



- (51) **International Patent Classification:**
A01H 5/12 (2006.01) *C12N 15/82* (2006.01)
- (21) **International Application Number:**
PCT/GB2016/050323
- (22) **International Filing Date:**
11 February 2016 (11.02.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
1502409.4 13 February 2015 (13.02.2015) GB
- (71) **Applicant:** **BRITISH AMERICAN TOBACCO (INVESTMENTS) LIMITED** [GB/GB]; Globe House, 1 Water Street, London WC2R 3LA (GB).
- (72) **Inventors:** **GRONLUND, Jesper Tranholm**; c/o British American Tobacco (Investments) Limited, Globe House, 1 Water Street, London WC2R 3LA (GB). **SANCHEZ TAMBURRINO, Juan Pablo**; c/o British American Tobacco (Investments) Limited, Globe House, 1 Water Street, London WC2R 3LA (GB). **HUMPHRY, Matthew Edward**; c/o British American Tobacco (Investments) Limited, Globe House, 1 Water Street, London WC2R 3LA (GB).
- (74) **Agent:** **WILLIAMS, Aylsa**; 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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, 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).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** METHOD

(57) **Abstract:** The present invention relates to a method of reducing heavy metal levels in at least the aerial parts (preferably the leaves) of a plant comprising modifying the plant by increasing the expression of a *GSH1* gene or the activity of the protein encoded by the *GSH1* gene in said plant. The invention further relates to a modified plant obtained or obtainable by the method of the present invention having a reduced heavy metal level in at least the aerial parts (particularly the leaves) compared with the aerial parts (particularly the leaves) of an unmodified plant, which is due to the increased expression of a *GSH1* gene or the activity of the protein encoded by the *GSH1* gene in said modified plant; and also consumable plant products comprising the aerial parts of the plant (preferably the leaves) harvested from the plant obtained or obtainable by the present invention.



WO 2016/128750 A1

METHOD

FIELD OF THE INVENTION

The present invention relates to methods for reducing heavy metals and metalloids (hereafter referred to as 'heavy metals') in the aerial parts, particularly in the leaves, of a plant. The present invention further relates to a plant with reduced heavy metal levels in their aerial parts, and particularly in their leaves.

BACKGROUND

The enzyme glutamate-cysteine ligase (GSH1) catalyses the first and rate-limiting step of glutathione biosynthesis. The *gsh1* gene is encoded by a single copy gene in *Arabidopsis* (locus At4g23100). The GSH1 polypeptide also has been called γ -glutamylcysteine synthetase (γ -ECS), cadmium insensitive 2 (CAD2), phytoalexin deficient 2 (PAD2) and root meristemless 1 (RML1). The *Arabidopsis* GSH1 polypeptide has a transit peptide and is targeted to the plastid (e.g. chloroplast).

Plants can be grown in regions where there are high heavy metal contents in the environment, e.g. soil. Plants can take up these heavy metals from their environment. It is undesirable for plants intended for human consumption to have high heavy metal content.

SUMMARY OF THE INVENTION

According to a first aspect the present invention provides a method of reducing heavy metal levels in at least the aerial parts (preferably the leaves) of a plant comprising modifying the plant by increasing the expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said plant.

According to another aspect of the present invention there is provided a construct or vector comprising a *gsh1* gene operably linked with a root-specific promoter or root-preferred promoter.

The present invention yet further provides a modified plant obtained or obtainable by the method according to the present invention having a reduced heavy metal level in at least the aerial parts (particularly the leaves) compared with the aerial parts (particularly the leaves) of an unmodified plant, which is due to the increased expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said modified plant.

The present invention further provides the use of increased expression of a *gsh1* gene or increased activity of the protein encoded by the *gsh1* gene in a plant for reducing heavy metal levels in at least the aerial parts (preferably the leaves) of said plant.

In a yet further aspect, the present invention provides a plant cell (e.g. a tobacco plant cell):

- a. comprising an exogenous *gsh1* gene or an exogenous *gsh1* gene and an exogenous *gsh2* gene;
- b. comprising a construct or vector according to the present invention; and/or
- c. obtainable (e.g. obtained by) a method or use according to the present invention.

The present invention yet further provides a plant (e.g. a tobacco plant):

- a. comprising an exogenous *gsh1* gene or an exogenous *gsh1* gene and an exogenous *gsh2* gene;
- b. which has been modified to achieve a reduction in heavy metal levels in at least the aerial parts (preferably the leaves) of said plant compared with the aerial parts (particularly the leaves) of an unmodified plant, where in the modification is an increase in the expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said plant;
- c. obtained or obtainable by the method according to the present invention;
- d. comprising a construct or vector according to the present invention;
- e. comprising a plant cell according to the present invention.

In a yet further aspect, the present invention provides a plant propagation material (e.g. a plant seed) obtainable from a plant according to the present invention.

The present invention yet further provides the use of a plant cell (e.g. a tobacco cell) according to the present invention for the production of a consumable plant product (e.g. a tobacco product).

In a further aspect, the present invention provides the use of a plant (e.g. a tobacco plant) according to the present invention to breed a plant (e.g. a tobacco plant).

In another aspect, the present invention provides the use of plant (e.g. a tobacco plant) according to the present invention: a) for production of a consumable plant product (e.g. a tobacco product) b) to grow a crop; c) to produce a consumable leaf (e.g. a processed (preferably cured) leaf).

The present invention yet further provides a harvested leaf of a plant (e.g. a tobacco plant)

according to the present invention or obtainable from a plant (e.g. a tobacco plant) propagated from a propagation material according to the present invention or obtainable from a plant (e.g. a tobacco plant) obtainable by a use according to the present invention.

The present invention further provides a cut harvested leaf or cut processed leaf of a plant (e.g. a tobacco plant).

In a further aspect the present invention provides a processed leaf (e.g. a processed tobacco leaf, preferably a non-viable processed tobacco leaf):

- a. comprising a plant cell according to the present invention;
- b. obtainable from a plant obtainable from a use according to the present invention;
- c. obtainable from processing a plant (e.g. tobacco plant) according to the present invention;
- 5 d. obtainable from a plant (e.g. tobacco plant) propagated from a plant propagation material according to the present invention;
- e. obtainable by processing a harvested leaf according to the present invention.

A consumable plant product:

- 10 a. prepared from the aerial parts of the plant (preferably the leaves) harvested from the plant obtained or obtainable by the method according to the present invention;
- b. prepared from the aerial parts of the plant (preferably the leaves) harvested from the plant according to the present invention;
- c. prepared from the aerial parts of the plant (preferably the leaves) propagated from a plant propagation material according to the present invention;
- 15 d. prepared from a harvested leaf according to the present invention;
- e. prepared from a processed leaf according to the present invention;
- f. prepared from or comprising a plant extract obtained from a modified plant or a portion thereof according to the present invention.

20 The present invention yet further provides a plant extract obtained from a modified plant or a portion thereof in accordance with the present invention.

The present invention also provides a consumable plant product comprising the aerial parts of the plant (preferably the leaves) harvested from the plant obtained or obtainable by the method of the present invention, or a modified plant according to the present invention, or a plant extract obtained from a modified plant or a portion thereof according to the present
25 invention, having a reduced heavy metal level due to the increased expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said plant from which the leaves were harvested.

Yet further the present invention provides a smoking article comprising consumable plant (e.g. tobacco) comprising the aerial parts of the plant (preferably the leaves) harvested from
30 the plant (e.g. tobacco) obtained or obtainable by the method of the present invention, or a modified plant (e.g. a modified tobacco plant) according to the present invention, or a plant extract obtained from a modified plant or a portion thereof according to the present invention, having a reduced heavy metal level due to the increased expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said plant (e.g. tobacco plant) from which
35 the aerial parts of the plant (e.g. the leaves) were harvested.

BRIEF DESCRIPTION OF THE DRAWINGS

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

Figure 1 shows a schematic representation depicting the pathway of glutathione biosynthesis from constituent amino acids;

Figure 2 shows cadmium levels detected in leaves of tobacco plants carrying either the empty vector control or an *A622-GSH1* construct. The presence of the *A622-GSH1* construct results in up to 61% wt/wt reduction of cadmium concentrations in the leaves, compared to leaves of plants carrying an empty vector control construct;

Figure 3 shows a genomic DNA for *Arabidopsis thaliana gsh1* (At4g23100.1)_(SEQ ID No. 1) (introns are shown in lower case and *italics*);

Figure 4 shows a coding sequence for *Arabidopsis thaliana gsh1* (At4g23100.1)_(SEQ ID No. 2);

Figure 5 shows a polypeptide sequence for the protein of *Arabidopsis thaliana GSH1* (At4g23100.1)_(SEQ ID No. 3) translated from the cDNA (SEQ ID No. 2);

Figure 6 shows a genomic DNA for *Arabidopsis thaliana gsh2* (At5g27380.1)_(SEQ ID No. 4) (introns are shown in lower case and *italics*);

Figure 7 shows a coding sequence for *Arabidopsis thaliana gsh2* (At5g27380.1)_(SEQ ID No. 5);

Figure 8 shows a polypeptide sequence for the protein of *Arabidopsis thaliana GSH2* (At5g27380.1)_(SEQ ID No. 6) translated from the cDNA (SEQ ID No. 5);

Figure 9 shows a polynucleotide sequence of the root-specific A622 promoter (SEQ ID No. 7);

Figure 10 shows a coding sequence for *Arabidopsis lyrata* subsp. *lyrata gsh1* (SEQ ID No. 8);

Figure 11 shows a polypeptide sequence (SEQ ID No. 9) for the protein of *Arabidopsis lyrata* subsp. *lyrata GSH1* (SEQ ID No. 8) (UniProt ID: D7MDW8);

Figure 12 shows a coding sequence for *Arabidopsis thaliana gsh1* (SEQ ID No. 10);

Figure 13 shows a polypeptide sequence (SEQ ID No. 11) for the protein of *Arabidopsis thaliana GSH1* (SEQ ID No. 10) (UniProt ID: F4JMS5);

Figure 14 shows a coding sequence for *Citrus clementina* (CICLE_v10019696mg) *gsh1* (SEQ ID No. 12);

Figure 15 shows a polypeptide sequence (SEQ ID No. 13) for the protein of *Citrus clementina* *GSH1* (SEQ ID No. 12) (UniProt ID: V4TSB7);

Figure 16 shows a coding sequence for *Glycine max* (Glyma) *gsh1* (SEQ ID No. 14);

Figure 17 shows a polypeptide sequence (SEQ ID No. 15) for the protein of *Glycine max* (Glyma) *GSH1* (SEQ ID No. 14) (UniProt ID: I1K6J1);

Figure 18 shows a coding sequence for *Medicago truncatula* *gsh1* (SEQ ID No. 16);

Figure 19 shows a polypeptide sequence (SEQ ID No. 17) for the protein of *Medicago truncatula* *GSH1* (SEQ ID No. 16) (UniProt ID: G7LHC5);

Figure 20 shows a coding sequence for *Solanum lycopersicum* *gsh1* (SEQ ID No. 18);

Figure 21 shows a polypeptide sequence (SEQ ID No. 19) for the protein of *Solanum lycopersicum* *GSH1* (SEQ ID No. 18) (UniProt ID: O22493);

Figure 22 shows a coding sequence for *Solanum tuberosum* *gsh1* (SEQ ID No. 20);

Figure 23 shows a polypeptide sequence (SEQ ID No. 21) for the protein of *Solanum tuberosum* *GSH1* (SEQ ID No. 20) (UniProt ID: M1AWB9);

Figure 24 shows a coding sequence for *Brachypodium distachyon* (*Bradi*) *gsh1* (SEQ ID No. 22);

Figure 25 shows a polypeptide sequence (SEQ ID No. 23) for the protein of *Brachypodium distachyon* (*Bradi*) *GSH1* (SEQ ID No. 22) (UniProt ID: I1HMP3);

Figure 26 shows a coding sequence for *Oryza sativa* *gsh1* (SEQ ID No. 24);

Figure 27 shows a polypeptide sequence (SEQ ID No. 25) for the protein of *Oryza sativa* *GSH1* (SEQ ID No. 24) (UniProt ID: B7EUG4);

Figure 28 shows a coding sequence for *Zea mays* (*Zeama*) *gsh1* (SEQ ID No. 26);

Figure 29 shows a polypeptide sequence (SEQ ID No. 27) for the protein of *Zea mays* *GSH1* (SEQ ID No. 26) (UniProt ID: Q8W4W3);

Figure 30 shows a coding sequence for *Physcomitrella patens* *subsp. patens* *gsh1* (SEQ ID No. 28);

Figure 31 shows a polypeptide sequence (SEQ ID No. 29) for the protein of *Physcomitrella patens* *subsp. patens* *GSH1* (SEQ ID No. 28) (UniProt ID: A9RU07);

Figure 32 shows a coding sequence for *Chlamydomonas reinhardtii* *gsh1* (SEQ ID No. 30);

Figure 33 shows a polypeptide sequence (SEQ ID No. 31) for the protein of *Chlamydomonas reinhardtii* *GSH1* (SEQ ID No. 30) (UniProt ID: A8IA77);

Figure 34 shows a coding sequence for *Oestrecoccus lucimarinus* *gsh1* (SEQ ID No. 32);

Figure 35 shows a polypeptide sequence (SEQ ID No. 33) for the protein of *Oestrecoccus lucimarinus* *GSH1* (SEQ ID No. 32) (UniProt ID: A4RZD5);

Figure 36 shows a coding sequence for *Arabidopsis lyrata gsh2* (SEQ ID No. 34);

Figure 37 shows a polypeptide sequence (SEQ ID No. 35) for the protein of *Arabidopsis lyrata GSH2* (SEQ ID No. 34) (UniProt ID: D7M652);

Figure 38 shows a coding sequence for *Citrus_clementina* (CICLE) *gsh2* (SEQ ID No. 36);

5 **Figure 39** shows a polypeptide sequence (SEQ ID No. 37) for the protein of *Citrus_clementina GSH2* (SEQ ID No. 36) (UniProt ID: V4SFK5);

Figure 40 shows a coding sequence for *Medicago truncatula gsh2* (SEQ ID No. 38);

Figure 41 shows a polypeptide sequence (SEQ ID No. 39) for the protein of *Medicago truncatula GSH2* (SEQ ID No. 38) (UniProt ID: A2Q600);

10 **Figure 42** shows a coding sequence for *Solanum lycopersicum gsh2* (SEQ ID No. 40);

Figure 43 shows a polypeptide sequence (SEQ ID No. 41) for the protein of *Solanum lycopersicum GSH2* (SEQ ID No. 40) (UniProt ID: O22494);

Figure 44 shows a coding sequence for *Brachypodium distachyon* (*Bradi*) *gsh2* (SEQ ID No. 42);

15 **Figure 45** shows a polypeptide sequence (SEQ ID No. 43) for the protein of *Brachypodium distachyon* (*Bradi*) *GSH2* (SEQ ID No. 42) (UniProt ID: I1IJU8);

Figure 46 shows a coding sequence for *Oryza sativa gsh2* (SEQ ID No. 44);

Figure 47 shows a polypeptide sequence (SEQ ID No. 45) for the protein of *Oryza sativa GSH2* (SEQ ID No. 44) (UniProt ID: Q0IP23);

20 **Figure 48** shows a coding sequence for *Zea mays* (*Zeama*) *gsh2* (SEQ ID No. 46);

Figure 49 shows a polypeptide sequence (SEQ ID No. 47) for the protein of *Zea mays GSH2* (SEQ ID No. 46) (UniProt ID: B4FWG0);

Figure 50 shows a coding sequence for *Physcomitrella patens subsp. patens gsh2* (SEQ ID No. 48);

25 **Figure 51** shows a polypeptide sequence (SEQ ID No. 49) for the protein of *Physcomitrella patens subsp. patens GSH2* (SEQ ID No. 48) (UniProt ID: A9RTT8);

Figure 52 shows a coding sequence for *Chlamydomonas reinhardtii gsh2* (SEQ ID No. 50);

Figure 53 shows a polypeptide sequence (SEQ ID No. 51) for the protein of *Chlamydomonas reinhardtii GSH2* (SEQ ID No. 50) (UniProt ID: A8IR74);

30 **Figure 54** shows a coding sequence for *Oestreococcus lucimarinus gsh2* (SEQ ID No. 52);

Figure 55 shows a polypeptide sequence (SEQ ID No. 53) for the protein of *Oestreococcus lucimarinus GSH2* (SEQ ID No. 52) (UniProt ID: A4RYM3).

Figure 56 shows an alignment of some GSH1 protein sequences and shows the conserved residues of two metal-binding sites (see underlined residues in the consensus sequence),

35 namely n1 metal binding site: Glu-115, Glu-167, Glu-173; and n2 metal binding site Glu-113,

Gln-254 and Glu-393 (wherein the numbering is that of the *Arabidopsis thaliana* GSH1 ("Arath_gsha (1)") sequence shown herein as SEQ ID No. 3).

DETAILED DESCRIPTION

5 A seminal finding of the present invention is that by increasing the expression of at least glutamate-cysteine ligase (*GSH1*) in a plant (e.g. a tobacco plant) – and particularly in plant roots, e.g. tobacco roots – heavy metal transport to the aerial parts of the plant, e.g. in the leaves, the seeds, the fruits and/or the stems, and particularly in the leaves of the plant can be reduced. Thus, the level of heavy metals in the arial parts of the plant, e.g. in the leaves,
10 the seeds, the fruits and/or the stems, and particularly in the plant leaves (particularly in the harvested plant leaves) can be significantly reduced.

The present invention provides a method of reducing heavy metal levels in at least the aerial parts (preferably the leaves) of a plant comprising modifying the plant by increasing the expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said
15 plant.

In one embodiment the increased expression of the *gsh1* gene is achieved by overexpressing or upregulating the *gsh1* gene.

In a preferred embodiment the method of the present invention comprises modifying the plant by increasing the expression of the *gsh1* gene in said plant root (e.g. in a plant root cell or
20 plant root tissue). Preferably the increased expression is predominantly in the plant root (e.g. in a plant root cell or plant root tissue). Preferably the increased expression is not constitutive. Preferably the increased expression is not in the aerial parts of the plant, e.g. is not in the leaves of the plant.

