTCGATCAGCCGGGGATGCATCAGGCCGACAGTCGGA
ACTTCGGGTCCCCGACCTGTACCATTCGGTGAGCAATGGATAGGGGAGTTGATATCGT
CAACGTTCACTTCTAAAGAAATAGCGCCACTCAGCTTCCTCAGCGGCTTTATCCAGCGA
TTTCTATTATGTCGGCATAGTTCTCAAGATCGACAGCCTGTCACGGTTAAGCGAGAAA
TGAATAAGAAGGCTGATAATTCGGATCTCTGCGAGGGAGATGATATTTGATCACAGGCA
GCAACGCTCTGTCAATCGTTACAATCAACATGCTACCCTCCGCGAGATCATCCGTGTTTC
AAACCCGGCAGCTTAGTTGCCGTTCTTCCGAATAGCATCGGTAACATGAGCAAAGTCTG
CCGCCTTACAACGGCTCTCCCGCTGACGCCGTCCCGGACTGATGGGCTGCCTGTATC
GAGTGGTGATTTTGTGCCGAGCTGCCGGTCGGGGAGCTGTTGGCTGGCTGGTGGCAG
GATATATTGTGGTGTAAACAAATTGACGCTTAGACAACCTAATAACACATTGCGGACGTT
TTAATGTACTGGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCC
TTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGC
AGGCGAAAATCCTGTTTGATGGTGGTTCCGAAATCGGCAAAATCCCTTATAAATCAAAA
GAATAGCCCGAGATAGGGTTGAGTGTGTTCCAGTTTGAACAAGAGTCCACTATTAAA
GAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTA
CGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCG

FIG. 30 (Continued)

23/31

GAACCCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGC
GAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCCATTCAGGCTGCGCAACTGTT
GGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGAT
GTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAA
AACGACGGCCAGTGAATTCGAGCTCGGTACCCCCC

FIG. 30 (Continued)

TRV2 (SEQ ID No. 27):

GTCTCAGAAGACCAAAGGGCTATTGAGACTTTTTCAACAAAGGGTAATTTTCGGGAAACCT
CCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAG
GTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCATTCAAGATGCCTCT
GCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAGAAAGAA
GACGTCCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGACATCTCCACTGACGTAAAG
GGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCTTCTATATAAGGAAGTTCATT
TCATTTGGAGAGGACAGCCCAAGCTTTCTAGAGGATCCATAAAACATTTCAATCCTTTGA
ACGCGGTAGAACGTGCTAATTGGATTTTGGTGAGAACGCGGTAGAACGTACTTATCACC
TACAGTTTTATTTGTTTTCTTTTTGGTTTAATCTATCCAGCTTAGTACCGAGTGGGGGA
AAGTGACTGGTGTGCCTAAAACCTTTTTCTTTGATACTTTGTAAAAATACATACAGATACA
ATGGCGAACGGTAACCTTCAAGTTGTCTCAATTGCTCAATGTGGACGAGATGTCTGCTGA
GCAGAGGAGTCATTTCTTTGACTTGATGCTGACTAAACCTGATTGTGAGATCGGGCAAA
TGATGCAAAGAGTTGTTGTTGATAAAGTCGATGACATGATTAGAGAAAGAAAGACTAAA
GATCCAGTGATTGTTTCATGAAGTTCTTTCTCAGAAGGAACAGAACAAGTTGATGGAAATT
TATCCTGAATTCATATCGTGTTTAAAGACGACAAAAACATGGTTCATGGGTTTGCGGCT
GCTGAGCGAAAACCTACAAGCTTTATTGCTTTTAGATAGAGTTCCTGCTCTGCAAGAGGT
GGATGACATCGGTGGTCAATGGTCGTTTTGGGTAAGTACTAGAGGTGAGAAAAGGATTTCATT
CCTGTTGTCCAAATCTAGATATTCGGGATGATCAGAGAGAAATTTCTCGACAGATATTTT
TACTGCTATTGGTGATCAAGCTAGAAGTGGTAAGAGACAGATGTCGGAGAATGAGCTG
TGGATGTATGACCAATTTCTGTGAAAATATTGCTGCGCCTAACGCGGTTAGGTGCAATAA
TACATATCAGGGTTGTACATGTAGGGGTTTTTCTGATGGTAAGAAGAAAGGCGCGCAGT
ATGCGATAGCTCTTCACAGCCTGTATGACTTCAAGTTGAAAGACTTGATGGCTACTATG
GTTGAGAAGAAAACCTAAAGTGGTTCATGCTGCTATGCTTTTTGCTCCTGAAAGTATGTTA
GTGGACGAAGGTCCATTACCTTCTGTTGACGGTTACTACATGAAGAAGAACGGGAAGAT
CTATTTTCGGTTTTTGAGAAAGATCCTTCCTTTCTTACATTCATGACTGGGAAGAGTACAA
GAAGTATCTACTGGGGAAGCCAGTGAGTTACCAAGGGAATGTGTTCTACTTCGAACCGT
GGCAGGTGAGAGGAGACACAATGCTTTTTTCGATCTACAGGATAGCTGGAGTTCCGAG
GAGGTCTCTATCATCGCAAGAGTACTACCGAAGAATATATATCAGTAGATGGGAAAGCA
TGGTTGTTGTCCCAATTTTCGATCTGGTCAATCAACGCGAGAGTTGGTCAAGAAAGAC
CTGTTTGTAGAGAAACAATTCATGGACAAGTGTGGATTACATAGCTAGGTATCTGAC
CAGCAGCTGACCATAAGCAATGTTAAATCATACTTGAGTTCAAATAATTGGGTCTTATTC
ATAAACGGGGCGGCCGTGAAGAACAAGCAAAGTGTAGATTCTCGAGATTTACAGTTGTT
GGCTCAAACCTTTGCTAGTGAAGGAACAAGTGGCGAGACCTGTCATGAGGGAGTTGCGT
GAAGCAATTCTGACTGAGACGAAACCTATCACGTCATTGACTGATGTGCTGGGTTTAAT
ATCAAGAAAACCTGTGGAAGCAGTTTGCTAACAAGATCGCAGTCGGCGGATTCTGTTGGC
ATGGTTGGTACTCTAATTGGATTCTATCCAAAGAAGGTAACCTGGGCGAAGGACAC
ACCAAATGGTCCAGAATCTATGTTACGAGAACTCGCACAAAACCAAGGTGATAGTATTTT
TGAGTGTGTTGTATGCCATTGGAGGAATCACGCTTATGCGTCGAGACATCCGAGATGG
ACTGGTGAAAAAACTATGTGATATGTTTGATATCAAACGGGGGGCCCATGTCTTAGACG
TTGAGAATCCGTGCCGCTATTATGAAATCAACGATTTCTTTAGCAGTCTGTATTCGGCAT
CTGAGTCCGGTGAGACCGTTTTACCAGATTTATCCGAGGTAAGGCAAGTCTGATAAG
CTATTGCAGCAGAAGAAAGAAATCGCTGACGAGTTTCTAAGTGCAAATTTCTCTAACTAT
TCTGGCAGTTCCGGTGAGAACTTCTCCACCATCGGTGGTTCGTTTCATCTCGAAGCGGAC
TGGGTCTGTTGTTGGAAGACAGTAACGTGCTGACCCAAGCTAGAGTTGGAGTTTCAAG
AAAGGTAGACGATGAGGAGATCATGGAGCAGTTTCTGAGTGGTCTTATTGACACTGAAG
CAGAAATTGACGAGGTTGTTTCAGCCTTTTCAGCTGAATGTGAAAGAGGGGGAACAAGC