In one embodiment the method of the present invention comprises expressing within a plant
25 a polynucleotide (e.g. an exogenous polynucleotide) comprising a nucleic acid sequence encoding a GSH1 polypeptide.

In a further embodiment the polynucleotide (e.g. exogenous polynucleotide) for use in the present invention comprises a nucleic acid sequence encoding a GSH1 polypeptide operably linked with a heterologous promoter for directing transcription of said nucleic acid sequence
30 in said plant. In one embodiment preferably the promoter is non-constitutive (e.g. in the sense that it does not direct expression in all or virtually all tissues), but has at least expression in the root (e.g. in a root cell or root tissue). Suitably the promoter (e.g. heterologous promoter) for use in the present invention is a root-specific promoter or root-preferred promoter.

35 The term "constitutive" in relation to the promoter as used herein means an unregulated promoter that allows for continual transcription of its associated gene. The term "constitutive

promoters” as used herein means promoters that direct expression in virtually all tissues and are largely, if not entirely, independent of environmental and developmental factors.

Examples of constitutive promoters include CaMV35S, CaMV19S, CERV, TMV, UBQ5, UBQ10, ACT2. In one embodiment preferably the promoter for use in the present invention is not a constitutive promoter. In one embodiment preferably the promoter for use in the present invention is not one of the following constitutive promoters: CaMV35S, CaMV19S, CERV, TMV, UBQ5, UBQ10, ACT2.

The method and uses of the present invention comprise increasing the expression or the overexpression of at least GSH1. The overexpression can be achieved by any means known to the person skilled in the art.

The term “overexpression” or “increased expression” as used herein means any form of expression that is additional to the original wild-type expression level.

“Overexpression” means that a plant is increased in the mRNA level or the protein level in comparison with an expression level of a GSH1 of a parent plant of the same breed. The expression level of GSH1 is compared with that of GSH1 at a corresponding part in the parent plant of the same breed cultured under the same condition. A case where the expression level increases at least 1.1 times greater than that of the parent plant is preferably considered as a case where the expression level is increased. Here, it is more preferable that the expression level of the plant has a significant difference of 5% by a t-test compared with that of the parent plant, in order to be considered that there is an increase in the expression level. It is preferable that the expression levels of the plant and the parent plant be measured at the same time by the same method. However, data stored as background data may be also used.

Methods for increasing expression of genes or gene products are well documented in the art and include, for example, overexpression driven by appropriate promoters, the use of transcription enhancers or translation enhancers. Isolated nucleic acids which serve as promoter or enhancer elements may be introduced in an appropriate position (typically upstream) of a non-heterologous form of a polynucleotide so as to upregulate expression of a nucleic acid encoding the polypeptide of interest. For example, endogenous promoters may be altered in vivo by mutation, deletion, and/or substitution (see, US 5,565,350; WO9322443), or isolated promoters may be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3'

end sequence to be added may be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

An intron sequence may also be added to the 5' untranslated region (UTR) or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold (Buchman and Berg (1988) Mol. Cell biol. 8: 4395-4405; Callis et al. (1987) Genes Dev 1 :1183-1200). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of the maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. For general information see: The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, N.Y. (1994).

In one embodiment the overexpression may be achieved by the use of gene-editing or targeted mutagenesis.

Gene-editing may be carried out using any method known in the art. A few non-limiting examples are presented here including use of the CRISPR-Cas9 system. CRISPR/Cas9 genomic editing tools are available commercially such as "Guide-it" from Clontech (Avenue du President Kennedy 78100 Saint-Germain-en-Laye, France). Another method of gene-editing includes the use of TALEN (transcription activator-like effector nuclease) technology with kits available commercially (e.g. from Addgene, 1Kendall Sq. Ste. B7102, Cambridge, MA 02139, USA). A further method comprises the use of Zinc Finger Nucleases such as the CompoZr[®] Zinc Finger Nuclease Technology available from Sigma-Aldrich. Another method comprises the use of meganucleases (or a further method) described in Silva *et al* Curr Gene Ther. Feb 2011; 11(1): 11-27 (the teaching of which is incorporated herein by reference). A yet further method is oligonucleotide-directed mutagenesis (ODM) such as KeyBase[®] available from Keygene (Agro Business Park 90, 6708 PW Wageningen, The Netherlands). Suitably, gene-editing may be used to alter the sequence of a *gsh1* or *gsh2* *in vivo*.

In one embodiment, the skilled person may induce mutations, deletions or insertions either within the endogenous *gsh1* gene (and/or *gsh2* gene) itself or in its regulatory sequences (e.g. the promoter and/or the 3'UTR) in order to more highly express the endogenous gene and/or to render the protein produced by the gene more active. These mutations, deletions or insertions may be introduced by any known gene editing method, e.g. EMS mutation, gamma irradiation, CRISPR/Cas 9 technology, TALENs and/or meganuclease technology.

The present invention yet further provides a construct or vector comprising a *gsh1* gene or a nucleic acid sequence encoding a GSH1 polypeptide operably linked with a root-specific promoter or root-preferred promoter.

5 In one embodiment the construct or vector of the present invention further comprises (e.g. in addition to a *gsh1* gene or a nucleic acid sequence encoding a GSH1 polypeptide) a *gsh2* gene or a nucleic acid sequence encoding a GSH2 polypeptide operably linked with the same or a different root-specific promoter or root-preferred promoter.

10 In a further embodiment the present invention provides a modified plant obtained or obtainable by a method of the present invention wherein the modified plant has a reduced heavy metal level in at least its aerial parts (particularly the leaves) compared with the aerial parts (particularly the leaves) of an unmodified plant, which is due to the increased expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said modified plant.

15 The present invention yet further provides consumable plant product comprising the aerial parts of the plant (preferably the leaves) harvested from the plant obtained or obtainable by the method of the present invention or a modified plant according to the present invention having a reduced heavy metal level due to the increased expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said plant from which the leaves were harvested.

20 In the present invention, the heavy metal may be one selected from the group consisting of: cadmium, arsenic, chromium, copper, lead, zinc, mercury, nickel, selenium and a combination thereof.

In one embodiment the heavy metal may be selected from the group consisting of cadmium, arsenic, chromium or a combination thereof.

25 In a yet further embodiment the heavy metal may be cadmium.

In a yet further embodiment the heavy metal may be arsenic.

If a yet further embodiment the heavy metal may be chromium.

30 The present invention yet further provides a smoking article comprising said tobacco prepared in according with the present invention or a modified tobacco in accordance with the present invention.

Advantageously, and preferably, the modified tobacco plant may have been transformed with a genetic construct or vector of the invention.

The term "smoking article" as used herein may be a cigarette, cigar, cigarillo, or rolling tobacco, or the like.

The enzyme glutamate-cysteine ligase (GSH1; E.C. 6.3.2.2), which catalyses the first and rate-limiting step of glutathione biosynthesis, is also known as γ -glutamylcysteine synthetase (γ -ECS), cadmium insensitive 2 (CAD2), phytoalexin deficient 2 (PAD2) and root meristemless 1 (RML1) (see Figure 1).

- 5 A polypeptide with "GSH 1 activity" is a polypeptide with glutamate-cysteine ligase activity or γ -glutamylcysteine synthetase activity (E.C. 6.3.2.2). Enzymatic assays are available for determining GSH1 activity (Noctor and Foyer, 1998, Anal. Biochem. 264:98-110; Noctor et al., 2002, Exp. Bot. 53:1283-1304; Hothorn et al., 2006, J. Biol. Chem. 281: 27557-27565). In one embodiment the *GSH1* enzyme according to the present invention is one classified as
10 E.C. 6.3.2.2.

As shown in Figure 1 glutathione synthetase (GSH2; E.C. 6.3.2.3) catalyses the second step of glutathione biosynthesis.

In one embodiment the *GSH2* enzyme according to the present invention is one classified as E.C. 6.3.2.3.

- 15 The term "aerial parts" of the plant as used herein means the part(s) of the plant above the ground and may include the leaves, the fruit, the seed and the stem. In one embodiment the term aerial parts of the plant as used herein means leaf only or leaf and/or stem of the plant. Without wishing to be bound by theory by overexpression or upregulating the enzymes involved in the biosynthesis of glutathione, e.g. GSH1 and/or GSH2, heavy metals may be
20 sequestered in the plant roots (e.g. the vacuole of the plant roots) and thus be removed from further transport in the plant, e.g. to the plants aerial parts, such as into the leaves, the seeds, the fruits and/or the stems^{^^^}, and particularly into the leaves.

This has significant advantages in the production of tobacco and other crops where the leaves are consumed, e.g. in salad crops, such as lettuces (including for example baby leaf
25 spinach, Cos, Frisée, Iceberg lettuce, Lamb's lettuce, Little Gem, Mizuna, Radicchio, Red mustard, Red oak leaf, Rocket (Roquette), Ruby chard, Sweet Romaine), watercress and cress; as well as in leaf vegetables, such as cabbages, spinach, kale, tea, chicory (or curly endive), Phak chet, and herbs (such as basil, mint and oregano).

In addition, this has significant advantages in the production of crops where fruits and/or
30 seeds (e.g. aerial fruits and/or seeds) are consumed.

Suitably, the terms "fruits" and "seeds" may include legumes, grain, and nuts.

The plant (or part thereof) or plant cell or plant propagation material according to the present invention may be a monocotyledonous (monocot) plant or a dicotyledonous (dicot) plant.

In one embodiment the plant, plant cell or plant tissue according to the present invention may
35 be a dicot.

The plant (or part thereof) or plant cell or plant propagation material according to the present invention may be a crop plant, e.g. a fruit crop, a seed crop, a legume or a nut crop.

The crop plant in accordance with the present invention may be selected from the group consisting of tomato, strawberry, cherry, redcurrant, blackcurrant, gooseberry, raspberry,

5 mulberry, peppers (*Capsicum*), peppers (*Piper*), water melon, melon, squash, gourd or aubergine (eggplant), olive, radish, horseradish, banana, apple, pears, peach, grape vine, citrus species, wheat, oat, barley, triticale, rice, quinoa (*Chenopodium quinoa*), fonio (*Digitaria*) maize, sorghum, rye, onion, leek, millet, buckwheat, sugarcane, sunflower, oilseed rapeseed (including canola), okra, coffee, and cocoa (*Theobroma cacao*), palm, cotton,
10 coconut, sesame, safflower, flax, kapok, mustard, nutmeg, jojoba, peas, beans, alfalfa, lentils, soybeans, peanuts, almonds, pecans, pistachios, walnuts, Brazil nuts, hazelnuts, macadamia nuts, cashew nut, acorn, beechnuts, filbert nuts and chestnuts.

The plant (or part thereof) or plant cell or plant propagation material according to the present invention may be a fruit crop. A fruit crop in accordance with the present invention may be

15 selected from the group consisting of: tomato, strawberry, cherry, redcurrant, blackcurrant, gooseberry, raspberry, mulberry, peppers (*Capsicum*), peppers (*Piper*), water melon, melon, squash, gourd, aubergine (eggplant), olive, radish, horseradish, banana, apple, pears, peach, grape vine and citrus species.

The plant (or part thereof) or plant cell or plant propagation material according to the present

20 invention may be a seed crop. A seed crop in accordance with the present invention may be a cereal or grain crop, e.g. one selected from the group consisting of: wheat, oat, barley, triticale, rice, quinoa (*Chenopodium quinoa*), fonio (*Digitaria*), maize, sorghum, rye, onion, leek, millet, buckwheat, sugarcane. In another embodiment the seed crop in accordance with the present invention may be selected from the group consisting of: a sunflower, oilseed
25 rapeseed (including canola), okra, coffee, cocoa (*Theobroma cacao*), palm, cotton, coconut, sesame, safflower, flax, kapok, mustard, nutmeg and jojoba.

The plant (or part thereof) or plant cell or plant propagation material according to the present invention may be a legume. A legume in accordance with the present invention may be selected from the group consisting of peas, beans, alfalfa, lentils, soybeans and peanuts.

30 The plant (or part thereof) or plant cell or plant propagation material according to the present invention may be a fruit or seed crop which is a nut crop. A nut crop in accordance with the present invention may be selected from the group consisting of almonds, pecans, pistachios, walnuts, Brazil nuts, hazelnuts, macadamia nuts, cashew nut, acorn, beechnuts, filbert nuts and chestnuts.

Preferably, the plant, plant cell or plant tissue, or host plant is a crop plant. By crop plant is meant any plant which is grown on a commercial scale for human or animal consumption or use for animal feed.

In such plants heavy metals can be sequestered into the roots of the plant (and thus appear in the harvested or consumable arial parts (e.g. leaves) of the plants.

Companies wishing to reduce heavy metal levels in plant leaves – which can be consumed, e.g. used in smoking articles, such as cigarettes and cigars, can use the present invention.

The overexpression of at least *GSH1* (and/or *GSH2*), e.g. in the roots of the plant, is an elegant solution to this problem as the transport of heavy metals to the arial parts (e.g.

leaves) of the plant can be interrupted and hence the heavy metal levels in the arial parts (e.g. leaves) of the plant can be reduced. Heavy metal ions bind to glutathione and can be bound in the roots of the plant.

In one embodiment the *gsh1* gene is the Arabidopsis gene, e.g. locus At4g23100, or an orthologue thereof. In one embodiment, the *gsh1* gene for use in the present invention is encoded by SEQ ID No. 1 or SEQ ID No. 2 taught herein.

The *gsh1* gene for use in the present invention may be from any source, e.g. from a plant (e.g. a dicot plant or a monocot plant); from a moss; from an algae, from a bacteria (including cyanobacteria); or from an animal.

In some embodiments, the *gsh1* gene for use in the present invention may be encoded by one or more of the following sequences identified by their Uniprot sequence identifier:

D7MDW8, F4JMS5, P46309, R0F4V5, V4TMB4, V4TSB7, V4VJD2, V4MPS5, I1K6J1, G7LHC5, Q9ZNX6, B9GKI0, B9GWI2, B9RCB6, O22493, M1AWB9, D7TVZ7, I1HMP3, B7EUG4, Q688Q9, K3Z5M4, C5YZF7, Q8W4W3, A9RU07, A9U3U7, D8R0Y2, D8S150, A8IA77, D8TT59, C1N1W0, C1FFN8, A4RZD5, E1ZNL3, I0YUN5.

In one embodiment the *gsh1* gene for use in the present invention is encoded by a nucleotide sequence taught in any of the following as: SEQ ID No. 1 or SEQ ID No. 2 or SEQ ID No. 8 or SEQ ID No. 10 or SEQ ID No. 12 or SEQ ID No. 14 or SEQ ID No. 16 or SEQ ID No. 18 or SEQ ID No. 20 or SEQ ID No. 22 or SEQ ID No. 24 or SEQ ID No. 26 or SEQ ID No. 28 or SEQ ID No. 30 or SEQ ID No. 32.

The term “orthologue” as used herein means an homologous gene sequence found in a different organism, e.g. a different species. The term orthologue of the Arabidopsis *gsh1* gene may include one of the following genes from another organism (e.g. another species): SEQ ID No. 8 or SEQ ID No. 10 or SEQ ID No. 12 or SEQ ID No. 14 or SEQ ID No. 16 or SEQ ID No. 18 or SEQ ID No. 20 or SEQ ID No. 22 or SEQ ID No. 24 or SEQ ID No. 26 or SEQ ID No. 28 or SEQ ID No. 30 or SEQ ID No. 32.

In one embodiment the *gsh1* gene according to the present invention encodes a GSH1 protein sequence which has the following conserved residues: an n1 metal binding site comprised of Glu-115, Glu-167 and Glu-173; an n2 metal binding site comprised of Glu-113, Gln-254 and Glu-393; or a combination thereof. The numbering of the conserved residues means the effective equivalent position when the GSH1 protein is aligned with the *Arabidopsis thaliana* GSH1 ("Arath_gsha (1)") protein sequence shown herein as SEQ ID No. 3.

Without wishing to be bound by theory these metal binding sites are form the catalytic centre of the active site, where n1 site may bind free metal and the n2 site may interact with MgATP. Active site analysis of GSH1 enzymes was studied in Abbott *et al.* J. of Biol. Chem. Vol 276, No. 45, 9, pp42099-42107 (2001) which is hereby incorporated by reference.

In a further embodiment the *gsh1* gene or the nucleic acid sequence encoding said GSH1 polypeptide for use in the present invention comprises:

- i) a polynucleotide sequence shown herein as SEQ ID No. 1 or SEQ ID No. 2 or SEQ ID No. 8 or SEQ ID No. 10 or SEQ ID No. 12 or SEQ ID No. 14 or SEQ ID No. 16 or SEQ ID No. 18 or SEQ ID No. 20 or SEQ ID No. 22 or SEQ ID No. 24 or SEQ ID No. 26 or SEQ ID No. 28 or SEQ ID No. 30 or SEQ ID No. 32, or
- ii) a functional fragment of the polynucleotide sequence shown in i) which functional fragment encodes a functional GSH1 polypeptide, or
- iii) a polynucleotide which encodes a polypeptide comprising the amino acid sequence shown herein as SEQ ID No. 3 or SEQ ID No. 9 or SEQ ID No. 11 or SEQ ID No. 13 or SEQ ID No. 15 or SEQ ID No. 17 or SEQ ID No. 19 or SEQ ID No. 21 or SEQ ID No. 23 or SEQ ID No. 25 or SEQ ID No. 27 or SEQ ID No. 29 or SEQ ID No. 31 or SEQ ID No. 33, or
- iv) a polynucleotide sequence which can hybridize to the polynucleotide taught in i), ii) or iii) above under high stringency conditions, or
- v) a polynucleotide sequence which has at least 70% (preferably 85%, more preferably 90%) identity with the polynucleotide shown in i), ii) or iii) above, or
- vi) a polynucleotide sequence which differs from polynucleotide shown in i), ii) or iii) due to degeneracy of the genetic code.

In a further embodiment the *gsh1* gene or the nucleic acid sequence encoding said GSH1 polypeptide for use in the present invention comprises:

15

- i) a polynucleotide sequence shown herein as SEQ ID No. 1 or SEQ ID No. 2, or
- ii) a functional fragment of the polynucleotide sequence shown in i) which functional fragment encodes a functional GSH1 polypeptide, or
- 5 iii) a polynucleotide which encodes a polypeptide comprising the amino acid sequence shown herein as SEQ ID No. 3, or
- iv) a polynucleotide sequence which can hybridize to the polynucleotide taught in i), ii) or iii) above under high stringency conditions, or
- v) a polynucleotide sequence which has at least 70% (preferably 85%,
10 more preferably 90%) identity with the polynucleotide shown in i), ii) or iii) above, or
- vi) a polynucleotide sequence which differs from polynucleotide shown in i), ii) or iii) due to degeneracy of the genetic code.

In one embodiment of the present invention the protein encoded by the *gsh1* gene or the
15 nucleic acid encodes a polypeptide sequence shown herein as SEQ ID No. 3 or SEQ ID No. 9 or SEQ ID No. 11 or SEQ ID No. 13 or SEQ ID No. 15 or SEQ ID No. 17 or SEQ ID No. 19 or SEQ ID No. 21 or SEQ ID No. 23 or SEQ ID No. 25 or SEQ ID No. 27 or SEQ ID No. 29 or SEQ ID No. 31 or SEQ ID No. 33, or polypeptide sequence which comprises SEQ ID No. 3 or SEQ ID No. 9 or SEQ ID No. 11 or SEQ ID No. 13 or SEQ ID No. 15 or SEQ ID No. 17 or
20 SEQ ID No. 19 or SEQ ID No. 21 or SEQ ID No. 23 or SEQ ID No. 25 or SEQ ID No. 27 or SEQ ID No. 29 or SEQ ID No. 31 or SEQ ID No. 33 with a conservative substitution of at least one of the amino acids, or a polypeptide having at least 70% identity with SEQ ID No. 3 or SEQ ID No. 9 or SEQ ID No. 11 or SEQ ID No. 13 or SEQ ID No. 15 or SEQ ID No. 17 or SEQ ID No. 19 or SEQ ID No. 21 or SEQ ID No. 23 or SEQ ID No. 25 or SEQ ID No. 27 or
25 SEQ ID No. 29 or SEQ ID No. 31 or SEQ ID No. 33.

In one embodiment of the present invention the protein encoded by the *gsh1* gene or the nucleic acid encodes a polypeptide having at least 85% identity with SEQ ID No. 3 or SEQ ID No. 9 or SEQ ID No. 11 or SEQ ID No. 13 or SEQ ID No. 15 or SEQ ID No. 17 or SEQ ID No. 19 or SEQ ID No. 21 or SEQ ID No. 23 or SEQ ID No. 25 or SEQ ID No. 27 or SEQ ID
30 No. 29 or SEQ ID No. 31 or SEQ ID No. 33.

In one embodiment of the present invention the protein encoded by the *gsh1* gene or the nucleic acid encodes a polypeptide having at least 90 or 95% identity with SEQ ID No. 3 or SEQ ID No. 9 or SEQ ID No. 11 or SEQ ID No. 13 or SEQ ID No. 15 or SEQ ID No. 17 or SEQ ID No. 19 or SEQ ID No. 21 or SEQ ID No. 23 or SEQ ID No. 25 or SEQ ID No. 27 or
35 SEQ ID No. 29 or SEQ ID No. 31 or SEQ ID No. 33.

In one embodiment of the present invention the protein encoded by the *gsh1* gene or the nucleic acid encodes a polypeptide having at least 98% or 99% identity with SEQ ID No. 3 or SEQ ID No. 9 or SEQ ID No. 11 or SEQ ID No. 13 or SEQ ID No. 15 or SEQ ID No. 17 or SEQ ID No. 19 or SEQ ID No. 21 or SEQ ID No. 23 or SEQ ID No. 25 or SEQ ID No. 27 or
 5 SEQ ID No. 29 or SEQ ID No. 31 or SEQ ID No. 33.

In one embodiment of the present invention the protein encoded by the *gsh1* gene or the nucleic acid encodes a polypeptide sequence shown herein as SEQ ID No. 3, or polypeptide sequence which comprises SEQ ID No. 3 with a conservative substitution of at least one of the amino acids, or a polypeptide having at least 70% (preferably 85%, more preferably 90%)
 10 identity with SEQ ID No. 3.

In one embodiment of the present invention in addition to increasing the expression of the *gsh1* gene (or the activity of the GSH1 protein) the plant is further modified to increase the expression of a *gsh2* gene or the activity of the protein encoded by the *gsh2* gene in said plant.

15 In one embodiment the increased expression of the *gsh2* gene is achieved by overexpressing or upregulating the *gsh2* gene.

In another embodiment the method of the present invention comprises in addition to increasing the expression of the *gsh1* gene (or the activity of the GSH1 protein), modifying the plant by increasing the expression of the *gsh2* gene in said plant root (e.g. in a plant root
 20 cell or plant root tissue).

In one embodiment the method of the present invention comprises expressing within a plant a polynucleotide (e.g. an exogenous polynucleotide) comprising a nucleic acid sequence encoding a GSH2 polypeptide.

In one embodiment the *gsh2* gene is the Arabidopsis *gsh2* gene, or an orthologue thereof.

25 The *gsh2* gene for use in the present invention may be from any source, e.g. from a plant (e.g. a dicot plant or a monocot plant); from a moss; from an algae, from a bacteria (including cyanobacteria); or from an animal.

In some embodiments, the *gsh2* gene for use in the present invention may be encoded by one or more of the following *gsh2* sequences identified by their Uniprot sequence identifier:

30 *D7M652, P46416, R0H761, V4SFK5, V4K412, A2Q600, U5FUH4, U5GDQ3, B9RF11, O22494, I1IJU8, Q0IP23, Q0IRE6, C5YP88, C5Y715, B4FWG0, Q7XB40, A9RTT8, D8SN58, D8SR77, A8IR74, D8UB55, C1MUC5, C1E4E9, A4RYM3, Q017J0, E1ZHD9, I0ZAA4.*

In one embodiment the *gsh2* gene for use in the present invention is encoded by a nucleotide
 35 sequence taught in any of the following as: SEQ ID No. 4 or SEQ ID No. 5 or SEQ ID No. 34

or SEQ ID No. 36 or SEQ ID No. 38 or SEQ ID No. 40 or SEQ ID No. 42 or SEQ ID No. 44 or SEQ ID No. 46 or SEQ ID No. 48 or SEQ ID No. 50 or SEQ ID No. 52.

The term "orthologue" as used herein means an homologous gene sequence found in different species. The term orthologue of the Arabidopsis *gsh2* gene may include one of the following genes from other species: SEQ ID No. 34 or SEQ ID No. 36 or SEQ ID No. 38 or
 5 SEQ ID No. 40 or SEQ ID No. 42 or SEQ ID No. 44 or SEQ ID No. 46 or SEQ ID No. 48 or SEQ ID No. 50 or SEQ ID No. 52.

In a further embodiment the polynucleotide (e.g. exogenous polynucleotide) for use in the present invention comprises a nucleic acid sequence encoding a GSH2 polypeptide operably
 10 linked with a heterologous promoter for directing transcription of said nucleic acid sequence in said plant. In one embodiment preferably the promoter is non-constitutive (e.g. in the sense that it does not direct expression in all or virtually all tissues), but has at least expression in the root (e.g. in a root cell or root tissue). Suitably the promoter (e.g. heterologous promoter) operably linked with the nucleic acid sequence encoding the GSH2
 15 polypeptide is a root-specific promoter or root-preferred promoter.

In one embodiment the *gsh2* gene or the nucleic acid sequence encoding said GSH2 polypeptide comprises:

- i) a polynucleotide sequence shown herein as SEQ ID No. 4 or SEQ ID No. 5 or SEQ ID No. 34 or SEQ ID No. 36 or SEQ ID No. 38 or SEQ ID
 20 No. 40 or SEQ ID No. 42 or SEQ ID No. 44 or SEQ ID No. 46 or SEQ ID No. 48 or SEQ ID No. 50 or SEQ ID No. 52, or
- ii) a functional fragment of the polynucleotide sequence shown in i) which functional fragment encodes a functional GSH2 polypeptide, or
- iii) a polynucleotide which encodes a polypeptide comprising the amino
 25 acid sequence shown herein as SEQ ID No. 6 or SEQ ID No. 35 or SEQ ID No. 37 or SEQ ID No. 39 or SEQ ID No. 41 or SEQ ID No. 43 or SEQ ID No. 45 or SEQ ID No. 47 or SEQ ID No. 49 or SEQ ID No. 51 or SEQ ID No. 53, or
- iv) a polynucleotide sequence which can hybridize to the polynucleotide
 30 taught in i), ii) or iii) above under high stringency conditions, or
- v) a polynucleotide sequence which has at least 70% (preferably 85%, more preferably 90%) identity with the polynucleotide shown in i), ii) or iii) above, or
- vi) a polynucleotide sequence which differs from polynucleotide shown in
 35 i), ii) or iii) due to degeneracy of the genetic code.

In one embodiment the *gsh2* gene or the nucleic acid sequence encoding said GSH2 polypeptide comprises:

- i) a polynucleotide sequence shown herein as SEQ ID No. 4 or SEQ ID No. 5, or
- 5 ii) a functional fragment of the polynucleotide sequence shown in i) which functional fragment encodes a functional GSH2 polypeptide, or
- iii) a polynucleotide which encodes a polypeptide comprising the amino acid sequence shown herein as SEQ ID No. 6, or
- 10 iv) a polynucleotide sequence which can hybridize to the polynucleotide taught in i), ii) or iii) above under high stringency conditions, or
- v) a polynucleotide sequence which has at least 70% (preferably 85%, more preferably 90%) identity with the polynucleotide shown in i), ii) or iii) above, or
- 15 vi) a polynucleotide sequence which differs from polynucleotide shown in i), ii) or iii) due to degeneracy of the genetic code.

In one embodiment the protein encoded by the *gsh2* gene or the nucleic acid encodes a polypeptide sequence shown herein as SEQ ID No. 6 or SEQ ID No. 35 or SEQ ID No. 37 or SEQ ID No. 39 or SEQ ID No. 41 or SEQ ID No. 43 or SEQ ID No. 45 or SEQ ID No. 47 or SEQ ID No. 49 or SEQ ID No. 51 or SEQ ID No. 53, or polypeptide sequence which
 20 comprises SEQ ID No. 6 or SEQ ID No. 35 or SEQ ID No. 37 or SEQ ID No. 39 or SEQ ID No. 41 or SEQ ID No. 43 or SEQ ID No. 45 or SEQ ID No. 47 or SEQ ID No. 49 or SEQ ID No. 51 or SEQ ID No. 53 with a conservative substitution of at least one of the amino acids, or a polypeptide having at least 70% identity with SEQ ID No. 6 or SEQ ID No. 35 or SEQ ID No. 37 or SEQ ID No. 39 or SEQ ID No. 41 or SEQ ID No. 43 or SEQ ID No. 45 or SEQ ID
 25 No. 47 or SEQ ID No. 49 or SEQ ID No. 51 or SEQ ID No. 53.

In one embodiment of the present invention the protein encoded by the *gsh2* gene or the nucleic acid encodes a polypeptide having at least 85% identity with SEQ ID No. 6 or SEQ ID No. 35 or SEQ ID No. 37 or SEQ ID No. 39 or SEQ ID No. 41 or SEQ ID No. 43 or SEQ ID No. 45 or SEQ ID No. 47 or SEQ ID No. 49 or SEQ ID No. 51 or SEQ ID No. 53.

30 In one embodiment of the present invention the protein encoded by the *gsh2* gene or the nucleic acid encodes a polypeptide having at least 90 or 95% identity with SEQ ID No. 6 or SEQ ID No. 35 or SEQ ID No. 37 or SEQ ID No. 39 or SEQ ID No. 41 or SEQ ID No. 43 or SEQ ID No. 45 or SEQ ID No. 47 or SEQ ID No. 49 or SEQ ID No. 51 or SEQ ID No. 53.

In one embodiment of the present invention the protein encoded by the *gsh2* gene or the
 35 nucleic acid encodes a polypeptide having at least 98% or 99% identity with SEQ ID No. 6 or

SEQ ID No. 35 or SEQ ID No. 37 or SEQ ID No. 39 or SEQ ID No. 41 or SEQ ID No. 43 or SEQ ID No. 45 or SEQ ID No. 47 or SEQ ID No. 49 or SEQ ID No. 51 or SEQ ID No. 53.

In one embodiment the protein encoded by the *gsh2* gene or the nucleic acid encodes a polypeptide sequence shown herein as SEQ ID No. 6, or polypeptide sequence which

- 5 comprises SEQ ID No. 6 with a conservative substitution of at least one of the amino acids, or a polypeptide having at least 70% (preferably 85%, more preferably 90%) identity with SEQ ID No. 6.

In one embodiment, the *gsh2* gene for use in the present invention is encoded by SEQ ID No. 4 or SEQ ID No. 5 taught herein.

- 10 In one embodiment at least GSH1 is overexpressed and/or upregulation.

In another embodiment GSH1 and GSH2 are overexpressed and/or upregulated.

In another embodiment in addition to increasing the expression of (e.g. overexpressing and/or upregulating) GSH1 and/or GSH2, the expression of one of the following may also be increased (e.g. overexpressed and/or upregulated): phytochelatin synthase (PCS) (E.C.

- 15 2.3.2.15) (e.g. PSC1), an ABCC transporter (e.g. ABCC1 and/or ABCC2).

Without wishing to be bound by theory, phytochelatin synthase (PCS) synthesizes phytochelatin from glutathione (e.g. as prepared by the glutathione biosynthesis disclosed herein). The amount of phytochelatin increases when the cell needs more phytochelatin, e.g. in high concentrations of heavy metal ions. The increase in PCS thus converts the metal

- 20 bound glutathione to a metal-phytochelatin complex. In a further embodiment of the present invention addition to overexpressing and/or upregulating GSH1 and/or GSH2, one may further overexpress and/or upregulate PCS.

Without wishing to be bound by theory, the metal-phytochelatin complex can be sequestered into the vacuole of the plants root cells by ABCC transporters, such as ABCC1 and/or

- 25 ABCC2. In one embodiment in addition to overexpressing and/or upregulating GSH1 and/or GSH2 the overexpression of ABCC transporters ABCC1 and/or ABCC2 is also contemplated. A gene of interest in accordance with the present invention is preferably a *gsh1* gene. In some embodiments the gene of interest may be a *gsh2* gene, and/or an *ABCC transporter* gene, and/or a *PCS* gene.

- 30 A protein of interest in accordance with the present invention is preferably GSH1. In some embodiments the protein of interest may be a GSH2, and/or an ABCC transporter, and/or PCS.

The "increase in an expression level of GSH1" means that a plant is increased in the mRNA level or the protein level in comparison with an expression level of GSH1 of a parent plant of the same breed. The expression level of GSH1 is compared with that of GSH1 at a

35 corresponding part in the parent plant of the same breed cultured under the same condition.

A case where the expression level increases at least 1.1 times greater than that of the parent plant is preferably considered as a case where the expression level is increased. Here, it is more preferable that the expression level of the plant has a significant difference of 5% by a t-test compared with that of the parent plant, in order to be considered that there is an increase in the expression level. It is preferable that the expression levels of the plant and the parent plant be measured at the same time by the same method. However, data stored as background data may be also used.

The plant of the present invention may be any plant where the aerial parts of the plant (and particularly the leaves of the plant) are consumed.

- 10 In one embodiment the plant (e.g. the modified, consumable or unmodified plant) according to the present invention is a plant selected from the group consisting of tobacco, a salad leaf crop and a leaf vegetable.

In one embodiment the plant (e.g. the modified, consumable or unmodified plant) may be a salad leaf crop, such as a lettuce (including for example baby leaf spinach, Cos, Frisée, Iceberg lettuce, Lamb's lettuce, Little Gem, Mizuna, Radicchio, Red mustard, Red oak leaf, Rocket (Roquette), Ruby chard, Sweet Romaine), watercress and cress.

In another embodiment the plant (e.g. the modified, consumable or unmodified plant) may be a leaf vegetable, such as a cabbage, spinach, kale, tea, chicory (or curly endive), Phak chet, or a herb (such as basil, mint or oregano).

- 20 In one embodiment, the plant (e.g. the modified, consumable or unmodified plant) is of the family Solanaceae, more preferably of the subfamily Cestoideae, more preferably of the genus *Nicotiana*, and most preferably the plant (e.g. the modified, consumable or unmodified plant) is *Nicotiana tabacum* or *N. rustica*.

In one preferred embodiment the plant (e.g. the modified, consumable or unmodified plant) is a tobacco plant. In one preferred embodiment the plant (e.g. the modified, consumable or unmodified plant) is an oriental tobacco plant.

In one embodiment, the plant (e.g. the modified, consumable or unmodified plant) is a tobacco plant, which refers to a plant belonging to the genus *Nicotiana*. Preferred species, cultivars, hybrids, and varieties of tobacco plant are described herein.

- 30 The disclosed compositions and methods can be applied to any species of the genus *Nicotiana*, including *N. rustica* and *N. tabacum* (for example, LA B21, LN KY171, TI 1406, Basma, Galpao, Perique, Beinhart 1000-1, and Petico). Other species include *N. acaulis*, *N. acuminata*, *N. acuminata var. multiflora*, *N. africana*, *N. alata*, *N. amplexicaulis*, *N. arentsii*, *N.*

attenuata, *N. benavidesii*, *N. benthamiana*, *N. bigelovii*, *N. bonariensis*, *N. cavicola*, *N. clevelandii*, *N. cordifolia*, *N. corymbosa*, *N. debneyi*, *N. excelsior*, *N. forgetiana*, *N. fragrans*, *N. glauca*, *N. glutinosa*, *N. goodspeedii*, *N. gossei*, *N. hybrid*, *N. ingulba*, *N. kawakamii*, *N. knightiana*, *N. langsdorffii*, *N. linearis*, *N. longiflora*, *N. maritima*, *N. megalosiphon*, *N. miersii*,
 5 *N. noctiflora*, *N. nudicaulis*, *N. obtusifolia*, *N. occidentalis*, *N. occidentalis subsp. hesperis*, *N. otophora*, *N. paniculata*, *N. pauciflora*, *N. petunioides*, *N. plumbaginifolia*, *N. quadrivalvis*, *N. raimondii*, *N. repanda*, *N. rosulata*, *N. rosulata subsp. ingulba*, *N. rotundifolia*, *N. setchellii*, *N. simulans*, *N. solanifolia*, *N. spegazzinii*, *N. stocktonii*, *N. suaveolens*, *N. sylvestris*, *N. thyrsoiflora*, *N. tomentosa*, *N. tomentosiformis*, *N. trigonophylla*, *N. umbratica*, *N. undulata*, *N.*
 10 *velutina*, *N. wigandioides*, and *N. x sanderae*.