FIG. 31

24/31

GGTACAAAGGTGTTGTGTAAACCTTTAACGCCACCAGGATTTGAGAACGTGTTGCCAGC
TGTCAAACCTTTGGTCAGCAAAGGAAAAACGGTCAAACGTGTCGATTACTTCCAAGTGA
TGGGAGGTGAGAGATTACCAAAAAGGCCGGTTGTCACTGGAGACGATTCTGTGGACGC
TAGAAGAGAGTTTCTGTACTACTTAGATGCGGAGAGAGTCGCTCAAAATGATGAAATTA
TGTCTCTGTATCGTGACTATTCGAGAGGAGTTATTCGAACTGGAGGTCAGAATTACCCG
CACGGACTGGGAGTGTGGGATGTGGAGATGAAGAAGTGGTGCATACGTCCAGTGGTC
ACTGAACATGCTTATGTGTTCCAACAGACAAACGTATGGATGATTGGTCGGGATACTT
AGAAGTGGCTGTTTGGGAACGAGGTATGTTGGTCAACGACTTCGCGGTGCAAAGGATG
AGTGATTATGTCATAGTTTGCGATCAGACGTATCTTTGCAATAACAGGTAATAATCCTCT
CTCTTGATATTTTTAAATTATAGAATTAATTAGTTTACTTTATTCTTTACTATATGATTTAA
TAGTTTAATCTTGTTTTTGAGTAACTATTCGATTTTGATATTTGTATTCGTCCTACAAAG
TTGGAAATACTGATGATATTTCTTTGAACGTGATACCTACCAATACTAATCTTACGGAA
TCTTTTAATAGAGCACTAATCAACATGGAACATAAGACCAATTCTTAAGTGTCTCTGTTG
TACAGTTCATTTTAGTAGTGCGTTTAAAGTATTATTATCTCCCTTCATGCGGGGCAATTAT
GTAGATTAAAATCGAAATTATATAAAATTTACATAAGTCTAAGTCTAGGGTCTCCAGCTA
ATTGTTATTTTTTAAACGATGTTGACTAAAGCAATAACGACGTTGACTTGTGTTAAACAG
GTTGATCTTGGACAATTTAAGTGCCCTGGATCTAGGACCAGTTAACTGTTCTTTTGAATT
AGTTGACGGTGTACCTGGTTGTGGTAAGTCGACAATGATTGTCAACTCAGCTAATCCTT
GTGTCGATGTGGTTCTCTCTACTGGGAGAGCAGCAACCGACGACTTGATCGAGAGATT
CGCGAGCAAAGGTTTTCCATGCAAATTGAAAAGGAGAGTGAAGACGGTTGATTCTTTTT
TGATGCATTGTGTCGATGGTTCTTTAACCGGAGACGTGTTGCATTTTCGACGAAGCTCTC
ATGGCCCATGCTGGTATGGTGTACTTTTGCGCTCAGATAGCTGGTGCTAAACGATGTAT
CTGTCAAGGAGATCAGAATCAAATTTCTTTCAAGCCTAGGGTATCTCAAGTTGATTTGAG
GTTTTCTAGTCTGGTTCGGAAGTTTGACATTGTTACAGAAAAAAGAGAACTTACAGAAG
TCCAGCAGATGTGGCTGCCGTATTGAACAAGTACTATACTGGAGATGTCAGAACACATA
ACGCGACTGCTAATTCGATGACGGTGAGGAAGATTGTGTCTAAAGAACAGGTTTCTTTG
AAGCCTGGTGCTCAGTACATAACTTTCTTCAGTCTGAGAAGAAGGAGTTGGTAAATTT
GTTGGCATTGAGGAAAGTGGCAGCTAAAGTGAGTACAGTACACGAGTCGCAAGGAGAG
ACATTCAAAGATGTAGTCCTAGTCAGGACGAAACCTACGGATGACTCAATCGCTAGAGG
TCGGGAGTACTTAATCGTGGCATTGTGCGCTCACACACAATCACTTGTGTATGAAACTG
TGAAAGAGGACGATGTAAGCAAAGAGATCAGGGAAAGTGCCGCGCTTACGAAGGCGG
CTTTGGCAAGATTTTTTGTACTGAGACCGTCTTATGACGGTTTCGGTCTAGGTTTGATG
TCTTTAGACATCATGAAGGGCCTTGCGCCGTTCCAGATTCAGGTACGATTACGGACTTG
GAGATGTGGTACGACGCTTTGTTTCCGGGAAATTCGTTAAGGAGACTCAAGCCTAGACG
GGTATTTGGTGGCAACGACTGATTGCAATTTGCGATTAGACAATGTTACGATCAAAAGT
GGAAACTGGAAAGACAAGTTTGCTGAAAAAGAAACGTTTCTGAAACCGGTTATTCGTAC
TGCTATGCCTGACAAAAGGAAGACTACTCAGTTGGAGAGTTTGTAGCATTGCAGAAAA
GGAACCAAGCGGCACCCGATCTACAAGAAAATGTGCACGCGACAGTTCTAATCGAAGA
GACGATGAAGAAGCTGAAATCTGTTGTCTACGATGTGGGAAAAATTCGGGCTGATCCTA
TTGTCAATAGAGCTCAAATGGAGAGATGGTGGAGAAATCAAAGCACAGCGGTACAGGC
TAAGGTAGTAGCAGATGTGAGAGAGTTACATGAAATAGACTATTCGTCTTACATGTATAT
GATCAAATCTGACGTGAAACCTAAGACTGATTTAACACCGCAATTTGAATACTCAGCTCT
ACAGACTGTTGTGTATCACGAGAAGTTGATCAACTCGTTGTTCCGGTCCAATTTTCAAAGA
AATTAATGAACGCAAGTTGGATGCTATGCAACCACATTTTGTGTTCAACACGAGAATGA
CATCGAGTGATTTAAACGATCGAGTGAAGTTCTTAAATACGGAAGCGGCTTACGACTTT
GTTGAGATAGACATGTCTAAATTCGACAAGTCGGCAAATCGCTTCCATTTACAACCTGCA
GCTGGAGATTTACAGGTTATTTGGGCTGGATGAGTGGGCGGCCTTCTTTGGGAGGTG
TCGCACACTCAAACCTACTGTGAGAGATATTCAAATGGTATGATGGCGCATATTTGGTA
CCAACAAAAGAGTGGAGATGCTGATACTTATAATGCAAATTCAGATAGAACACTGTGTG
CACTCTTGTCTGAATTACCATTGGAGAAAGCAGTCATGGTTACATATGGAGGAGATGAC
TCACTGATTGCGTTTTCTAGAGGAACGCAAGTTTGTGATCCGTGTCCAAAGTTGGCTAC
TAAGTGGAATTTTCGAGTGCAAGATTTTTAAGTACGATGTCCCAATGTTTTGTGGGAAGTT
CTTGCTTAAGACGTCATCGTGTTACGAGTTCGTGCCAGATCCGGTAAAAGTTCTGACGA
AGTTGGGGAAAAAGAGTATAAAGGATGTGCAACATTTAGCCGAGATCTACATCTCGCTG
AATGATTCCAATAGAGCTCTTGGGAACACATGGTGGTATCCAAACTGTCCGAGTCTGT