Particularly useful *Nicotiana tabacum* varieties include Burley type, dark type, flue-cured type, and Oriental type tobaccos.

In some embodiments the present invention is particularly relevant for Oriental type tobacco. Non-limiting examples of varieties or cultivars are: Izmir, BD 64, CC 101, CC 200, CC 27, CC
 15 30 1, CC 400, CC 500, CC 600, CC 700, CC 800, CC 900, Coker 176, Coker 319, Coker 371 Gold, Coker 48, CD 263, DF911, 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 50 1 LC, K 149, K 326, K 346, K 358, K394, K 399, K 730, KDH 959, KT
 20 200, 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-7371LC, 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, 'Perique' tobacco, PVH03, PVH09, PVH19, PVH50, PVH51, R 610, R 630, R 7-11, R 7-12,
 25 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,
 30 Burley 21, KY 8959, Y 9, MD 609, PG 0 1, PG 04, P0 1, P0 2, P0 3, RG 11, 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
 35 21, NC 2326, NC 297, PVH 2110, 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.

In a particularly preferred embodiment the tobacco plant is a Burley type tobacco plant. Suitably a Burley PH2517.

5 As used herein, the term 'plant' refers to any plant at any stage of its life cycle or development and its progenies. In general, unless otherwise specified, when referring to a "plant" it is intended to cover a plant at any stage of development, including single cells and seeds. Thus in particular embodiments, the present invention provides a plant cell, e.g. an isolated plant cell, having one or more characteristics of a "modified plant" as defined herein.

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

The term "plant propagation material" as used herein refers to any plant matter taken from a plant from which further plants may be produced. Suitably the plant propagation material may be a seed.

15 In one embodiment the modified plant is a transgenic plant.

The term "consumed" as used herein means ingested by a human or animal (preferably human). The term "consumable" as used herein means being ingestable by a human or animal (preferably human). Ingestion may be in the form of eating, e.g. entering the body via the mouth for digestion and absorption, in which case the plants will be edible plants. In
20 other embodiments the plants may be consumed or consumable by burning or heating the plant material or an extract (e.g. a tobacco extract) thereof and inhaling the fumes or smoke thus produced. In the latter case the consumption may be via the mouth and lungs.

The present invention further provides for harvested leaf of a plant in accordance with the present invention. Suitably the harvested leaf may be a harvested tobacco leaf.

25 The term harvested means that the leaf or leaves of the plant are removed from the roots of the plant. The harvested leaf may be comprised of leaf and stem material.

Suitably the harvested leaf (e.g. tobacco leaf) may be subjected to downstream processing. Thus in one embodiment the harvested leaf may be processed to produce a processed leaf.

The term "processed tobacco leaf" as used herein refers to a tobacco leaf that has
30 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.

Suitably the leaf (e.g. the tobacco leaf) may be subjected to curing, fermenting, pasteurising or combinations thereof.

In some embodiments the leaf (e.g. tobacco leaf) is subjected to curing followed by either fermentation or pasteurisation or a combination thereof.

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.

Preferably the processed leaf (e.g. processed tobacco leaf) does not comprise material which is capable of reproduction.

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. Most preferably no cells are viable in the processed plant material, e.g. processed tobacco material.

Suitably the processed tobacco leaf may be processed by curing.

These processing methods preferably result in the processed leaf being non-viable.

Leaves (e.g. tobacco leaf) may be cured by any method known in the art. In one embodiment leaf (e.g. 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.

Suitably the leaf (e.g. tobacco leaf) may be air cured.

Typically air curing is achieved by hanging leaf (e.g. 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 leaf (e.g. tobacco leaf) may be fire cured. Fire curing is typically achieved by hanging leaf (e.g. 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 plant (e.g. type of tobacco).

In another embodiment the leaf (e.g. tobacco leaf) may be flue cured. Flue curing of tobacco 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 leaf (e.g. tobacco leaf) may be sun cured. This method typically involves exposure of uncovered plants or harvested leaf (e.g. tobacco) to the sun.

Suitably the processed leaf (e.g. tobacco leaf) may be processed by fermenting.

Fermentation can be carried out in any manner known in the art. Typically during fermentation of tobacco, 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 leaf (e.g. tobacco leaf) may be processed by pasteurising. Pasteurising may be particularly preferred when the tobacco leaf will be used to make a smokeless tobacco 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 harvested leaf or processed leaf is cut into cut leaf. The leaf may be cut before or after processing. Suitably tobacco leaf may be cut after processing. Cut tobacco leaf as used herein means tobacco leaf which has been cut such that it can be used to produce tobacco products (such as cigarettes, tobacco heating devices and the smokeless tobacco products and the like).

Plant extracts (e.g. tobacco extracts) may be produced by any known method. One example of a suitable extraction process is taught in EP 0 862 865, which method is hereby incorporated 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.

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.

In one embodiment a consumable plant product may be prepared from the plant or a part thereof according to the present invention.

5 In one embodiment the consumable plant product is a tobacco product.

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

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

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

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

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

15 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 product may be a smokeless tobacco product.

20 The term “smokeless tobacco product” as used herein refers to a tobacco product that is not intended to be smoked and/or subjected to combustion. In one embodiment a smokeless tobacco product may include snus, snuff, chewing tobacco or the like.

In a further embodiment the tobacco product may be a tobacco heating device.

25 Typically in heated smoking articles, 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.

30 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.

35 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 product in accordance with the present invention.

- 5 The terms “root-specific promoter” or “root-preferred promoter” means plant promoters containing elements that drive the expression of genes of interest in root cells and tissues.

In one embodiment the term “root-preferred” means that the expression driven by a plant promoter is selectively enhanced in the roots in comparison to other tissues of the plant. In some embodiments root-preferred promoters may drive low levels of expression in tissues
10 other than root cells and tissues; however, they predominantly drive expression in the roots.

In one embodiment the term “root-specific” means that the expression driven by a plant promoter is exclusively or substantially exclusively in the roots in comparison with other tissues of the plant.

- 15 Any root-specific or root-preferred promoter may be used in accordance with the present invention.

- In some embodiments the root-specific or root-preferred promoter may be one selected from the group consisting of *A622*; *AtWRKY6* (see Werner *et al.* (2011), "Root-Specific Reduction of Cytokinin Causes Enhanced Root Growth, Drought Tolerance, and Leaf Mineral Enrichment in Arabidopsis and Tobacco", *Plant Cell*, **22**(12), 3905-3920); *NsPMT1* (see Shoji
20 *et al.* (2000), "Jasmonate induction of putrescine N-methyltransferase genes in the root of *Nicotiana sylvestris*", *Plant Cell Physiol.*, **41**(7), 831-839); *NsPMT2* (see Shoji *et al.* (2000), "Jasmonate induction of putrescine N-methyltransferase genes in the root of *Nicotiana sylvestris*", *Plant Cell Physiol.*, **41**(7), 831-839); *NsPMT3* (see Shoji *et al.* (2000), "Jasmonate induction of putrescine N-methyltransferase genes in the root of *Nicotiana sylvestris*", *Plant*
25 *Cell Physiol.*, **41**(7), 831-839); *NtRB7* (see Conkling *et al.* (1990), "Isolation of Transcriptionally Regulated Root-Specific Genes From Tobacco", *Plant Physiol.*, **93**(3), 1203-1211); *PaHB* (see Bogusz *et al.* (1990), "Nonlegume Hemoglobin Genes Retain Organ-Specific Expression in Heterologous Transgenic Plants", *Plant Cell*, **2**(7), 633-641); *TtHb* (see Bogusz *et al.* (1990), "Nonlegume Hemoglobin Genes Retain Organ-Specific
30 Expression in Heterologous Transgenic Plants", *Plant Cell*, **2**(7), 633-641); *AtADH* (see Hossain *et al.* (2006), "Root specific expression of Na⁺ / H⁺ antiporter gene from *Synechocystis* sp. PCC 6803 confers salt tolerance of tobacco plant", *Plant Biotechnology*, **23**, 275-281); *EgTIP2* (see Rodrigues *et al.* (2013), "The tonoplast intrinsic aquaporin (TIP) subfamily of *Eucalyptus grandis*: Characterisation of *EgTIP2*, a root-specific and osmotic stress-responsive gene", *Plant Science*, **213**, 106-113); *AhAsy* (see Geng *et al.* (2014),
35

"Minig Tissue-specific Contigs from Peanut (*Archis hypogaea* L.) for Promoter Cloning by Deep Transcriptome Sequencing", *Plant Science*, **55**, 1793-1801).

Root cells and tissues include any part of the roots, and cover primary, lateral and adventitious roots.

- 5 One method for modifying the plant is through the use of genetic engineering techniques.

The soil bacterium *Agrobacterium tumifaciens* provides the tools for stable insertion of foreign genes into a plant and has been used in the transformation of many plant species, including tobacco, potato, tomato, *A. thaliana*, eucalyptus, etc. (Hoekema et al, 1983, *Nature*, 10 303:179-180, Bendahmane et al., 2000 *The Plant Journal*, 21 (1): 73-81)). The *A. tumifaciens* naturally transfers its own plasmid DNA into plant genomes as a means of infecting the plant. *A. tumifaciens* contains a plasmid separate from the bacterial chromosome, known as the Ti plasmid. Within the Ti plasmid there is a region of DNA which can be transferred to the infected plant known as transfer-DNA (T-DNA). Also contained in 15 the Ti plasmid are genes which facilitate the transfer of the T-DNA such as the vir region (a region which confers virulence for infection). A specific gene of interest (or genes) can be inserted into the transfer-DNA (T-DNA) of *A. tumifaciens* and this is then used to infect plants and generate transgenic populations.

As well as the gene of interest, a selectable marker gene may be used as part of the T-DNA, 20 such as neomycin phosphotransferase II (NPTII) which confers kanamycin antibiotic resistance to the plants expressing that gene, allowing a method of selection of transformed plants (Angenon et al., 1994 *Plant Mol. Biol Manu.* C1:1-13). The use of selectable markers in the present invention is not essential.

By "unmodified plant" it is intended to refer to a plant before transformation with the 25 exogenous gene. In other words, "modified" refers to a plant which has been modified to increase the expression of the gene of interest (e.g. *gsh1* gene), e.g. a transgenic plant created by transformation with the exogenous gene. Thus "unmodified" does not limit the nature of the plant in any other way. The unmodified plant may be a wild type plant derived from any species or strain, or may be plant which has already been modified by one or more 30 previous genetic modifications, including the introduction of other transgenes or the deletion or inactivation of endogenous genes.

A transgenic plant is generated by introduction of an exogenous gene of interested (e.g. encoding GSH1 enzyme) into the unmodified plant. By "exogenous gene" it is meant that the gene is transformed into the modified plant from an external source. The exogenous gene 35 may have a nucleic acid sequence identical to or different to an endogenous gene encoding the protein of interest (e.g. GSH1 or GSH2) in the modified plant. The exogenous gene may,

for example, be derived from a genomic DNA or cDNA sequence encoding the protein of interest (e.g. GSH1 or GSH2) from any species. Typically the exogenous gene is derived from a different source and has a sequence different to the endogenous gene. Alternatively, introduction of an exogenous gene having a sequence identical to the endogenous gene may be used to increase the number of copies of the endogenous gene sequence present in the plant. As a yet further alternative, introduction of an exogenous gene may mean introducing a gene which is identical (or functionally identical) to the endogenous gene but which is operably linked to an exogenous (non-natural or non-endogenous) promoter

In one embodiment the exogenous gene is not identical to an endogenous gene encoding the protein of interest (e.g. GSH1 or GSH2) in the plant. For instance, the exogenous gene preferably has less than 95% sequence identity with an endogenous gene encoding the protein of interest (e.g. GSH1 or GSH2) in the unmodified plant. More preferably the exogenous gene has less than 90%, less than 80%, less than 70%, less than 60%, less than 50% or less than 40% sequence identity with the endogenous gene of interest, e.g. *gsh1* gene or *gsh2* gene.

Similarly, it is preferred that the protein of interest, e.g. GSH1 or GSH2, encoded by the exogenous gene is not identical (at the amino acid/polypeptide level) to the protein (e.g. the GSH1 protein or the GSH2 protein) encoded by an endogenous gene in the plant. For instance, the exogenous gene product preferably has less than 95% sequence identity with an endogenous gene product in the unmodified plant. More preferably the exogenous gene has less than 90%, less than 80%, less than 70%, less than 60%, less than 50% or less than 40% sequence identity with the endogenous gene, e.g. *gsh1* gene or *gsh2* gene.

Preferably the exogenous gene encoding the protein of interest (e.g. GSH1 or GSH2) is a heterologous gene, which means that the exogenous gene is derived from a species different to the species of the unmodified plant. In one preferred embodiment, the heterologous gene is derived from a donor plant of the genus *Arabidopsis*, more preferably from *Arabidopsis thaliana*.

The genomic and cDNA sequences of a GSH1 from *Arabidopsis thaliana* are defined in SEQ ID No. 1 and SEQ ID No. 2, respectively.

The amino acid sequence of GSH1 encoded by these sequences is defined in SEQ ID No. 3. In one embodiment, a functional fragment of the nucleic acid as taught herein may be the cDNA sequence.

The term "functional fragment" as used herein refers to a portion of a polynucleotide that is capable of encoding GSH1 (or GSH2) that retains its activity. In other words a functional fragment a polynucleotide is able to encode a GSH1 (or GSH2) that has GSH1 (or GSH2) activity as defined herein.

In one embodiment the functional fragment may be a portion of a polynucleotide of the invention comprising at least 50 nucleotides, at least 75 nucleotides or at least 100 nucleotides. In some embodiments the functional fragment may comprise at least 150 nucleotides of a polynucleotide of the invention.

- 5 The term "degeneracy of the genetic code" as used herein refers to the redundancy in codons encoding polypeptide sequences exhibited as the multiplicity of three-codon combinations specifying an amino acid. For example in an mRNA molecule encoding a polypeptide having an isoleucine amino acid, isoleucine can be encoded by AUU, AUC or AUA. This means that a DNA molecule encoding the RNA can have multiple sequences yet
10 the resulting polypeptide will have the same sequence. In other words polymorphic nucleotide sequences can encode the same polypeptide product. This means that one nucleic acid sequence can comprise a sequence with very low sequence identity to a second sequence while encoding the same polypeptide sequence.

- Variants and fragments of the nucleic acid sequences according to the present invention
15 include genomic DNA sequences (e.g. comprising a full length gene locus including 5' and/or 3' non-translated regions), DNA sequences comprising a coding sequence and introns (e.g. excluding 5' and/or 3' non-translated regions), and cDNA sequences associated with the specified SEQ ID Nos.

- In one embodiment, the exogenous gene comprises one or more introns, which may be the
20 introns found in the genomic sequence from which the exogenous gene is derived.

In another embodiment, the exogenous gene comprises one or more introns, which may not be the introns found in the genomic sequence from which the exogenous gene is derived.

- The inclusion of introns in the exogenous gene may improve the stability of RNA transcribed therefrom in the transgenic plant, thereby enhancing expression levels of the protein of
25 interest, e.g. GSH1 or GSH2. The excision of introns involves a complex of proteins around the RNA, which may protect the RNA from degradation by enzymes which recognise it as "foreign".

- The exogenous gene may be introduced into the modified plant by any suitable transformation technique, provided that this leads to expression of the protein of interest (e.g.
30 GSH1 or GSH2) encoded by the exogenous gene in the transgenic plant. Typically the exogenous gene is a chimeric gene comprising the protein of interest (e.g. GSH1 or GSH2) coding sequence fused to a promoter sequence derived from a different gene. Such a chimeric gene may be cloned into any construct suitable for transforming plants.

- In embodiments of the invention, the heavy metal levels in the aerial parts (e.g. the leaves) of
35 the transgenic plant grown in a medium, e.g. soil, comprising heavy metals is reduced relative to an unmodified plant that is compared to a plant before transformation with the

exogenous gene. Heavy metal accumulation is compared by growing the plants under the same environmental conditions, including the same exposure to heavy metals of interest. To test the plants ability to not transport and/or accumulate heavy metals in the aerial parts (e.g. leaves thereof) the plants (both modified and unmodified) should be grown in an environment which contains one or more heavy metals, e.g. which contains one or more heavy metals in the growth media, e.g. soil, or tissue culture media. Since any reduction in heavy metal levels is desirable, the level of heavy metals in the aerial parts (e.g. leaves) of the transgenic plant is not particularly limiting provided that it is detectably lower than that of the unmodified plant. In one embodiment the heavy metal reduction is a value of at least about 30%, preferably at least about 50%, more preferably at least about 60%, more preferably at least about 70%, more preferably at least about 80%, most preferably at least about 90% or even 95% w/w. In one embodiment the heavy metal reduction is a value of at least about 60% w/w. The reduction is compared with an unmodified plant (e.g. one which does not have increased expression of at least *gsh1* gene (or the *gsh1* and *gsh2* genes) or one where the activity of the protein encoded by the *gsh1* gene (or the *gsh1* and *gsh2* genes) is not increased) under the same growing conditions.

Heavy metal levels may be measured by any suitable technique. An official method available from Health Canada – Protocol T-306 “Determination of Ni, Pb, Cd, Cr, As, Se and Hg in Whole tobacco – 31 December 1999” may be used. Suitably, Inductively Coupled Argon Plasma (ICP) with the use of HCl in the digestion procedure, Atomic Emission Spectroscopy (AES) may be used. Alternatives may include the use of Atomic Absorption Spectroscopy (AA). In one preferred embodiment the method detailed in Health Canada – Protocol T-306 is used to measure heavy metal levels (this method may be adjusted to account for smaller sample sizes as required).

In some embodiments of the present invention, the introduction of the exogenous gene into an unmodified plant generates a primary transgenic plant, i.e. a T0 plant produced by direct transformation with the exogenous gene. A secondary transgenic plant (i.e. a T1 plant) may be produced by propagation of the primary transgenic plant, for instance by sexual or asexual reproduction thereof. Preferably the secondary transgenic plant is produced by selfing the primary transgenic plant, i.e. by self-fertilization or self-pollination. In this way, it is possible to generate secondary transgenic plants which are homozygous for the exogenous gene of interest (e.g. *gsh1* or *gsh2*). It is expected that 25% of secondary transgenic plants generated by selfing a primary transgenic plant will be homozygous for the exogenous gene.

In one embodiment, preferably the plants in accordance with the present invention are prepared using double haploids. Double haploids may be prepared by using tissue culture

techniques referred to as “anther culture” and “isolated microspore culture” where immature pollen grains can grow to product colonies of cells. The colonies may be transferred to media with different plant growth regulators and sugars to induce growth of shoots and then roots.

The present method preferably involves generating a plurality of transgenic plants by

5 independent transformation of a plurality of unmodified plants with the exogenous gene.

In other words, the method may be repeated on multiple individual plants to produce a series of transgenic plants derived from individual transformation events. Plant lines derived from each of these transgenic plants may differ in their properties, including the extent to which heavy metal accumulation in the aerial parts (e.g. leaves) is reduced. Thus in some

10 embodiments, the method involves screening the transgenic plants produced by the method (either primary, secondary or subsequent generation plants) and selecting those plants having desirable properties (for example a reduction in heavy metal levels in the aerial parts – e.g. leaves – of the plant) for further propagation.

Preferably a primary transgenic plant generated by introduction of the exogenous gene

15 contains a single copy of the exogenous gene. The method preferably involves detecting the copy number of the exogenous gene in the primary transgenic plant, for instance before selecting primary transgenic plants having a single copy for propagation.

In another embodiment, the method further comprises determining the level of one or more heavy metals in the aerial parts, e.g. leaves, of each of a plurality of transgenic plants

20 generated by independent transformation events. Heavy metal levels may be determined in the primary, secondary or subsequent generation transgenic plants. The method preferably further comprises selecting one or more transgenic plants having reduced levels of one or more heavy metals relative to an unmodified plant, and propagating the transgenic plants having reduced levels.

25 In one embodiment the modified plants of the present invention have reduced levels of at least 1 heavy metal.

In another embodiment the modified plants of the present invention have reduced levels of at least 2 heavy metals.

In a yet further embodiment the modified plants of the present invention have reduced levels

30 of at least 3 heavy metals.

In certain embodiments of the present invention, chimeric genes encoding a protein of interest (e.g. GSH1, or GSH1 and GSH2) may be transformed into plant cells leading to controlled expression of the protein of interest (e.g. GSH1, or GSH1 and GSH2) under the direction of a promoter. The promoters may be obtained from different sources including

35 animals, plants, fungi, bacteria, and viruses. Promoters may also be constructed synthetically.

Exogenous genes 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 gene of interest (e.g. GSH1, or GSH2) coding sequence, optionally including introns, 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 expression cassette 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. 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 *GSH1* or *GSH2*) 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*. When referring to a "chimeric gene", it is meant that the nucleic acid sequence encoding a gene of interest (e.g. *GSH1* or *GSH2*) is derived from a different origin (e.g. from a different gene, or from a different species) to the promoter sequence which directs its expression.

The expression cassette may be incorporated into a basic plant transformation vector, such as *pBIN 19 Plus*, *pBI 101*, 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.

The term "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. The term "incorporated" preferably covers stable incorporation into the genome.

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).

Typically, in *Agrobacterium-mediated* transformation a binary vector carrying a foreign DNA of interest, i.e. a chimeric gene, 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) 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 D. N. et al., (1980), *Tissue Culture Methods for Plant Pathologists*, eds.: D. S. Ingrams and J.P. Helgeson, 203-208.

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, including the silicon carbide whisker technique are taught in Frame B R, Drayton P R, Bagnaall S V, Lewnau C J, Bullock

W P, Wilson H M, Dunwell J M, Thompson J A & Wang K (1994). Production of fertile transgenic maize plants by silicon carbide whisker-mediated transformation is taught in *The Plant Journal* 6: 941-948) and viral transformation techniques is taught in for example Meyer P, Heidmm I & Niedenhof I (1992). The use of cassava mosaic virus as a

vector system for plants is taught in *Gene* 110: 213-217. Further teachings on plant

transformation may be found in EP-A-0449375.

In a further aspect, the present invention relates to a vector system which carries a nucleotide sequence encoding a gene of interest (e.g. *gsh1* or *gsh2*) and introducing it into the genome of an organism, such as a 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

One extensively employed system for transformation of plant cells uses the Ti plasmid from *Agrobacterium tumefaciens* or a Ri plasmid from *Agrobacterium rhizogenes* Anetal., (1986), *Plant Physiol.* 81, 301-305 and Butcher D. N. et al., (1980), *Tissue Culture Methods for Plant Pathologists*, eds.: D. S. Ingrams and J.P. Helgeson, 203-208. 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, in: *The Binary Plant Vector System* Offset-drukkerij Kanters B. B., Amsterdam, 1985, Chapter V; Fraley, et al., *Crit. Rev. Plant Sci.*, 4:1-46; and Anetal., *EMBO J* (1985) 4:277-284.

Plant cells transformed with an exogenous gene encoding a protein of interest (e.g. GSH1 or GSH2) 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 an exogenous gene encoding a gene of interest, e.g. *gsh1* or *gsh2*, according to the present invention. Preferably the exogenous gene is incorporated in the genome of the plant.

The terms "transgenic plant" and "chimeric gene" do 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 nucleic acid sequence, chimeric gene, plant transformation vector or 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 nucleic acid sequence, chimeric gene, plant transformation vector or 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.

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.

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 does not include the native nucleotide sequence according to the present invention when in its natural environment and when it is 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 "non-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 naturally associated, which promoter is also in its native environment. However, the amino acid sequence encompassed by scope the present invention can be isolated and/or purified post expression of a nucleotide sequence in its native organism. Preferably, however, the amino acid sequence encompassed by scope of the present invention may be expressed by a nucleotide sequence in its native organism but wherein the nucleotide sequence is not under the control of the promoter with which it is naturally associated within that organism.

Typically, the nucleotide sequence encompassed by the scope of the present invention is prepared using recombinant DNA techniques (i.e. recombinant DNA). However, in an alternative embodiment of the invention, the nucleotide sequence could be synthesised, in whole or in part, using chemical methods well known in the art (see Caruthers MH *et al.*, (1980) *Nuc Acids Res Symp Ser* 215-23 and Horn T *et al.*, (1980) *Nuc Acids Res Symp Ser* 225-232).

A nucleotide sequence encoding either a protein which has the specific properties 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.

By way of further example, a genomic DNA and/or cDNA library may be constructed using chromosomal DNA or messenger RNA from the organism producing the enzyme. If the

amino acid sequence of the enzyme is known, labelled oligonucleotide probes may be synthesised and used to identify enzyme-encoding clones from the genomic library prepared from the organism. Alternatively, a labelled oligonucleotide probe containing sequences homologous to another known enzyme gene could be used to identify enzyme-encoding clones. In the latter case, hybridisation and washing conditions of lower stringency are used.

In a yet further alternative, the nucleotide sequence encoding the enzyme may be prepared synthetically by established standard methods, e.g. the phosphoroamidite method described by Beaucage S.L. *et al.*, (1981) *Tetrahedron Letters* **22**, p 1859-1869, or the method described by Matthes *et al.*, (1984) *EMBO J.* **3**, p 801-805. In the phosphoroamidite method, oligonucleotides are synthesised, e.g. in an automatic DNA synthesiser, purified, annealed, ligated and cloned in appropriate vectors.

The nucleotide sequence may be of mixed genomic and synthetic origin, mixed synthetic and cDNA origin, or mixed genomic and cDNA origin, prepared by ligating fragments of synthetic, genomic or cDNA origin (as appropriate) in accordance with standard techniques. Each ligated fragment corresponds to various parts of the entire nucleotide sequence. The DNA sequence may also be prepared by polymerase chain reaction (PCR) using specific primers, for instance as described in US 4,683,202 or in Saiki R K *et al.*, (*Science* (1988) **239**, pp 487-491).

The scope of the present invention also encompasses amino acid sequences of enzymes having the specific properties as defined herein.

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 amino acid sequence may be prepared/isolated from a suitable source, or it may be made synthetically or it may be prepared by use of recombinant DNA techniques.

Preferably the amino acid sequence when relating to and when encompassed by the *per se* scope of the present invention is not a native enzyme. In this regard, the term "native enzyme" means an entire enzyme that is in its native environment and when it has been expressed by its native nucleotide sequence.

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 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 enzyme.

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.

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

In one embodiment the present invention relates to a protein whose amino acid sequence is represented herein or a protein derived from this (parent) protein by substitution, deletion or addition of one or several amino acids, such as 2, 3, 4, 5, 6, 7, 8, 9 amino acids, or more amino acids, such as 10 or more than 10 amino acids in the amino acid sequence of the parent protein and having the activity of the parent protein.

In one embodiment the present invention relates to a nucleic acid sequence (or gene) encoding a protein whose amino acid sequence is represented herein or encoding a protein derived from this (parent) protein by substitution, deletion or addition of one or several amino acids, such as 2, 3, 4, 5, 6, 7, 8, 9 amino acids, or more amino acids, such as 10 or more than 10 amino acids in the amino acid sequence of the parent protein and having the activity of the parent protein.

Homology or 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.

5 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
10 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
15 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
20 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
25 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
30 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 %
35 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	

5

In one embodiment, CLUSTAL may be used with the gap penalty and gap extension set as defined above.

Suitably, the degree of identity with regard to a nucleotide sequence is determined over at least 20 contiguous nucleotides, preferably over at least 30 contiguous nucleotides, preferably over at least 40 contiguous nucleotides, preferably over at least 50 contiguous nucleotides, preferably over at least 60 contiguous nucleotides, preferably over at least 100 contiguous nucleotides.

Suitably, the degree of identity with regard to a nucleotide sequence may be determined over the whole sequence.

15 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

20 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.

Conservative substitutions may be made, for example according to the Table below. Amino

25 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

40

	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.

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, 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

RJ *et al.*, *PNAS* (1992) **89**(20), 9367-9371 and Horwell DC, *Trends Biotechnol.* (1995) **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 stringent 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 stringent conditions (e.g. 65°C and 0.1xSSC {1xSSC = 0.15 M NaCl, 0.015 M Na₃citrate pH 7.0}).

In one aspect the sequence for use in the present invention is a synthetic sequence – i.e. a sequence that has been prepared by *in vitro* chemical or enzymatic synthesis. It includes, but is not limited to, sequences made with optimal codon usage for host organisms.

The term "expression vector" means a construct capable of *in vivo* or *in vitro* expression.

Preferably, the expression vector is incorporated into the genome of a suitable host organism.

The term "incorporated" preferably covers stable incorporation into the genome.

The nucleotide sequence of 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 vectors for use in the present invention may be transformed into a suitable host cell as described herein to provide 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.

Thus, in a further embodiment, the invention provides a method of making nucleotide sequences of the present invention by introducing a nucleotide sequence of the present invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector.

The vector may further comprise a nucleotide sequence enabling the vector to replicate in the host cell in question. Examples of such sequences are the origins of replication of plasmids pUC19, pACYC177, pUB110, pE194, pAMB1 and pIJ702.

In some applications, the nucleotide sequence for use in the present invention is operably linked 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 the present invention operably linked to such a regulatory sequence, i.e. the vector is an expression vector.

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 condition 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.

Enhanced expression of the nucleotide sequence encoding the enzyme of the present invention may also be achieved by the selection of heterologous regulatory regions, e.g. promoter, secretion leader and terminator regions.

Preferably, the nucleotide sequence according to the present invention is operably linked to at least a promoter.

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

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

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.

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

- 10 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.

- 15 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
20 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.

- 25 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
30 with the term "peptide". In some instances, the term "amino acid sequence" is synonymous with the term "enzyme".

- The terms "protein" and "polypeptide" are used interchangeably herein. 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
35 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.

ADVANTAGES

One advantage of the present invention is that plants grown in contaminated soils, e.g. soils contaminated with at least one heavy metal, have reduced transportation and/or assimilation of those heavy metals into the aerial parts (e.g. leaves) of the plants. This has significant advantages for crops and plant whose aerial parts (particularly their leaves) are consumed, e.g. by humans or animals (preferably humans).

Modification of the plant by increasing the expression of at least a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene may have advantages over other enzymes in the roots because it is capable of reducing the transportation and/or assimilation of at least 1, preferably at least 2, more preferably at least 3 different heavy metals into the aerial parts of the plant (e.g. the leaves). This contrasts sharply previous attempts to reduce heavy metal uptake in plants which have used systems which are restricted to reducing a single heavy metal species.

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

The invention will now be described, by way of example only, with reference to the following

5 Figures and Examples.

EXAMPLES

Example 1 – Overexpression of *GSH1* under the control of a root specific promoter reduces leaf heavy metal levels.

Materials and Methods

Tobacco plants (*Nicotiana tabacum* cv. Virginia 40), agrobacterium transformed with a *GSH1* gene (*Arabidopsis thaliana* wild type coding sequence) under control of the tobacco *A622* promoter (*Nicotiana tabacum* cv. Burley 21) are sown in soil as cress. Empty vector (*A622* promoter with no gene attached) agrobacterium transformed tobacco plants (*Nicotiana*
15 *tabacum* cv. Virginia 40) are sown similarly to be used as controls. After 14-21 days, depending on development, the seedlings are individually extracted from the soil, the roots washed and the seedlings transferred to 200g of moler clay with a holding capacity of 1mL/g. The plants are then grown for four weeks after which they are treated with a single dose of
20 10mL of 100µM CdCl₂, corresponding to 50ppm. After two weeks all leaf material from each plant is collected, with each plant constituting a single biological replicate. For each line five biological replicates are obtained. The leaf material is freeze dried and weighed and the amount of cadmium present, expressed as ng cadmium/dry weight leaf material is determined according to the method set out in Health Canada – Official Method Protocol T-
25 306 “Determination of Ni, Pb, Cd, Cr, As, Se and Hg in Whole tobacco – 31 December 1999” modified to allow analysis of smaller sample amounts.

Results and Discussion

A number of independent transgenic tobacco lines carrying either the empty vector control or
30 the *A622-GSH1* construct are screened as described above. It is found that the presence of the *A622-GSH1* construct results in up to 61% reduction of cadmium concentrations in the leaves, compared to leaves of plants carrying the empty vector. Therefore the *A622-GSH1* construct, when inserted as a transgene, is clearly able to significantly reduce the amount of cadmium present in the leaves of the plants in which this construct has been inserted,
35 showing that expression of the *GSH1* gene in the roots of the plants is sufficient to achieve the effect of significantly reduced concentrations of cadmium in the leaves. Figure 2 shows

cadmium levels detected in leaves of tobacco plants carrying either the empty vector control or a *A622-GSH1* construct. The presence of the *A622-GSH1* construct results in up to 61% reduction of cadmium concentrations in the leaves, compared to leaves of plants carrying an empty vector control construct.

Therefore, the upregulation of GSH1 resulted in a significant reduction in heavy metals, e.g. cadmium, in the leaves of the tobacco. Thus, overexpression of *GSH1* using a root-specific or root-preferred promoter can lead to a significant reduction in the level of heavy metals in the leaves (e.g. the harvested leaves) of the tobacco plant.

Example 2 – Overexpression of *GSH1* & *GSH2* under the control of a root specific promoter reduces leaf heavy metal levels.

Materials and Methods

Tobacco plants (*Nicotiana tabacum* cv. Virginia 40), agrobacterium transformed with a *GSH1* gene (*Arabidopsis thaliana* wild type coding sequence) under control of the tobacco *A622* promoter (*Nicotiana tabacum* cv. Burley 21) are crossed with tobacco plants (*Nicotiana tabacum* cv. Virginia 40), agrobacterium transformed with a *GSH2* gene (*Arabidopsis thaliana* wild type coding sequence) under control of the tobacco *A622* promoter (*Nicotiana tabacum* cv. Burley 21). Plants homozygous for both the *A622-GSH1* and the *A622-GSH2* construct are identified in the F2 population and the offspring from these are used for screening by sowing in soil as cress. Empty vector (*A622* promoter with no gene attached) agrobacterium transformed tobacco plants (*Nicotiana tabacum* cv. Virginia 40) are sown similarly to be used as controls. After 14-21 days, depending on development, the seedlings are individually extracted from the soil, the roots washed and the seedlings transferred to 200g of moler clay with a holding capacity of 1mL/g. The plants are then grown for four weeks after which they are treated with a single dose of 10mL of 100µM CdCl₂, corresponding to 50ppm. After two weeks all leaf material from each plant is collected, with each plant constituting a single biological replicate. For each line five biological replicates are obtained. The leaf material is freeze dried and weighed and the amount of cadmium present, expressed as ng cadmium/dry weight leaf material is determined according to the method set out in Health Canada – Official Method Protocol T-306 “Determination of Ni, Pb, Cd, Cr, As, Se and Hg in Whole tobacco – 31 December 1999” modified to allow analysis of smaller sample amounts.

Results and Discussion

A number of independent transgenic tobacco lines carrying either the empty vector control or the *A622-GSH1* and the *A622-GSH2* construct (*A622-GSH1/A622-GSH1*, *A622-GSH2/A622-GSH2* genotype) are screened as described above. The parental lines for the *A622-GSH1/A622-GSH1*, *A622-GSH2/A622-GSH2* plants are also screened as described above to allow identification of cumulative effects of the two transgenes.

Example 3 – Overexpression of at least *GSH1* under the control of a root specific promoter reduces leaf heavy metal levels (including Cd, Cr, As).