FIG. 31 (Continued)

25/31

TTCAGACCGGTATTTGTACAAAGGTGATTCTGTTTCATGCGCTTTGTGCGCTATGGAAGC
ATATTAAGAGTTTTACAGCTCTGTGTACATTATTCCGAGACGAAAACGATAAGGAATTGA
ACCCGGCTAAGGTTGATTGGAAGAAGGCACAGAGAGCTGTGTCAAACCTTTTACGACTG
GTAATATGGAAGACAAGTCATTGGTCACCTTGAAGAAGAAGACTTTTTCGAAGTCTCAAAA
TTCTCAAATCTAGGGGCCATTGAATTGTTTGTGGACGGTAGGAGGAAGAGACCGAAGT
ATTTTCACAGAAGAAGAGAAACTGTCCTAAATCATGTTGGTGGGAAGAAGAGTGAACAC
AAGTTAGACGTTTTTACCAAAGGGATTACAAAATGATTAAATCTTACGCGTTTTCTAAAG
GTAGTAGGTGTACAACCTAGTTGTAACATCACATCTACCTGCAGATACGCCTGGGTTCAT
TCAAATCGATCTGTTGGATTTCGAGACTTACTGAGAAAAAGAAAGAGAGGAAAGACTATTC
AGAGATTCAAAGCTCGAGCTTGCGATAACTGTTTCAGTTGCGCAGTACAAGGTTGAATAC
AGTATTTCCACACAGGAGAACGTACTTGATGTCTGGAAGGTGGGTTGTATTTCTGAGGG
CGTTCCGGTCTGTGACGGTACATACCCTTTCAGTATCGAAGTGTGCGCTAATATGGGTTG
CTACTGATTTCGACTAGGCGCCTCAATGTGGAAGAACTGAACAGTTCGGATTACATTGAA
GGCGATTTTACCGATCAAGAGGTTTTCCGGTGAGTTCATGTCTTTGAAACAAGTGGAGAT
GAAGACGATTGAGGCGAAGTACGATGGTCCTTACAGACCAGCTACTACTAGACCTAAG
TCATTATTGTCAAGTGAAGATGTTAAGAGAGCGTCTAATAAGAAAAAACTCGTCTTAATGC
ATAAAGAAATTTATTGTCAATATGACGTGTGTACTCAAGGGTTGTGTGAATGAAGTCACT
GTTCTTGGTCACGAGACGTGTAGTATCGGTTCATGCTAACAAATTGCGAAAGCAAGTTGC
TGACATGGTTGGTGTACACGTAGGTGTGCGGAAATAATTGTGGATGGTTTGTCTGTG
TTGTTATCAATGATTTTACTTTTGATGTGTATAATTGTTGTGGCCGTAGTCACCTTGAAAA
GTGTCGTAAACGTGTTGAAACAAGAAATCGAGAAATTTGAAACAATTCGACGAAATC
AAGCTGAAACATGTCTGCGACAGCTAAAAAGTCTCATAATTCGAAGACCTCTAAGAAG
AAATTCAAAGAGGACAGAGAATTTGGGACACCAAAAAGATTTTAAAGAGATGATGTTCT
TTCGGGATTGATCGTTTGTTTGCTTTTTGATTTTATTTTATATTGTTATCTGTTTCTGTGTA
TAGACTGTTTGAGATTGGCGCTTGCGCCGACTCATTGTCTTACCATAGGGGAACGGACTT
TGTTTGTGTTGTTATTTTATTTGATTTTATTTAAATTTCTCAATGATCTGAAAAGGCCTCG
AGGCTAAGAGATTATTGGGGGGTGAGTAAGTACTTTTAAAGTGATGATGGTTACAAAGG
CAAAAGGGGTAAACCCCTCGCCTACGTAAGCGTTATTACGCCCGGATCCCCCGGGGA
GCTCGAATTCGCTGAAATCACCGAGTCTCTCTCTACAAATCTATCTCTCTATTTTTTCCA
TAAATAATGTGTGAGTAGTTTCCCGATAAGGGAAATTAGGGTTCTTATAGGGTTTCGCTC
ATGTGTTGAGCATATAAGAAACCCTTAGTATGTATTTGTATTTGTAAATACTTCTATTAT
CAATAAAATTTCTAATTCCTAAAACCAAATCCAGTACTAAAATCCAGATCTCCTAAAGTC
CCTATAGATCTTTGTGCTGAATATAAACCAGACACGAGACGACTAAACCTGGAGCCCAG
ACGCCGTTTCGAAGCTAGAAGTACCGCTTAGGCAGGAGGCCGTTAGGGAAAAGATGCTA
AGGCAGGGTTGGTTACGTTGACTCCCCCGTAGGTTTGGTTTAAATATGATGAAGTGGAC
GGAAGGAAGGAGGAAGACAAGGAAGGATAAGGTTGCAGGCCCTGTGCAAGGTAAGAA
GATGGAAATTTGATAGAGGTACGCTACTATACTTATACTATACGCTAAGGGAATGCTTGT
ATTTATACCCTATACCCCTAATAACCCCTTATCAATTTAAGAAATAATCCGCATAAGCC
CCCGCTTAAAAATTGGTATCAGAGCCATGAATAGGTCTATGACCAAACTCAAGAGGAT
AAAACCTCACCAAAATACGAAAGAGTTCTTAACCTCTAAAGATAAAAGATCTTTCAAGATC
AAAACCTAGTTCCCTCACACCGGAGCATGCGATATCCTCGACCTGCAGGCATGCAAGCT
TGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCAC
ACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTA
ACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGGCGGTTTGCCTATTGGGCCAA
AGACAAAAGGGCGACATTCAACCGATTGAGGGAGGGGAAGGTAAATATTGACGGAAATT
ATTCATTAAAGGTGAATTATCACCGTCACCGACTTGAGCCATTTGGGAATTAGAGCCAG
CAAAATCACCGAGTAGCACCATTAACATTAGCAAGGCCGGAACGTCACCAATGAAACCA
TCGATAGCAGCACCGTAATCAGTAGCGACAGAATCAAGTTTGCCTTTAGCGTTCAGACTG
TAGCGCGTTTTTCATCGGCATTTTCGGTCATAGCCCCCTTATTAGCGTTTGCCTATCTTTT
ATAATCAAAATCACCGGAACCGAGGCCACCGGGAACCGCCTCCCTCAGAGCCGCCA
CCCTCAGAACCGCCACCCTCAGAGGCCACCACCCTCAGAGCCGCCACCAGAACCACCA
CCAGAGCCGCCGCCAGCATTGACAGGAGGCCCGATCTAGTAACATAGATGACACCGC
GCGCGATAATTTATCCTAGTTTGCAGCGCTATATTTTGTCTTCTATCGCGTATTAAATGTAT
AATTGCGGGACTCTAATCATAAAAACCATCTCATAAATAACGTCATGCATTACATGTTA