Materials and Methods

Tobacco plants (*Nicotiana tabacum* cv. Virginia 40), agrobacterium transformed with a *GSH1* gene (*Arabidopsis thaliana* wild type coding sequence) under control of the tobacco *A622* promoter (*Nicotiana tabacum* cv. Burley 21) are crossed with tobacco plants (*Nicotiana tabacum* cv. Virginia 40), agrobacterium transformed with a *GSH2* gene (*Arabidopsis thaliana* wild type coding sequence) under control of the tobacco *A622* promoter (*Nicotiana tabacum* cv. Burley 21). Plants homozygous for both the *A622-GSH1* and the *A622-GSH2* construct are identified in the F2 population and the offspring from these are used for screening by sowing in soil as cross. Additionally the parental lines used for crossing, are sown. Finally, empty vector (*A622* promoter with no gene attached) agrobacterium transformed tobacco plants (*Nicotiana tabacum* cv. Virginia 40) are sown similarly to be used as controls. After 14-21 days, depending on development, the seedlings are individually extracted from the soil, the roots washed and the seedlings transferred to 200g of molar clay with a holding capacity of 1mL/g. The plants are then grown for four weeks after which they are treated either with a single dose of 10mL of 100µM CdCl₂, 10mL of 100µM K₂CrO₄ or 10mL 100µM KH₂AsO₄ or 10mL of 100µM CdCl₂, 100µM K₂CrO₄ and 100µM KH₂AsO₄. After two weeks all leaf material from each plant is collected, with each plant constituting a single biological replicate. For each line five biological replicates are obtained. The leaf material is freeze dried and weighed and the amount of cadmium, chromium and/or arsenic present, expressed as ng cadmium/g dry weight leaf material, chromium/g dry weight leaf material and/or arsenic/g dry weight leaf material is determined according to the method set out in Health Canada – Official Method Protocol T-306 “Determination of Ni, Pb, Cd, Cr, As, Se and Hg in Whole tobacco – 31 December 1999” modified to allow analysis of smaller sample amounts.

Results and Discussion

A number of independent transgenic tobacco lines carrying the empty vector control or the *A622-GSH1* and/or the *A622-GSH2* construct are screened as described above. The parental lines for the *A622-GSH1/A622-GSH1*, *A622-GSH2/A622-GSH2* plants are also screened as described above to allow identification of cumulative effects of the two transgenes.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the present invention will be apparent to those skilled in the art without departing from the scope and spirit of the present invention. Although the present invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in biochemistry and biotechnology or related fields are intended to be within the scope of the following claims.

CLAIMS

1. A method of reducing heavy metal levels in at least the aerial parts (preferably the leaves) of a plant comprising modifying the plant by increasing the expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said plant.
- 5 2. Use of increased expression of a *gsh1* gene or increased activity of the protein encoded by the *gsh1* gene in a plant for reducing heavy metal levels in at least the aerial parts (preferably the leaves) of said plant.
3. A method or use according to claim 1 or claim 2 wherein the *gsh1* gene is overexpressed in said plant root.
- 10 4. A method or use according to any one of claims 1 to 3 comprising expressing within the plant a polynucleotide (e.g. an exogenous polynucleotide) comprising a nucleic acid sequence encoding a GSH1 polypeptide.
5. A method or use according to claim 4 wherein said polynucleotide (e.g. exogenous polynucleotide) comprises a nucleic acid sequence encoding a GSH1 polypeptide operably linked with a heterologous promoter for directing transcription of said nucleic acid sequence in said plant.
- 15 6. A method or use according to claim 5 wherein said promoter is a root-specific promoter or root-preferred promoter.
7. A method or use according to any one of the preceding claims comprising further modifying the plant by increasing the expression of a *gsh2* gene or the activity of the protein encoded by the *gsh2* gene in said plant.
- 20 8. A method or use according to claim 7 wherein the *gsh2* gene is overexpressed in said plant.
9. A method or use according to claim 7 or claim 8 comprising expressing within the plant a polynucleotide (e.g. an exogenous polynucleotide) comprising a nucleic acid sequence encoding a GSH2 polypeptide.
- 25 10. A method or use according to claim 9 wherein said polynucleotide (e.g. exogenous polynucleotide) comprises a nucleic acid sequence encoding a GSH2 polypeptide operably linked with a heterologous promoter for directing transcription of said nucleic acid sequence in said plant.
- 30 11. A method or use according to claim 10 wherein said promoter is a root-specific promoter or root-preferred promoter.
12. A method or use according to any one of the preceding claims wherein the *gsh1* gene or the nucleic acid sequence encoding said GSH1 polypeptide:
 - 35 a. encodes a GSH1 polypeptide having which has the following conserved residues: n1 metal binding site comprised of Glu-115, Glu-167 and Glu-173;

50

an n2 metal binding site comprised of Glu-113, Gln-254 and Glu-393; or a combination thereof, wherein the numbering of the conserved residues means the effective equivalent position when the GSH1 protein is aligned with the *Arabidopsis thaliana* GSH1 protein sequence having SEQ ID No. 3; or

- b. is an *Arabidopsis gsh1* gene or an orthologue thereof; or
- c. comprises a polynucleotide sequence shown herein as SEQ ID No. 1 or SEQ ID No. 2 or SEQ ID No. 8 or SEQ ID No. 10 or SEQ ID No. 12 or SEQ ID No. 14 or SEQ ID No. 16 or SEQ ID No. 18 or SEQ ID No. 20 or SEQ ID No. 22 or SEQ ID No. 24 or SEQ ID No. 26 or SEQ ID No. 28 or SEQ ID No. 30 or SEQ ID No. 32, or
- d. comprises a functional fragment of the polynucleotide sequence shown in i) which functional fragment encodes a functional GSH1 polypeptide, or
- e. comprises a polynucleotide which encodes a polypeptide comprising the amino acid sequence shown herein as SEQ ID No. 3 or SEQ ID No. 9 or SEQ ID No. 11 or SEQ ID No. 13 or SEQ ID No. 15 or SEQ ID No. 17 or SEQ ID No. 19 or SEQ ID No. 21 or SEQ ID No. 23 or SEQ ID No. 25 or SEQ ID No. 27 or SEQ ID No. 29 or SEQ ID No. 31 or SEQ ID No. 33, or
- f. comprises a polynucleotide sequence which can hybridize to the polynucleotide taught in i), ii) or iii) above under high stringency conditions, or
- g. comprises a polynucleotide sequence which has at least 70% (preferably 85%, more preferably 90%) identity with the polynucleotide shown in i), ii) or iii) above, or
- h. comprises a polynucleotide sequence which differs from polynucleotide shown in i), ii) or iii) due to degeneracy of the genetic code.

13. A method or use according to any one claims 7 to 12 wherein the *gsh2* gene or the nucleic acid sequence encoding said GSH2 polypeptide:

- a. is an *Arabidopsis gsh2* gene or an orthologue thereof; or
- b. comprises a polynucleotide sequence shown herein as SEQ ID No. 4 or SEQ ID No. 5 or SEQ ID No. 34 or SEQ ID No. 36 or SEQ ID No. 38 or SEQ ID No. 40 or SEQ ID No. 42 or SEQ ID No. 44 or SEQ ID No. 46 or SEQ ID No. 48 or SEQ ID No. 50 or SEQ ID No. 52, or
- c. comprises a functional fragment of the polynucleotide sequence shown in i) which functional fragment encodes a functional GSH2 polypeptide, or
- d. comprises a polynucleotide which encodes a polypeptide comprising the amino acid sequence shown herein as SEQ ID No. 6 or SEQ ID No. 35 or SEQ ID No. 37 or SEQ ID No. 39 or SEQ ID No. 41 or SEQ ID No. 43 or SEQ

51

ID No. 45 or SEQ ID No. 47 or SEQ ID No. 49 or SEQ ID No. 51 or SEQ ID No. 53, or

- e. comprises a polynucleotide sequence which can hybridize to the polynucleotide taught in i), ii) or iii) above under high stringency conditions, or
- f. comprises a polynucleotide sequence which has at least 70% (preferably 85%, more preferably 90%) identity with the polynucleotide shown in i), ii) or iii) above, or
- g. comprises a polynucleotide sequence which differs from polynucleotide shown in i), ii) or iii) due to degeneracy of the genetic code.

14. A method or use according to any one of the preceding claims wherein the protein encoded by the *GSH1* gene comprises a polypeptide sequence shown herein as SEQ ID No. 3, or polypeptide sequence which comprises SEQ ID No. 3 with a conservative substitution of at least one of the amino acids, or a polypeptide having at least 70% (preferably 85%, more preferably 90%) identity with SEQ ID No. 3.

15. A method or use according to any one claims 7 to 14 wherein the protein encoded by the *GSH2* gene comprises a polypeptide sequence shown herein as SEQ ID No. 6, or polypeptide sequence which comprises SEQ ID No. 6 with a conservative substitution of at least one of the amino acids, or a polypeptide having at least 70% (preferably 85%, more preferably 90%) identity with SEQ ID No. 6.

16. A construct or vector comprising a *GSH1* gene operably linked with a root-specific promoter or root-preferred promoter.

17. A construct or vector according to claim 16 which further comprises a *GSH2* gene operably linked with the same or a different root-specific promoter or root-preferred promoter.

18. A plant cell (e.g. a tobacco plant cell):

- a. comprising an exogenous *gsh1* gene or an exogenous *gsh1* gene and an exogenous *gsh2* gene;
- b. comprising a construct or vector according to claim 16 or claim 17; and/or
- c. obtainable (e.g. obtained by) a method or use according to any one of claims 1-15.

19. A plant (e.g. a tobacco plant):

- a. comprising an exogenous *gsh1* gene or an exogenous *gsh1* gene and an exogenous *gsh2* gene;
- b. which has been modified to achieve a reduction in heavy metal levels in at least the aerial parts (preferably the leaves) of said plant compared with the

aerial parts (particularly the leaves) of an unmodified plant, where in the modification is an increase in the expression of a *gsh1* gene or the activity of the protein encoded by the *gsh1* gene in said plant;

- c. obtained or obtainable by the method according to any one of claims 1 to 15;
- d. comprising a construct or vector according to claim 16 or claim 17;
- e. comprising a cell according to claim 18.

20. A plant propagation material (e.g. a plant seed) obtainable from a plant according to claim 19.

21. The method or use according to any one of claims 1 to 15, or the plant cell according to claim 18, or the plant according to claim 19, or the plant propagation material according to claim 20 wherein said heavy metal is one selected from the group consisting of: cadmium, arsenic, chromium, copper, lead, zinc, mercury, nickel, selenium and a combination thereof.

22. The method or use according to any one of claims 1 to 15 or 21, or the plant cell according to claim 18 or claim 21, or the plant according to claim 19 or claim 21, or the plant propagation material according to claim 20 or claim 21 wherein the heavy metal reduction is a value of at least about 30% w/w, preferably at least about 50%, more preferably at least about 60%, more preferably at least about 70%, more preferably at least about 80%, most preferably at least about 90% or 95% w/w.

23. The method or use according to any one of claims 1 to 15 or 21 to 22, or the plant cell according to any one of claims 18, 21 or 22, or the plant according to any one of claims 19, 21 or 22, or the plant propagation material according to any one of claims 20-22 wherein said heavy metal reduction is a value of at least about 60% w/w.

24. The method or use according to any one of claims 1 to 15 or 21 to 23, or the plant cell according to any one of claims 18 or 21 to 23, or the plant according to any one of claims 19 or 21 to 23, or the plant propagation material according to any one of claims 20-23 wherein the plant is a plant having consumable aerial parts, e.g. consumable leaves.

25. The method or use according to any one of claims 1 to 15 or 21 to 24, or the plant cell according to any one of claims 18 or 21 to 24, or the plant according to any one of claims 19 or 21 to 24, or the plant propagation material according to any one of claims 20-23 wherein the plant is a plant selected from the group consisting of a plant from the family *Solanaceae*, a salad leaf crop and a leaf vegetable.

26. The method or use according to claim 25, or the plant cell according to claim 25, or the plant according to claim 25, or the plant propagation material according to claim 25 where in the plant from the family *Solanaceae* is from the genus *Nicotiana*.

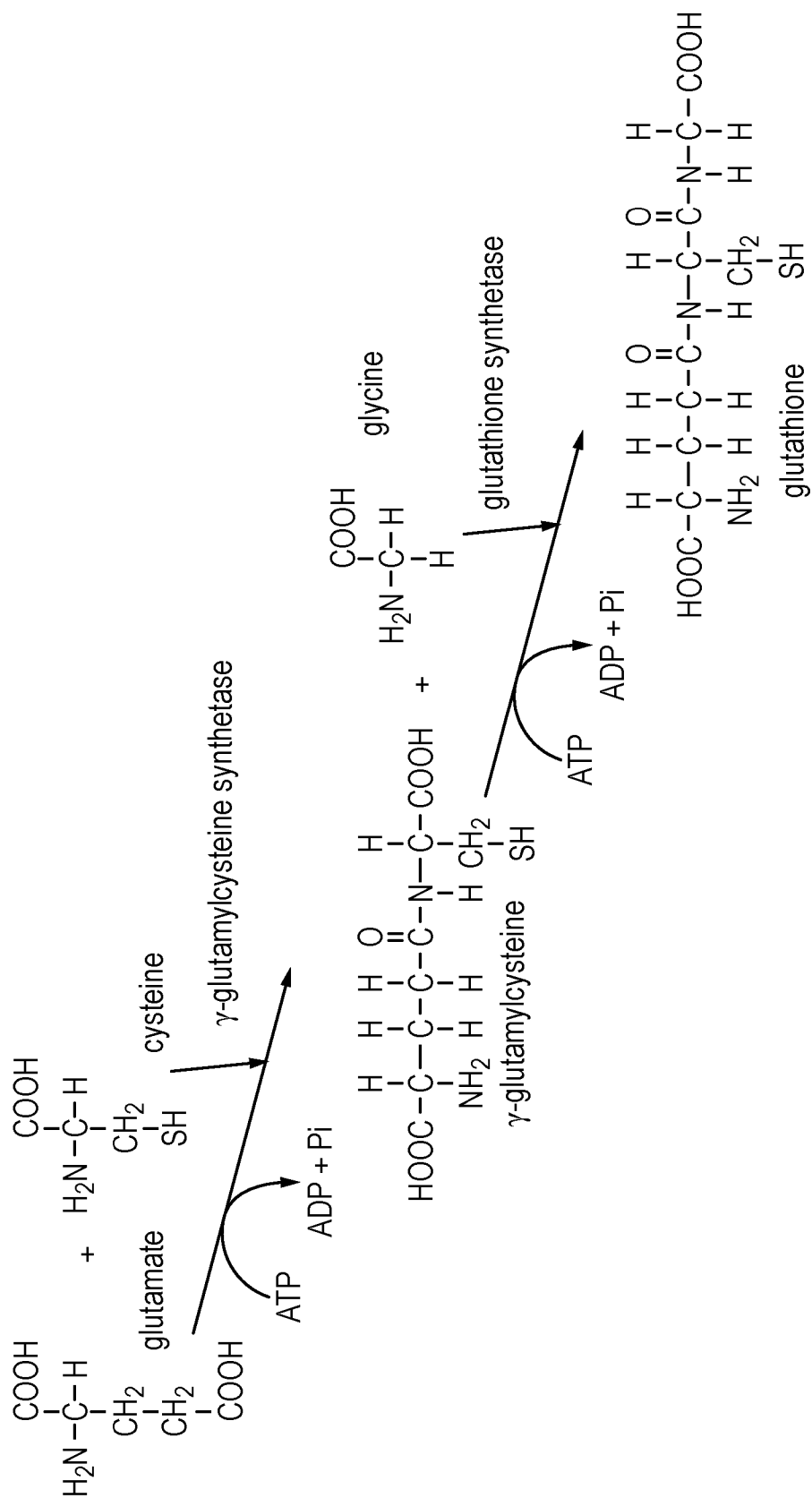
27. The method or use according to claim 25, or the plant cell according to claim 25, or the plant according to claim 25, or the plant propagation material according to claim 25 wherein the salad leaf crop is a crop selected from the group consisting of a lettuce (including for example baby leaf spinach, Cos, Frisée, Iceberg lettuce, Lamb's lettuce, Little Gem, Mizuna, Radicchio, Red mustard, Red oak leaf, Rocket (Roquette), Ruby chard, Sweet Romaine), watercress and cress.
28. The method or use according to claim 25, or the plant cell according to claim 25, or the plant according to claim 25, or the plant propagation material according to claim 25 wherein the leaf vegetable is a plant selected from the group consisting of a cabbage, spinach, kale, tea, chicory (or curly endive), Phak chet, or a herb (such as basil, mint or oregano).
29. The method or use according to claim 26, or the plant cell according to claim 26, or the plant according to claim 26, or the plant propagation material according to claim 26 wherein the plant is from the species *Nicotiana tabacum* or *Nicotiana rustica*.
30. Use of a plant cell (e.g. a tobacco cell) according to any one of claims 18 or 21-29 for the production of a consumable plant product (e.g. a tobacco product).
31. Use of a plant (e.g. a tobacco plant) according to any one of claims 19 or 21-29 to breed a plant (e.g. a tobacco plant).
32. Use of plant (e.g. a tobacco plant) according to any one of claims 19 or 21-29 for production of a consumable plant product (e.g. a tobacco product).
33. Use of plant (e.g. a tobacco plant) according to any one of claims 19 or 21-29 to grow a crop.
34. Use of plant (e.g. a tobacco plant) according to any one of claims 19 or 21-29 to produce a consumable leaf (e.g. a processed (preferably cured) leaf).
35. A harvested leaf of a plant (e.g. a tobacco plant) according to any one of claims 19 or 21-29 or obtainable from a plant (e.g. a tobacco plant) propagated from a propagation material according to any one of claims 20-29 or obtainable from a plant (e.g. a tobacco plant) obtainable by a use according to any one of claims 31 to 33.
36. A harvested leaf of a plant (e.g. a tobacco plant) according to claim 35 wherein the harvested leaf is a cut harvested leaf.
37. A processed leaf (e.g. a processed tobacco leaf, preferably a non-viable processed tobacco leaf):
- comprising a plant cell according to any one of claims 18 or 21-29;
 - obtainable from a plant obtainable from a use according to any one of claims 31-33;

- c. obtainable from processing a tobacco plant according to any one of claims 19 or 21-29;
 - d. obtainable from a plant (e.g. tobacco plant) propagated from a plant propagation material according to any one of claims 20-29;
 - 5 e. obtainable by processing a harvested leaf according to claim 35 or claim 36.
38. The processed leaf (e.g. processed tobacco leaf) according to claim 37, wherein the plant or leaf is processed by curing, fermentation, pasteurising or combinations thereof.
39. The processed leaf (e.g. processed tobacco leaf) according to claim 37 or claim 38
- 10 wherein the processed tobacco leaf is cut processed tobacco leaf.
40. A consumable plant product:
- a. prepared from the aerial parts of the plant (preferably the leaves) harvested from the plant obtained or obtainable by the method according to any one of claims 1 to 15;
 - 15 b. prepared from the aerial parts of the plant (preferably the leaves) harvested from the plant according to any one of claims 19 or 21 to 29;
 - c. prepared from the aerial parts of the plant (preferably the leaves) propagated from a plant propagation material according to any one of claims 20-29;
 - d. prepared from a harvested leaf according to any one of claims 35 or 36;
 - 20 e. prepared from a processed leaf according to any one of claims 37 to 39;
 - f. prepared from or comprising a plant extract obtained from a modified plant according to the present invention.
41. A consumable plant product according to claim 40 wherein the consumable plant product is a tobacco product.
- 25 42. A consumable plant product according to claim 40 or claim 41 wherein the tobacco product is a smoking article.
43. A consumable plant product according to claim 40 or claim 41 wherein the tobacco product is a smokeless tobacco product.
44. A consumable plant product according to claim 40 or claim 41 wherein the tobacco
- 30 product is a tobacco heating device, e.g. an aerosol-generating device.
45. A plant extract (e.g. tobacco extract) of said plant according to any one of claim 19 or 21-29 or of a portion of said plant.
46. A smoking article, smokeless tobacco product or tobacco heating device comprising a
- 35 plant or a portion thereof from the species *Nicotiana tabacum* or *Nicotiana rustica* according to claim 29 or an extract (e.g. a tobacco extract) thereof.

55

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

1/37



Where: γ -glutamylcystine synthetase (γ -ECS) = glutamate-cysteine ligase (GSH1)

FIG. 1

2/37

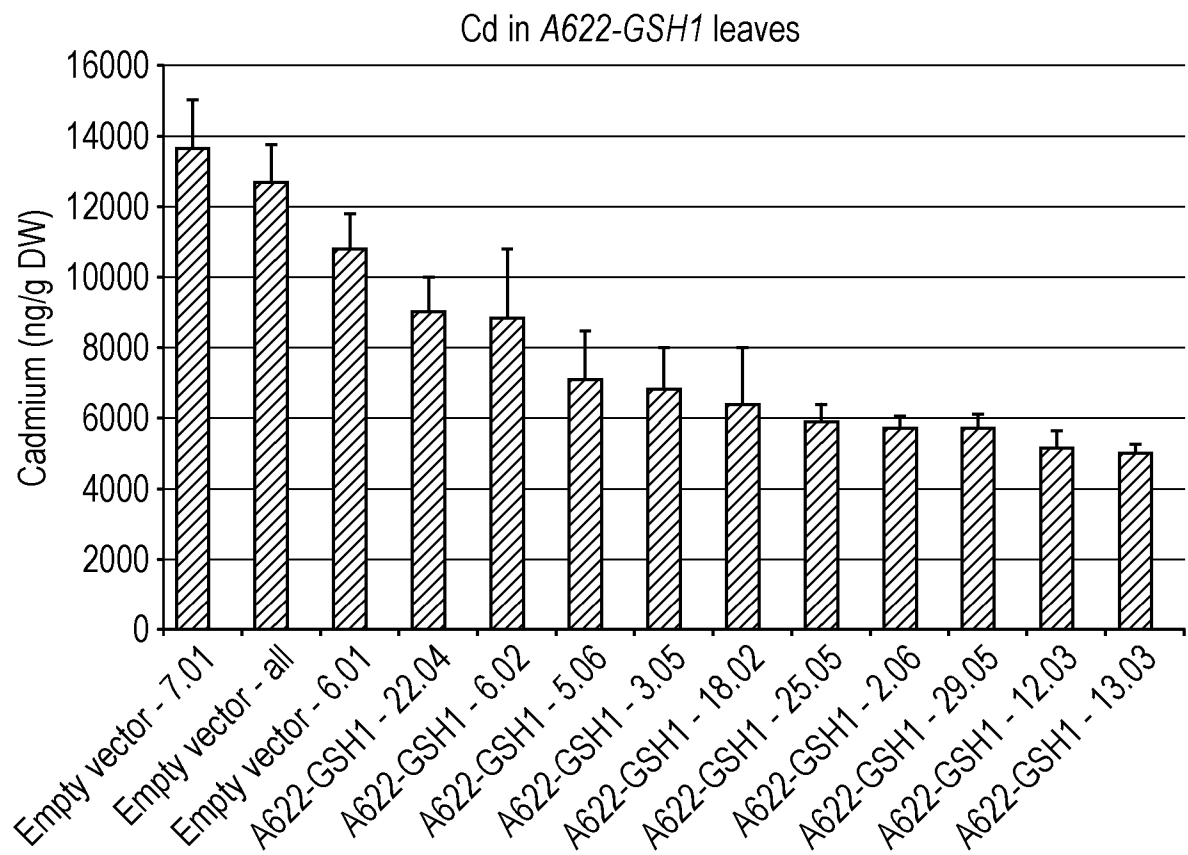


FIG. 2

3/37

(SEQ ID No. 1)

>*GSH1*_gDNA_Arabidopsis thaliana (At4g23100.1)_(introns in lower case *italics*)

ATGGCGCTCTTGTCTCAAGCAGGAGGATCATACACTGTTGTTCTCTGGAGTTTGTTCAAAGGCTGGAACATAAA
GCTGTTGTTTCGGGTGGCGTGAGGAATTTGGATGTTTTGAGGATGAAAGAAGCTTTTGGTAGCTCCTACTCTAGG
AGTCTATCTACCAAATCAATGCTTCTCCATTCTGTTAAGAGGAGTAAGAGAGGGGCATCAATTGATTGTTGCGGCA
AGTCCTCCAACGGAAGAGGCTGTAGTTGCAACTGAGCCGTTGACGAGAGAGGATCTCATTGCCTATCTTGCCTCT
GGATGCAAAACAAAGGACAAATATAGgtacagtttgatgcttataatgttctgttataaatgctacttagtggttgt
aattgtgacataatttctttgtttgtgctcttttacggtgttggtattgtttgaattcatccagattgtttgtcta
ttggagaaatattctgttttgctaacttaacattgtgagtttccagAATAGGTACAGAACATGAGAAATTTGGTT
TTGAGGTCAATACTTTGCGCCCTATGAAGTATGATCAAATAGCCGAGCTTCTTAATGGTATCGCTGAAAGATTTG
AATGGGAAAAAGTAATGGAAGGTGACAAGATCATTTGGTCTGAAGCAGgttaacattctcctgtgtttcatctttt
gagccgtttattatattatatacatatttccagcctgtagaacatgttagcttcatatacaacacacaaatgatcggta
ttataaggtttcagtcctttttttttttttgtctgtagataaactcgttctgtcctttgttgaccagGGAAAGCAA
AGCATTTCACTTGAACCTGGGGTTCAGTTCGAGCTTAGTGGTGCACCTCTTGAGACTTTGCATCAAACCTTGTGCT
GAAGTCAATTACATCTTTATCAGgttaatttcaaagataattattgccaatgctctcaccctaaaaacgtgacaaa
taaccaactatctacggatgtcacagGTAAGCAGTTGCTGAGGAAATGGGAATTGGTTTCTTAGGAATTGGCT
TCCAGCCCAAATGGCGTCGGGAGGATATACCCATCATGCCAAAGgtatgacgctttctctattcttaagattgaa
gaaagttaacatctgtttgttttctaattcagataatattatctgattctacatagccaacggttcggattattct
tggtgtttgtgttaagttttcctatgacgagtcaggcagtaagatttgacatagtgtaatgtgttctgtgcagGG
GAGATACGACATTATGAGAACTACATGCCGAAAGTTGGTACCCTTGGTCTTGATATGATGCTCCGAACGTGTAC
TGTTCAAGgttagattaatattctgactttgagccttttcataatttcccttaaaactaaactttccagtttctcatca
ataatcttattttgaaacttatagaaactgcattacttgcaatcttgacagGTTAATCTGGATTTTAGCTCAGAA
GCTGATATGATCAGGAAGTTTCGTGCTGGTCTTGCTTTACAACCTgtgagaagctcgatctctttgtggcataac
cagaaacttttattaacgattttgactaaccattgcctcactattatctaaattcaacagATAGCAACGGCTCTA
TTTGCGAATTCCCCTTTTACAGAAGGAAAGCCAAACGGATTTCTCAGCATGAGAAGgtatctagattcagttggt
atgatccccgatattctttttgaattagtttccctcccaccttataactctggatgttttgacagCCACATATGG
ACAGACACTGACAAGGACCGCACAGGAATGCTACCATTTGTTTTCGATGACTCTTTTGGgtgagtatcgatgttt
aatctatgaaaaacaattacattacatttggttggtttggtatgagaagcttctcaaaaaccaatgactctgtctc
gcattcttgatgagttatttgaaaattgatattgtgctattctgtggattttcttacatatgtttgtgttcaagat
agaaagatgctttggatctttatccttctcttgaaacagcatggctaattcttaattttattgctctctttgttc
ttatctggtatagTTTTGAGCAGTATGTTGACTACGCACTCGATGTCCCTATGTACTTTGCCTACAGAAAGAACA
AATACATCGACTGTACTGGAATGACATTTCCGGgtcaaaatctatccttaattctttctgttaggttctataactg
tttcattgcaaggaattggcaaattgaaaaccatgggatcttcttgaaattataatctaatctcataataattcttca
tgctgcagCAATCTTGGCTGGAAAACCTCCCTGTCTCCCTGGTGAACCTGCCTTCATATAATGATTGGGAAAACC
ATCTGACAACAATATTCCCAGAGgttaattcatctgcctcaagagtccttcaatcagcctatgactttcatttccc
tgtcttatgttgaaagtctgacaacggttcataacactagGTTCCGTTGAAGAGATACTTGGAGATGAGAGGTGCT
GATGGAGGTCCCTGGAGGAGGCTGTGTGCCCTGCCAGCTTTCTGGgttaactacatctatttagcacatatataatt
cgtgcttaattgagaatgtttcaaaaagtattgactgtttgttttgattttatctacaaatttcagGTGGGTTTATT
ATATGATGATGATAGTCTCCAAGCTATCCTGGATCTGACAGCTGACTGGACTCCAGCAGAGAGAGAGATGCTAAG
GAACAAAgtaataacaagacgcaccacattttacattttcttattccacagttagaaaacatcttgtaataaacta
atacttgattttccaacagGTCCAGTTACTGGCTTAAAGACTCCTTTTAGGGATGGTTTGTAAAGCATGTGCT
GAAGATGTCCTGAAACTCGCAAAGgtgtgtttgtatgaaccctcttagaaacttcagaccagctttctatcccgt
ctggtccagcttctaactattggacttattttctctttgtttctcagGATGGTTTAGAGCGCAGAGGCTACAAGG
AAGCCGGTTTCTTGAACGCAGTCGATGAAGTGGTCAAGACAGgttaattgaagtttccctaatgcataattgaattc
cacaccgacattttgtttgataaacctggtggtgaaatttggttttgaaatgacagGAGTTACGCCTGCGGAGAAG
CTCTTGGAGATGTACAATGGAGAATGGGGACAAAGCGTAGATCCCGTGTTCGAAGAGCTGCTGTACTAAgaaaat
gggacgtgaacaaaagggtgtctataaacctttgggtgtgagtttatgtatctgaagaattcgagtctccggaat
aaggattttttttttgttgtaatcgatttttaaaaactgattttgttttagaaattcgaagcattgaaaagcag
aagaaaaattgtatgtactaaacgattttcgggtgtgggaaatcgtttgggagggtgtgtttggatcttgaataaat
taccatttttctgtcacaaatttgcctacatttaacgaaataaactcaaaactgatttcaaggcattctcatag
cttatgatgaattgttataattgcag

FIG. 3

4/37

(SEQ ID No. 2)

>*GSH1_CDS_Arabidopsis thaliana* (Arath)

ATGGCGCTCTTGTCTCAAGCAGGAGGATCATACACTGTTGTTCTTCTGGAGTTTGTTCAAAGGCTGGAACATAA
GCTGTTGTTTCGGGTGGCGTGAGGAATTTGGATGTTTTGAGGATGAAAGAAGCTTTTGGTAGCTCCTACTCTAGG
AGTCTATCTACCAAATCAATGCTTCTCCATTCTGTTAAGAGGAGTAAGAGAGGGCATCAATTGATTGTTGCGGCA
AGTCCTCCAACGGAAGAGGCTGTAGTTGCAACTGAGCCGTTGACGAGAGAGGATCTCATTGCCTATCTTGCCCTCT
GGATGCAAAACAAAGGACAAATATAGAATAGGTACAGAACATGAGAAATTTGGTTTTGAGGTCAATACTTTGCGC
CCTATGAAGTATGATCAAATAGCCGAGCTTCTTAATGGTATCGCTGAAAGATTTGAATGGGAAAAAGTAATGGAA
GGTGACAAGATCATTGGTCTGAAGCAGGGAAAAGCAAAGCATTTCACTTGAACCTGGGGGTCAAGTTCGAGCTTAGT
GGTGACCTCTTGAGACTTTGCATCAAACCTTGTGCTGAAGTCAATTCACATCTTTATCAGGTAAAAGCAGTTGCT
GAGGAAATGGGAATTGGTTTCTTAGGAATTGGCTTCCAGCCCAAATGGCGTCGGGAGGATATACCCATCATGCCA
AAGGGGAGATACGACATTATGAGAACTACATGCCGAAAGTTGGTACCCTTGGTCTTGATATGATGCTCCGAACG
TGTA CTGTT CAGGTTAATCTGGATTTTAGCTCAGAAGCTGATATGATCAGGAAGTTTCGTGCTGGTCTTGCTTTA
CAACCTATAGCAACGGCTCTATTTGCGAATTCCCCTTTTACAGAAGGAAAGCCAAACGGATTTCTCAGCATGAGA
AGCCACATATGGACAGACACTGACAAGGACCGCACAGGAATGCTACCATTGTTTTCGATGACTCTTTTGGGTTT
GAGCAGTATGTTGACTACGCACTCGATGTCCCTATGTA CTTTGCCTACAGAAAGAACAAATACATCGACTGTACT
GGAATGACATTTTCGGCAATTCTTGGCTGGAAAACCTCCCTGTCTCCCTGGTGAAGTGCCTTCATATAATGATTGG
GAAAACCATCTGACAACAATATTTCCAGAGGTTTCGGTTGAAGAGATACTTGGAGATGAGAGGTGCTGATGGAGGT
CCCTGGAGGAGGCTGTGTGCCCTGCCAGCTTTCTGGGTGGGTTTATTATATGATGATGATAGTCTCCAAGCTATC
CTGGATCTGACAGCTGACTGGACTCCAGCAGAGAGAGAGATGCTAAGGAACAAAGTCCCAGTTACTGGCTTAAAG
ACTCCTTTTAGGGATGGTTTGTAAAGCATGTGCTGAAAGATGTCCTGAAACTCGCAAAGGATGGTTTAGAGCGC
AGAGGCTACAAGGAAGCCGGTTTCTTGAACGCAGTCGATGAAGTGGTCAGAACAGGAGTTACGCCTGCGGAGAAG
CTCTTGGAGATGTACAATGGAGAATGGGGACAAAGCGTAGATCCCCTGTTTCGAAGAGCTGCTGTACTAA

FIG. 4

5/37

(SEQ ID No. 3)

*GSH1*_protein (translated from CDS)_Arath

MALLSQAGGSYTVVPSGVCCKAGTKAVVSGGVRNLDVLRMKEAFGSSYSRSLSTKSMLLHVKRSKRGHQLIVAA
SPPTEEAVVATEPLTREDLIAYLASGCKTKDKYRIGTEHEKFGFEVNTLRPMKYDQIAELLNGIAERFEWEKVME
GDKIIIGLKQGKQSSISLEPGGQFELSGAPLETLHQTCAEVNSHLYQVKAVAEEMGIGFLGIGFQPKWRREDIPIMP
KGRYDIMRNYMPKVGTGLDMMMLRTCTVQVNLDIFSSEADMIRKFRAGLALQPIATALFANSPTTEGKPNGFLSMR
SHIWTDTDKDRTGMLPFVFDSDSGFEQYVDYALDVPMYFAYRKNKYIDCTGMTFRQFLAGKLPCLPGELPSYNDW
ENHLTTIFFEVRLKRYLEMARGADGGPWRRLCALPAFWVGLLYDDDSLQAILDLTADWTPAEREMLRNKVPVTGLK
TPFRDGLLKHVAEDVLKLAKDGLERRGYKEAGFLNAVDEVVRTGVTPAEKLLEMYNGEWQSVDPVFEELLY*

FIG. 5

6/37

>*GSH2_gDNA_Arath*(At5g27380.1)_(introns in lower case *italics*) (SEQ ID No. 4)