FIG. 31 (Continued)

26/31

ATTATTACATGCTTAACGTAATTCAACAGAAATTATATGATAATCATCGCAAGACCGGCA
ACAGGATTCAATCTTAAGAACTTTATTGCCAAATGTTTGAACGATCGGGGATCATCCG
GGTCTGTGGCGGGAACCTCCACGAAAATATCCGAACGCAGCAAGATATCGCGGTGCATC
TCGGTCTTGCTGGGCAGTCGCCGCCGACGCCGTTGATGTGGACGCCGGGCCCCGATC
ATATTGTGCTCAGGATCGTGCGCTTGTGCTTGTGCGGCCGTTGCTGTGTAATGATATC
GGCACCTTCGACCGCCTGTTCCGCAGAGATCCCGTGGGCGAAGAACTCCAGCATGAG
ATCCCCGCGCTGGAGGATCATCCAGCCGGCGTCCCGGAAAACGATTCCGAAGCCCCAA
CCTTTCATAGAAGGCGGCGGTGGAATCGAAATCTCGTGATGGCAGGTTGGGCGTCGCT
TGGTCGGTCATTTGCAACCCCAAGAGTCCCGCTCAGAAGAACTCGTCAAGAAGGCGATA
GAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTC
AGCCCATTCGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCTGA
TAGCGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTT
CCACCATGATATTCGGCAAGCAGGCATCGCCATGGGTACGACGAGATCATCGCCGTC
GGGCATGCGCGCCTTGAGCCTGGCGAACAGTTCCGGCTGGCGCGAGCCCCTGATGCTC
TTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCG
ATGCGATGTTTCGCTTGGTGGTGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCC
GCCGCATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAG
GAGATCCTGCCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACA
ACGTCGAGCACAGCTGCGCAAGGAACGCCCGTCTGTGGCCAGCCACGATAGCCGCGCT
GCCTCGTCTGCAGTTCATTGAGGGCACCGGACAGGTGCGTCTTGACAAAAAGAACC
GGCGCCCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTT
GTGCCCAGTCATAGCCGAATAGCCTCTCCACCCAAAGCGGCCGGAGAACCTGCGTGCA
ATCCATCTTGTTCAATCATGCGAAACGATCCAGATCCGGTGCAGATTATTTGGATTGAG
AGTGAATATGAGACTCTAATTGGATACCGAGGGGAATTTATGGAACGTCAGTGGAGCAT
TTTTGACAAGAAATATTTGCTAGCTGATAGTGACCTTAGGCGACTTTTGAACGCGCAATA
ATGGTTTCTGACGTATGTGCTTAGCTCATTAAACTCCAGAAACCCGCGGCTGAGTGGCT
CCTTCAACGTTGCGGTTCTGTGAGTTCCAAACGTAAAACGGCTTGTCCCGCGTCATCGG
CGGGGGTCATAACGTGACTCCCTTAATTCTCCGCTCATGATCAGATTGTGTTTTCCGCG
CTTCAGTTTAAACTATCAGTGTTTGACAGGATATATTGGCGGGTAAACCTAAGAGAAAA
GAGCGTTTATTAGAATAATCGGATATTTAAAAGGGCGTGAAAAGGTTTATCCGTTCTGTC
CATTTGTATGTGCATGCCAACCACAGGGTTCCCCAGATCTGGCGCCGGCCAGCGAGAC
GAGCAAGATTGGCCGCCGCCGAAACGATCCGACAGCGCGCCAGCACAGGTGCGCA
GGCAAATTGCACCAACGCATACAGCGCCAGCAGAATGCCATAGTGGGCGGTGACGTC
GTTTCGAGTGAACCAGATCGCGCAGGAGGCCCGGCAGCACCGGCATAATCAGGCCGAT
GCCGACAGCGTCGAGCGCGACAGTGCTCAGAATTACGATCAGGGGTATGTTGGGTTTT
ACGTCTGGCCTCCGGACCAGCCTCCGCTGGTCCGATTGAACGCGCGGATTCTTTATCA
CTGATAAGTTGGTGGACATATTATGTTTATCAGTGATAAAGTGTAAGCATGACAAAGTT
GCAGCCGAATACAGTGATCCGTGCCGCCCTGGACCTGTTGAACGAGGTGCGCGTAGA
CGGTCTGACGACACGCAAACTGGCGGAACGGTTGGGGGTTTCAGCAGCCGGCGCTTTA
CTGGCACTTCAGGAACAAGCGGGCGCTGCTCGACGCACTGGCCGAAGCCATGCTGGC
GGAGAATCATACGCATTCGGTGCCGAGAGCCGACGACGACTGGCGCTCATTTCTGATC
GGGAATGCCCGCAGCTTCAGGCAGGCGCTGCTCGCCTACCGCGATGGCGCGCGCATC
CATGCCGGCACGCGACCGGGCGCACCGCAGATGGAAACGGCCGACGCGCAGCTTCG
CTTCCTCTGCGAGGCGGGTTTTTCGGCCGGGGACGCCGTCAATGCGCTGATGACAATC
AGCTACTTCACTGTTGGGGCCGTGCTTGAGGAGCAGGCCGGCGACAGCGATGCCGGC
GAGCGCGGCGGCACCGTTGAACAGGCTCCGCTCTCGCCGCTGTTGCGGGCCGCGATA
GACGCCCTTCGACGAAGCCGGTCCGGACGCGAGCGTTTCGAGCAGGGACTCGCGGTGATT
GTCGATGGATTGGCGAAAAGGAGGCTCGTTGTGAGGAACGTTGAAGGACCGAGAAAG
GGTGACGATTGATCAGGACCGCTGCCGGAGCGCAACCCACTCACTACAGCAGAGCCA
TGTAGACAACATCCCCCTCCCCCTTTCCACCGCGTCAGACGCCCGTAGCAGCCCGCTAC
GGGCTTTTTTCATGCCCTGCCCTAGCGTCCAAGCCTCACGGCCGCGCTCGGCCTCTCTG
GCGGCCCTTCTGGCGCTCTTCCGCTTCTCTGCTCACTGACTCGCTGCGCTCGGTGCTTC
GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCAAAAAGGCCAGGAACCG
TAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAC

FIG. 31 (Continued)

27/31

AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGCTTACCGG
ATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTTCCGCTGCATAACCCCTG
CTTCGGGGTCAATTATAGCGATTTTTTCGGTATATCCATCCTTTTTTCGCACGATATACAGG
ATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGC
AGGATAGGTGAAGTAGGCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATT
CGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCG
GCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCA
GCCACCTATCAAGGTGTACTGCCTTCAGACGAACGAAGAGCGATTGAGGAAAAGGC
GGCGGCGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTTCGGCCAGGGCTACA
AAATCACGGGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCG
ACCTGGGCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGG
CGCGGTTCCGGTATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGG
ACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGGCAGAGCCATGACTTT
TTAGCCGCTAAAACGGCCGGGGGGTGC GCGTGATTGCCAAGCACGTCCCCATGCGC
TCCATCAAGAAGAGCGACTTCGCGGAGCTGGTGAAGTACATCACCGACGAGCAAGGCA
AGACCGAGCGCCTTTGCGACGCTCACCGGGCTGGTTGCCCTCGCCGCTGGGCTGGCG
GCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGCGAGACAC
CGCGGCCGCCGGCGTTGTGGATACCTCGCGGAAACTTGGCCCTCACTGACAGATGA
GGGGCGGACGTTGACACTTGAGGGGCGGACTCACCCGGCGCGGGCGTTGACAGATGA
GGGGCAGGCTCGATTTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCG
AAAACGCCTGATTTTACGCGAGTTTCCACAGATGATGTGGACAAGCCTGGGGATAAGT
GCCCTGCGGTATTGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCGATCC
TTGACACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATTTGAGG
GGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGCCCGTTTTTTCGGCCA
CCGCTAACCTGTCTTTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAACAG
GGCTGCGCCCTGTGCGCGTGACCGCGCACGCGGAAGGGGGGTGCCCCCCCTTCTCG
AACCTCCCGGCCCGCTAACGCGGGCCTCCCATCCCCCAGGGGCTGCGCCCCCTCGG
CCGCGAACGGCCTCACCCCAAAAATGGCAGCGCTGGCAGTCCTTGCCATTGCCGGGA
TCGGGGCAGTAACGGGATGGGCGATCAGCCCGAGCGCGACGCCCGGAAGCATTGAC
GTGCCGACAGGTGCTGGCATCGACATTCAGCGACCAGGTGCCGGGCAGTGAGGGCGG
CGGCTGGGTGGCGGCCTGCCCTTCACTTCGGCCGTTCGGGGCATTACAGGACTTCAT
GGCGGGGGCCGGCAATTTTTACCTTGGGCATTCTTGGCATAGTGGTCCGCGGTGCCGT
GCTCGTGTTCCGGGGTGCGATAAACCCAGCGAACCATTTGAGGTGATAGGTAAGATTA
TACCGAGGTATGAAAACGAGAATTGGACCTTTACAGAATTACTCTATGAAGCGCCATAT
TAAAAAAGCTACCAAGACGAAGAGGATGAAGAGGATGAGGAGGCAGATTGCCTTGAAT
ATATTGACAATACTGATAAGATAATATATCTTTTATATAGAAGATATCGCCGTATGTAAGG
ATTTAGGGGGCAAGGCATAGGCAGCGCGCTTATCAATATATCTATAGAATGGGCAAA
GCATAAAAACCTTGATGGAATAATGCTTGAAACCCAGGACAATAACCTTATAGCTTGTA
ATTCTATCATAATTGGGTAATGACTCCAACCTTATTGATAGTGTTTTATGTTTCAGATAATGC
CCGATGACTTTGTATGCAGCTCCACCGATTTTGAGAACGACAGCGACTTCCGTCCCA
GCCGTGCCAGGTGCTGCCTCAGATTGAGGTTATGCCGCTCAATTTCGCTGCGTATATCG
CTTGCTGATTACGTGCAGCTTTCCCTTCAGGCGGGATTTCATACAGCGGCCAGCCATCC
GTCATCCATATCACCGAGTCAAAGGGTGACAGCAGGCTCATAAGACGCCCCAGCGTCG
CCATAGTGCGTTCACCGAATACGTGCGCAACAACCGTCTTCCGGAGACTGTCATACGC
GTAAAACAGCCAGCGCTGGCGCGATTTAGCCCCGACATAGCCCCACTGTTTCGTCCATT
TCCGCGCAGACGATGACGTCACTGCCCGGCTGTATGCGCGAGGTTACCGACTGCGGC
CTGAGTTTTTTAAGTGACGTAAAATCGTGTTGAGGCCAACGCCCATATGCGGGCTGTT
GCCCGGCATCCAACGCCATTGATGCCATATCAATGATTTTCTGGTGCGTACCGGGTT
GAGAAGCGGTGTAAGTGAAGTGCAGTTGCCATGTTTTACGGCAGTGAGAGCAGAGATA
GCGCTGATGTCCGGCGGTGCTTTTGCGGTTACGCACCAACCCCGTCAGTAGCTGAACAG
GAGGGACAGCTGATAGACACAGAAGCCACTGGAGCACCTCAAAAACACCATCATACAC
TAAATCAGTAAGTTGGCAGCATCACCCATAATTGTGGTTTTCAAAATCGGCTCCGTGAT
ACTATGTTATACGCCAATTTGAAAACAACCTTTGAAAAGCTGTTTTCTGGTATTTAAGG
TTTTAGAATGCAAGGAACAGTGAATTGGAGTTCGTCTTGTTATAATTAGCTTCTGGGGT

FIG. 31 (Continued)

28/31

ATCTTTAAATACTGTAGAAAAGAGGAAGGAAATAATAAATGGCTAAAATGAGAATATCAC
CGGAATTGAAAAAACTGATCGAAAAATACCGCTGCGTAAAAGATACGGAAGGAATGTCT
CCTGCTAAGGTATATAAGCTGGTGGGAGAAAATGAAAACCTATATTTAAAAATGACGGA
CAGCCGGTATAAAGGGACCACCTATGATGTGGAACGGGAAAAGGACATGATGCTATGG
CTGGAAGGAAAAGCTGCCTGTTCCAAAGGTCTCTGCACTTTGAACGGCATGATGGCTGGA
GCAATCTGCTCATGAGTGAGGCCGATGGCGTCTTTGCTCGGAAGAGTATGAAGATGA
ACAAAGCCCTGAAAAGATTATCGAGCTGTATGCGGAGTGCATCAGGCTCTTTCACTCCA
TCGACATATCGGATTGTCCCTATACGAATAGCTTAGACAGCCGCTTAGCCGAATTGGAT
TACTTACTGAATAACGATCTGGCCGATGTGGATTGCGAAAACCTGGGAAGAAGACACTCC
ATTTAAAGATCCGCGCGAGCTGTATGATTTTTTAAAGACGGAAAAGCCCGAAGAGGAAC
TTGTCTTTTCCACGGCGACCTGGGAGACAGCAACATCTTTGTGAAAAGATGGCAAAGTA
AGTGGCTTTATTGATCTTGGGAGAAGCGGCAGGGCGGACAAGTGGTATGACATTGCCT
TCTGCGTCCGGTCGATCAGGGAGGATATCGGGGAAGAACAGTATGTCCGAGCTATTTTT
TGACTTACTGGGGATCAAGCCTGATTGGGAGAAAATAAAAATATTATATTTTACTGGATGA
ATTGTTTTAGTACCTAGATGTGGCGCAACGATGCCGGCGACAAGCAGGAGCGCACCGA
CTTCTTCCGCATCAAGTGTTTTGGCTCTCAGGCCGAGGCCACGGCAAGTATTTGGGC
AAGGGGTCGCTGGTATTCTGTCAGGGCAAGATTGCGAATACCAAGTACGAGAAGGACG
GCCAGACGGTCTACGGGACCGACTTCATTGCCGATAAGGTGGATTATCTGGACACCAA
GGCACCAGGCGGGTCAAATCAGGAATAAGGGCACATTGCCCGGGCGTGAGTCGGGGC
AATCCCGCAAGGAGGGTGAATGAATCGGACGTTTGACCGGAAGGCATACAGGCAAGAA
CTGATCGACGCGGGGTTTTCCGCCGAGGATGCCGAAACCATCGCAAGCCGCACCGTC
ATGCGTGCGCCCCCGCAAACCTTCCAGTCCGTCGGCTCGATGGTCCAGCAAGCTACG
GCCAAGATCGAGCGCGACAGCGTGCAACTGGCTCCCCCTGCCCTGCCCGCGCCATCG
GCCGCCGTGGAGCGTTCGCGTCGTCTCGAACAGGAGGCGGCAGGTTTGGCGAAGTCG
ATGACCATCGACACGCGAGGAACTATGACGACCAAGAAGCGAAAAACCGCCGGCGAG
GACCTGGCAAAACAGGTCAGCGAGGCCAAGCAGGCCGCGTTGCTGAAACACACGAAG
CAGCAGATCAAGGAAATGCAGCTTTCCTTGTTTCGATATTGCGCCGTGGCCGGACACGA
TGCGAGCGATGCCAAACGACACGGCCCGCTCTGCCCTGTTACCCACGCGCAACAAGA
AAATCCCGCGCGAGGGCGCTGCAAAACAAGGTCAATTTCCACGTCAACAAGGACGTGAA
GATCACCTACACCGGCGTCGAGCTGCGGGCCGACGATGACGAACTGGTGTGGCAGCA
GGTGTGGAGTACGCGAAGCGCACCCCTATCGGGCAGCCGATCACCTTCACGTTCTAC
GAGCTTTGCCAGGACCTGGGCTGGTTCGATCAATGGCCGGTATTACACGAAGGCCGAG
GAATGCCTGTGCGGCCTACAGGCCGACGGCGATGGGCTTCACGTCCGACCGCGTTGGG
CACCTGGAATCGGTGTGCTGCTGCACCGCTTCCGCGTCTCTGGACCGTGGCAAGAAAA
CGTCCCGTTGCCAGGTCCTGATCGACGAGGAAATCGTCTGCTGTTTGGTGGCGACCA
CTACACGAAATTCATATGGGAGAAGTACCGCAAGCTGTCGCCGACGGCCCGACGGATG
TTCGACTATTTAGCTCGCACCGGGAGCCGTACCCGCTCAAGCTGGAAACCTTCCGCC
TCATGTGCGGATCGGATTCCACCCGCGTGAAGAAGTGGCGCGAGCAGGTGGCGAAG
CCTGCGAAGAGTTGCGAGGCAGCGGCCTGGTGGAACACGCCTGGGTCAATGATGACC
TGGTGCATTGCAAACGCTAGGGCCTTGTGGGGTCAGTTCCGGCTGGGGGTTCAGCAG
CCAGCGCTTTACTGGCATTTCAGGAACAAGCGGGCACTGCTCGACGCACTTGCTTCGC
TCAGTATCGCTCGGGACGCACGGCGCGCTCTACGAACTGCCGATAAACAGAGGATTAA
AATTGACAATTGTGATTAAGGCTCAGATTTCGACGGCTTGGAGCGGCCGACGTGCAGGA
TTTCCGCGAGATCCGATTGTGCGCCCTGAAGAAAGCTCCAGAGATGTTCCGGTCCGTT
TACGAGCACGAGGAGAAAAAGCCCATGGAGGCGTTCGCTGAACGGTTGCGAGATGCC
GTGGCATTGCGCGCCTACATCGACGGCGAGATCATTGGGCTGTGCGTCTTCAAACAGG
AGGACGGCCCCAAGGACGCTCACAAGGCGCATCTGTCCGGCGTTTTCTGGAGCCCG
AACAGCGAGGCCGAGGGGTGCGCCGGTATGCTGCTGCGGGCGTTGCCGGCGGGTTTAT
TGCTCGTGATGATCGTCCGACAGATTCCAACGGGAATCTGGTGGATGCGCATCTTCAT
CCTCGGCGCACTTAATATTTTCGCTATTCTGGAGCTTGTTGTTTATTTCCGGTCTACCGCCT
GCCGGGCGGGGTGCGGGCGACGGTAGGCGCTGTGCAGCCGCTGATGGTTCGTGTTCA
TCTCTGCCGCTCTGCTAGGTAGCCCGATACGATTGATGGCGGTCCTGGGGGCTATTTG
CGGAACTGCGGGCGTGGCGCTGTTGGTGTGACACCAAACGCAGCGCTAGATCCTGT
CGGCGTCGCAGCGGGCCTGGCGGGGGCGGTTTCCATGGCGTTCGGAACCGTGCTGA
CCCGCAAGTGGCAACCTCCCGTGCTCTGCTCACCTTTACCGCCTGGCAACTGGCGGC

FIG. 31 (Continued)

29/31

CGGAGGACTTCTGCTCGTTCCAGTAGCTTTAGTGTTTGATCCGCCAATCCCGATGCCTA
CAGGAACCAATGTTCTCGGCCTGGCGTGGCTCGGCCTGATCGGAGCGGGTTTAACCTA
CTTCCTTTGTTCCGGGGGATCTCGCGACTCGAACCTACAGTTGTTTCCTTACTGGGCT
TTCTCAGCCCCAGATCTGGGGTTCGATCAGCCGGGGATGCATCAGGCCGACAGTCGGA
ACTTCGGGTCCCCGACCTGTACCATTCCGGTGAGCAATGGATAGGGGAGTTGATATCGT
CAACGTTCACTTCTAAAGAAATAGCGCCACTCAGCTTCCTCAGCGGCTTTATCCAGCGA
TTTCCTATTATGTCGGCATAGTTCTCAAGATCGACAGCCTGTCACGGTTAAGCGAGAAA
TGAATAAGAAGGCTGATAATTCGGATCTCTGCGAGGGAGATGATATTTGATCACAGGCA
GCAACGCTCTGTCATCGTTACAATCAACATGCTACCCTCCGCGAGATCATCCGTGTTTC
AAACCCGGCAGCTTAGTTGCCGTTCTTCCGAATAGCATCGGTAACATGAGCAAAGTCTG
CCGCCTTACAACGGCTCTCCCGCTGACGCCGTCCCGGACTGATGGGCTGCCTGTATC
GAGTGGTGATTTTGTGCCGAGCTGCCGGTCCGGGAGCTGTTGGCTGGCTGGTGGCAG
GATATATTGTGGTGTAACAAATTGACGCTTAGACAACCTTAATAACACATTGCGGACGTT
TTTAATGTACTGGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCC
TTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGC
AGGCGAAAATCCTGTTTGATGGTGGTTCCGAAATCGGCAAAATCCCTTATAAATCAAAA
GAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTTAA
GAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTA
CGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCG
GAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGC
GAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCCATTCAGGCTGCGCAACTGTT
GGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGAT
GTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAA
AACGACGGCCAGTGAATTCGAGCTCGGTACCCCCC

FIG. 31 (Continued)

30/31

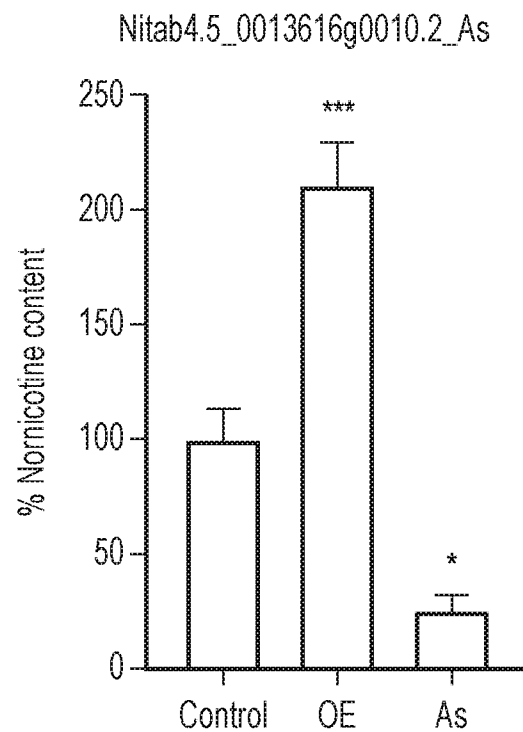


FIG. 32

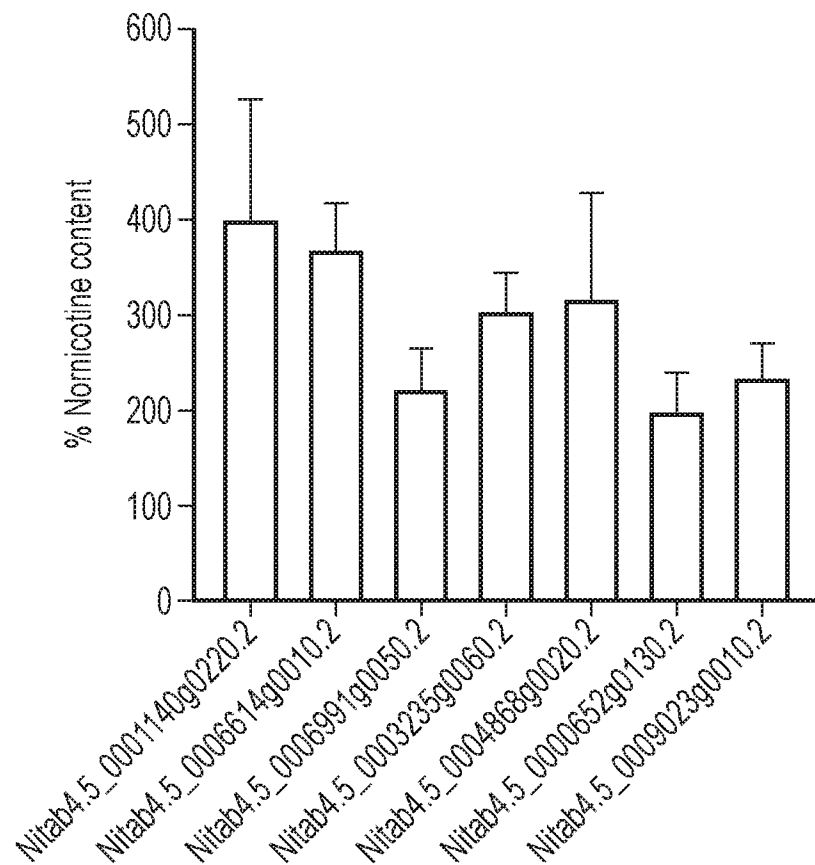


FIG. 33

31/31

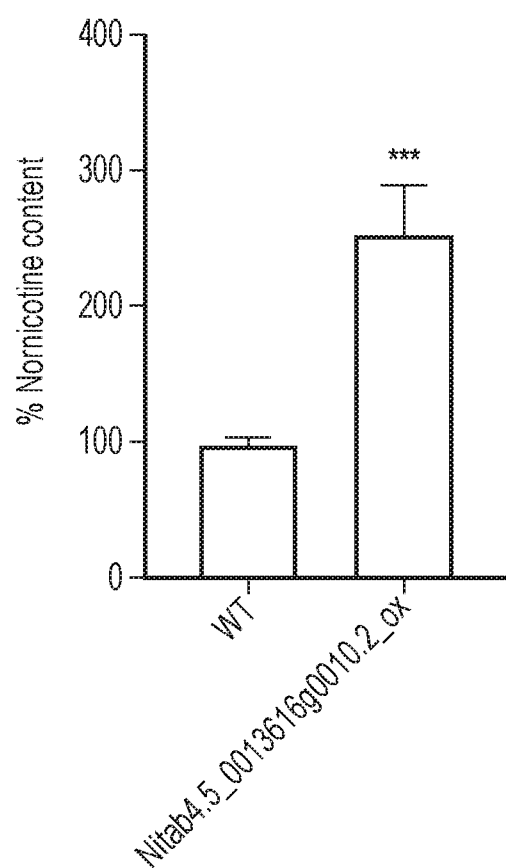


FIG. 34

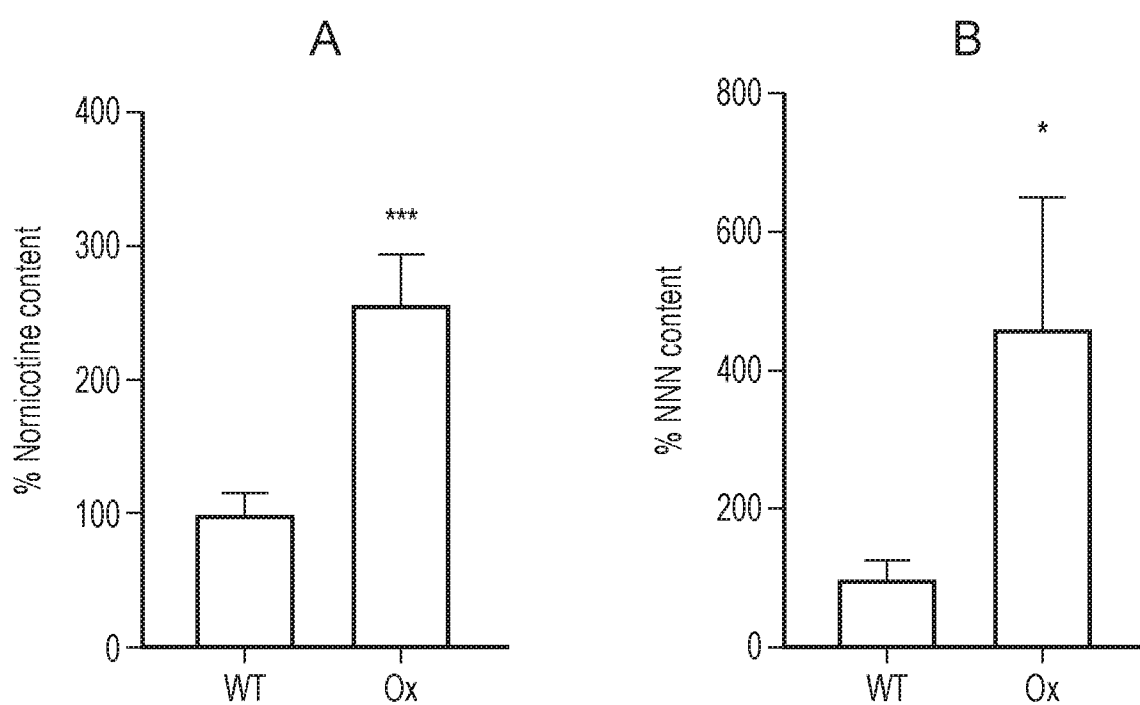


FIG. 35

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2020/050153

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/82 A01H5/12 A01H6/82
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N A01H

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

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

EP0-Internal, BIOSIS, CHEM ABS Data, EMBASE, WPI Data, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	W0 2015/054106 A1 (MONSANTO TECHNOLOGY LLC [US]) 16 April 2015 (2015-04-16)	1-41
Y	paragraph [0021] - paragraph [0032]; claims; table 20; sequence 32	1-41

X	W0 2018/116164 A1 (EVOGENE LTD [IL]) 28 June 2018 (2018-06-28)	1-41
Y	page 30, line 29 - page 31, line 21; claims; table 319	1-41

	-/--	

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

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

"E" earlier application or patent but published on or after the international filing date

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

22 April 2020

Date of mailing of the international search report

04/05/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Burkhardt, Peter

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2020/050153

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DATABASE UniProt [Online] 10 May 2017 (2017-05-10), "SubName: Full=heme-binding-like protein At3g10130, chloroplastic isoform X1, Reg_factor_effector_dom_sf, SOUL_haem-bd", XP055688238, retrieved from EBI accession no. UNIPROT:A0A1U7VHE2 Database accession no. A0A1U7VHE2 sequence	1-41
Y	----- DATABASE UniProt [Online] 12 April 2017 (2017-04-12), "SubName: Full=heme-binding-like protein At3g10130, chloroplastic, Reg_factor_effector_dom_sf, SOUL_haem-bd.", XP055688261, retrieved from EBI accession no. UNIPROT:A0A1S4AX09 Database accession no. Unreviewed sequence	1-41
A	----- W0 2018/067985 A1 (ALTRIA CLIENT SERVICES LLC [US]) 12 April 2018 (2018-04-12) the whole document	1-41
A	----- W0 2016/124932 A1 (BRITISH AMERICAN TOBACCO CO [GB]) 11 August 2016 (2016-08-11) the whole document -----	1-41

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2020/050153

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015054106 A1	16-04-2015	US 2016237447 A1	18-08-2016
		US 2019367935 A1	05-12-2019
		WO 2015054106 A1	16-04-2015

WO 2018116164 A1	28-06-2018	AU 2017381762 A1	01-08-2019
		BR 112019012622 A2	19-11-2019
		CA 3047074 A1	28-06-2018
		US 2019352661 A1	21-11-2019
		WO 2018116164 A1	28-06-2018

WO 2018067985 A1	12-04-2018	EP 3523439 A1	14-08-2019
		US 2018119163 A1	03-05-2018
		WO 2018067985 A1	12-04-2018

WO 2016124932 A1	11-08-2016	AR 103998 A1	21-06-2017
		BR 112017016815 A2	03-04-2018
		CN 107208109 A	26-09-2017
		EP 3253878 A1	13-12-2017
		JP 6368863 B2	01-08-2018
		JP 2018506967 A	15-03-2018
		US 2018037902 A1	08-02-2018
		WO 2016124932 A1	11-08-2016