ATGGGCAGTGGCTGCTCTTCTCTCTTACTCATCATCTTCCACTTGCAATGCCACTGTTTTCTCTATTTCTTCT
TCTCCATCATCTTCTTCTTCGCTGAAACTCAATCCAAGTAGCTTCCTTTTCCAGAACCCTAAAACCTCTGAGAAAT
CAGTCACCTTTGAGGTGTGGGAGATCATTTAAGATGGAATCACAGAAACCCATTTTCGATTTGGAGAAATTAGAT
GATGAGTTTGTTGAGAACTGGTTTACGATGCTCTCGTTTGGTCTTCTCTTCACGGACTCGTCGTTGGTGACAAA
AGTTATCAGgtttacttacttgcccaccactaaagtccctgtcttttgatcttgtgttctgataaagttagagaa
aaattcgaaaatttattatctttatgaaccttgatatagctgaatcatcattttcttcgtcctggattgtgttagtt
tcggtgtcttatagctttcgatgccatgtttgatctgaatccaaagcttgggtgttcttttgattgggtttgat
gattacagAAATCAGGAAATGTTCCAGGGGTGGATTGATGCACGCACCTATTGCATTGCTACCAACTGCATTCC
CAGAAGCTTATTGGAAGCAAGCTTGAATGTTACTCCTCTTTTCAATGAATTGATTGATCGTGTAGCTTGATG
GCAAATCTTACAGGATAGTCTCTCTAGgttcgtatcgcaatctatctgctttggattatatcatctaataacaa
gcgtaatgatcttttgattgattttctatgttcttagAACGAAGAAAGTTGATGTCTTTACATCTAGACTTCTTGA
CATTCACTCCAAGATGCTAGAAAGAAATAAGAAAGAGgtctttttacattatctctggtttcatgtttatcttga
ttctttgggttctggatgttgttcatgttgggtgtgttgtgatcgagGACATTCGTTTGGGTTTACACCGGTTTG
ATTATATGCTTGACGAAGAAACAAATCACTTCTTCAGATTGAGATGAATACTATCTCGTGTTCTGTTTCTGGCC
TTAGTCGTCTGTAGCCAGCTACATCAGtatgtaacagtgcagaccatcttctgcattacctgttgtgtttt
tttatacagcggaattttattccacttgtttgagGTCATTGCTTCGATCTTATGGGGATCAGATTGGCATAGAC
TCTGAACGTGTACCTATAAACACATCCACAATCCAATTTGCTGATGCATTGGCTAAAGCTTGGTTGGAGTACAGT
AACCCAAAGgttagcctggaacacaaatcaagaatcagtgatgtaatttaaagctttttgtgtttaaagctagagt
ttttcttggtttcagAGCGGTAGTCATGGTAATTGTACAGCCAGAAGAACGCAACATGTACGATCAACATTTGCT
GAGCAGTATATTGAGAGAAAAatatcctcttttgcttctgtttctttgggtttccaggggtcgatcttgacttctta
tgctataattttgtaccacaaaaaatggtgttgttcttcttagcttcagaacacGCATAATATAGTTGTCATCAG
GAAGACTCTAGCAGAAAGTCGAAAAAGAGGGAGGTACAAAGAGGATGAAACCCTTATTGTgtaagcatttgtttg
cattgtgggtttttgagaacttacaccttcatcttgaaagctatgttaagagttttcattgttgcagTGGCGGCC
AAGCAGTCGCAGTGGTTTATTTTCAGATCTGGCTATACTCCTAATGATCATCCGTCTGAATCagtaagttgccgtc
aagtttccaatattgtggctttactctttgggaaaatctgaatcatatcattgggtttatttgtttcaatttgcag
GAATGGAATGCTAGGCTGCTTATTGAGGAGTCTTCAGCTGTCAAATGCCGAGCATAGCTTACCATTTAAGTGGC
TCCAAGAAAAATCCAGCAAGAACTGGCTAAACCAGGTGTTCTCGAGAGgtaaatggaggctgatactgaagaccgt
ttccaaatatttcaaaggcgaaaacttttttaacttaaaacatgagaaacagGTTTCTGGACAACAAAGAGGACA
TTGCTAAGCTGAGGAAATGCTTTGCTGGGCTTTGGAGCTTGATGACTCAGAAATTGTCAAGCAGGCTATCGAAA
AACCCGGATTGTTTGTATGAAGCCTCAGAGAGAAGGCGGAGgttagaaccctaattgagtcacccaatatcttca
acttttcgggtcatatttctaagcctgtcttcttggttcttccagGAAACAACATCTATGGAGATGATGTGAGGG
AAAATCTTTTGAGACTGCAGAAAGAAGGAGAGGAAGGAAACGCTGCGTATATCCTGATGCAGAGGATATTCCCAA
AAGTCTCAAACATGTTCTTGGTGCGAGAAGGCGTTTACCATAAGCATCAAGCTATATCAGAACTCGGTGTCTATG
GTGCTTACCTCAGgtacagagcagatagatgaagaatttgggtatgcttcaataattctcaacttaatttgtgcc
ataatctttgattaaatgaaatatgtagGAGCAAAGACGAAGTTATAGTAAACGAGCAGAGTGGTTATCTAATGC
GCACGAAGATCGCATCATCAGATGAAGGCGGCGTTGCAGCTGGTTTTGGAGTCTTGGACAGCATATATCTGATTT
GAGatggataacattcacaacctcttgatcagttttcttctggtttattccttcttcttcttcttcttcttcttctt
taaagatttgtgtattaatcatgaagatcaataagactaggcaaagctctttacaattctttccacaggaccgca
aagcctacattgcttattgttatttttagaccgtgtgagatgattactcgtttataaagatcatgaattgaataa
gataagatcaatcatgg

FIG. 6

7/37

>*GSH2_CDS_Arath* (SEQ ID No. 5)

ATGGGCAGTGGCTGCTCTTCTCTCTCTTACTCATCATCTTCCACTTGCAATGCCACTGTTTTCTCTATTTCTTCT
TCTCCATCATCTTCTTCTTCGCTGAACTCAATCCAAGTAGCTTCCTTTTCCAGAACCCTAAAACCTCTGAGAAAT
CAGTCACCTTTGAGGTGTGGGAGATCATTTAAGATGGAATCACAGAAACCCATTTTCGATTTGGAGAAATTAGAT
GATGAGTTTGTTCAGAACTGGTTTACGATGCTCTCGTTTGGTCTTCTCTTCACGGACTCGTCGTTGGTGACAAA
AGTTATCAGAAATCAGGAAATGTTCCAGGGGTGGATTGATGCACGCACCTATTGCATTGCTACCAACTGCATTC
CCAGAAGCTTATTGGAAGCAAGCTTGTAATGTTACTCCTCTTTTCAATGAATTGATTGATCGTGTTAGCTTGGAT
GGCAAATTCTTACAGGATAGTCTCTCTAGAACGAAGAAAGTTGATGTCTTTACATCTAGACTTCTTGACATTAC
TCCAAGATGCTAGAAAGAAATAAGAAAGAGGACATTCTGTTTGGGTTTACACCGGTTTGATTATATGCTTGACGAA
GAAACAAATTCACCTCTTCAGATTGAGATGAATACTATCTCGTGTTCTGTTTCTGGCCTTAGTCGTCTTGTTAGC
CAGCTACATCAGTCATTGCTTCGATCTTATGGGGATCAGATTGGCATAGACTCTGAACGTGTACCTATAAACACA
TCCACAATCCAATTTGCTGATGCATTGGCTAAAGCTTGGTTGGAGTACAGTAACCCAAGAGCGGTAGTCATGGTA
ATTGTACAGCCAGAAGAACGCAACATGTACGATCAACATTTGCTGAGCAGTATATTGAGAGAAAAGCATAATATA
GTTGTCATCAGGAAGACTCTAGCAGAAGTCGAAAAAGAAGGGAGTGTACAAGAGGATGAAACCCCTTATTGTTGGC
GGCCAAGCAGTCGCAGTGGTTTATTTTCAGATCTGGCTATACTCCTAATGATCATCCGTCTGAATCAGAATGGAAT
GCTAGGCTGCTTATTGAGGAGTCTTCAGCTGTCAAATGCCCCGAGCATAGCTTACCATTTAACTGGCTCCAAGAAA
ATCCAGCAAGAACTGGCTAAACCAGGTGTTCTCGAGAGGTTTCTGGACAACAAAGAGGACATTGCTAAGCTGAGG
AAATGCTTTGCTGGGCTTTGGAGCTTGATGACTCAGAAATTGTCAAGCAGGCTATCGAAAAACCCGGATTGTTT
GTTATGAAGCCTCAGAGAGAAGGCGGAGGAAACAACATCTATGGAGATGATGTGAGGGAAAATCTTTTGAGACTG
CAGAAAGAAGGAGAGGAAGGAAACGCTGCGTATATCCTGATGCAGAGGATATTTCCAAAAGTCTCAAACATGTTT
TTGGTGCGAGAAGGCGTTTACCATAAGCATCAAGCTATATCAGAACTCGGTGTCTATGGTGCTTACCTCAGGAGC
AAAGACGAAGTTATAGTAAACGAGCAGAGTGTTATCTAATGCGCACGAAGATCGCATCATCAGATGAAGGCGGC
GTTGCAGCTGGTTTTGGAGTCTTGACAGCATATATCTGATTTGA

FIG. 7

8/37

>*GSH2*_protein (translated from CDS)_*Arabidopsis thaliana*

MSGCSSLSYSSSSSTCNATVFSISSSSPSSSSSSSLKLNPSFLFQNPKTLRNQSPLCGRSFKMESQKPIFDLEKLD
DEFVQKLVDALVWSSLHGLVVGDKSYQKSGNVPGVGLMHAPIALLPTAFPEAYWKQACNVTPLFNELIDRVSLD
GKFLQDSLRSRTKKVDVFTSRLLDIHSKMLERNKKEDIRLGLHRFDYMLDEETNSLLQIEMNTISCSFPGLSRLVS
QLHQSLLSYGDQIGIDSERVPINTSTIQFADALAKAWLEYSNPRAVVMVIVQPEERNMYDQHLLSSILREKHNI
VVIRKTLAEVEKEGSSVQEDELIVGGQAVAVVYFRSGYTPNDHPSESEWNARLLIEESSAVKCPSTIAYHLTGSKK
IQQELAKPGVLERFLDNKEDIAKLKCFAGLWSLDDSEIVKQAI EKPGLFVMKPQREGGGNNIYGDDVRENLLRL
QKEGEEGNAAAYILMQRIFPKVSNMFLVREGVYHKHQAI SELGVYGAYLRSKDEVIVNEQSGYLMRTKIASSEGG
VAAGFGVLD SIYLI*

FIG. 8

9/37

>A622_promoter (SEQ ID No. 7)

t c g a g g a g a t g c a a t g c t g a t a a a g g g a t g a a a a g c a a g g g a t a t t t g t a t t a g c a t a g a a c a c a g a a g a a g
c a g t a a g g c a g a t g c a t t a g c c g a t t a g a t t a t c a a t a g c c t t a a t a t a t g t a t a a t t c g a t c c g a a a a t t c a a
a a t a t g t a t c c a a t c c a a c t t t a a a a t t c g a t c a g a t c c g a t a t t a t t t c g g a t c g g a t t c g a a t t g c a t t t t g
t a g a a t c c g a a a t t c a a a a t t t t a a t c t g a a a t g t g c t a a a t c c g a t c c g a t t t g a t c c a a g c a c a c c c t a a t t
g t a a g t a t c t g a a a c t g a t t a t t t c a t t g a t t t g c t t c a a a a a c a a a g t a t t a g c a g a a a a t g a t a t t t t t a t a
t a g a a g t t a t g t t a a a g c t t t t a a a a g c c a c c g t t c c t a a t g c t t t t a t t a a
t t t c g c t g c c a c t c a a t t t c a g t t c t a g t c c c a g a t t t t c t c a c a t g g c c c t a t t t t c a t a a t g c a t g a a a t a a
g t g t t a t t a t a t g t a a c a c a a t c a c t a t g t t a t t a t t a a g t t t a g g t a g g t a c c a t g c c t c c t c a g t a a t g t t c c
t t g c a a t t g a t t c c t c c g g c t a a a t c t a g a t c t c t c t t a t g t t t c a t g c a c g t g g a a a a a a g a a g t t t g g t t t
t t t t t g g t t c t t a t t g c t a t t t a a a t g t g g t a g g t g a g a a c t g g a t t a a a t a a a a c a a a c c g t a t t c a a c t
t t t c c g t c c t g t a t t a t a t t t a t t a t t a c a g t a a a t a g c a c a a a a t a t t a a a a g a a a a t g a g a a t g a g g a t c t t
t t c a a a a t t a a a t t t c t t t a a a a t a a c a g a a t g t t g t t c c c t a c a g g c t a c a g c t a c a c g t a a g a a a g t a c a a
c a a g t c t t t t c t t t t g t c c t t a t c g a t a t c t g a a c g c a a t a g g t g g g a a t a t t g g a t c g a a a a t t a a c c c a c t
a t t g c c t g a a a g a t a a a a a t a t a a a a t t c t c t a a t a a t t a g t t c t t t a a t a c a a c a c a t g a c a a a t g t a a t t
c t t t a t a t t a t g a c a a g a t t c g a t t c c t t a t a c t a g c a c c a c t t t t g g c a g c a a g a a a t c a a t a t t t a a t a t t a a a
a c a a a a t a t g a c a c a g c a t t a t t t t t c t a t a t t a t g g c a a t a t a t a a t t c t t a a a a t t a g t t c t t t a a t a c a a c
a a t a t a t g a c a a a a t g t a a t t c t t t a t a t a a t g a c a a g a t t c g a t t t t t a g c a c c a c t t t t g g c a g c a a g a g a t a c
a t a t t t a a t a t t a a a a a a t a t g a a a a c a a t t a t a t t t t a t a t t a t g g c a a t a t a t a a t t c t t t a t a t t a g c a c
c a c t t t c c a g c a a g a t a a a a t a t t t a a a a t c a a t t t c t a c t t t g a a c c a c t a t c c c c t g c t c t a a a a t a a t t t g t
t t a t t t c a g g g g a t t g g a a a t g a c t t t a a g t t a t a t a t t g t t a g t g t a a a c g a a t t t g a t c a c c t a a a g a a t a
t t t a c a a g t a a g c c t c t a t a a a a g t a a t t a a g g t t t g t a a t c t t a a a g a g a g c a t a t t a c a t g t t g t a a t t g g
t a a a t g t g a c c t g a t a g t g t a c a a a a t a a t t t a a g c t a a c c g t g t a a c t a a t t a a a t t t a t a t a t a t a t a t a t
a t a c c t t t t c t t t c a t t a a a t a a t g t a a t
t t c c c g a g g a a g g a t t c a a c c c t c c c t t a t a t g c g g c t c g c t a t a t a a t a t t t t t t t t t a a t t a t t t t c c t a c a t g
c t c a t c t a g t a g g t a a t g a c a c a g c t a g a c g c a t g g a t c a a c a t t c t a a g a c a g a t c a t c a a a a t t t a t t a a t g
g t a t g a t a t c g t a a t t c c a t g t g t t a g g t t a a t t g t g c a a a a t t c c c t a g c t t g a a t t t g c c t a t a a g t a c c a g c
g t t a a t a t c a g a t t c t c c a c a g c a t t t c a t a c a a a a t c c g a t t t a a t t c c t a g t t t c t a g c c c c t c c a c c t t a a
c c c g a a g c t a c t t t t t t t t t c t t c c c c t a g a a g g c c t a g c g g t c g a c

FIG. 9

10/37

SEQ ID No. 8

>XM_002867684.1__Arabidopsis_lyrata_subsp._lyrata
(Araly)_hypothetical_protein,_mRNA

AGCTCAACTGATCGGCTAATGCTCTCACTCTTTCTTTTTTATCAATTTTCTCATTTCTCTCAGTCGCATATAAAT
TGAAACCGCCGATTCGTTTTTGGCTCTTAAGCTTTCTTCTTTATTTTCGACGGTGTCTCTCAATCTCCGTCAAG
GTTGACGAATTGCAGGAGCTATATAGACCATGGCGCTCTTGTCAACAGCAGGAGGATCATACACTGTTGTCCCTT
TGGGATGTTCAAAGACTGGAATAAATCAGTTTCTGGTGGTGTGAGGAATTTGGATGTGTTGAGGATGAAAGAAG
CTTTTGTTAGCTCCTATTCTAGGAGTCTATCTACAAAATCAATGCTTCTCCGTTCTGTTAAGAGGAGCAAGAGAG
AGCATCAGTTGATTGTTGCGGCAAGTCCGCCAACGGAAGAGGCTGTCGTTGCAACTGAGCCGTTGACGAGAGAGG
ATCTTCATTGCGTATCTTGCCCTCCGGATGCAAACTAAGGACAAATATAGAATAGGTACAGAACATGAGAAATTTG
GTTTTGAGGTCAATACTTTGCGCCCTATGAAGTATGAGCAAATAGCCGAGCTGCTTAATGGTATCGCTGAAAGAT
TTGAATGGGAAAAAGTAATGGAAAGTGACAAGATCATTGGTCTGAAGCAGGGAAAGCAAAGCATTCTACTGGAAC
CTGGGGGTCAATTCGAGCTTAGTGGTGCACCTCTTGAGACTTTGCACCAAACCTGTGCTGAAGTCAATTCACATC
TTTATCAGGTAAAAGCTGTTGCTGAGGAAATGGGAATTGGTTTCTTAGGAATCGGCTTCCAGCCCAAATGGCGTC
GGGAGGATATACCCATCATGCCAAAGGGGAGATACGACATTATGAGAACTACATGCCGAAAGTTGGCACCCCTTG
GACTTGATATGATGCTCCGAACGTGTACTGTTGAGGTTAATCTGGATTTTAGCACAGAAGCTGATATGATCAGGA
AGTTCGGTGCTGGTCTTGCTTTGCAACCTATAGCAACGGCTCTATTTGCAAATTCCTTTTTACAGAAGGAAAGC
CAAACGGATTTCTCAGCATGAGAAGCCACATATGGACAGACACTGACAAGGACCGCACAGGAATGCTACCGTTTG
TTTTCGATGACTCTTTTGGGTTTGAGCAGTATGTTGACTACGCACTCGATGTCCCTATGTACTTTGCCTACCGAA
AGAATAAATACATCGACTGTACTGGAATGACATTTGCGCAATTCCTTGGCTGGAAAACCTCCCTGTCTCCCTGGTG
AACTGCCTTCATATAATGATTGGGAAAACCATCTGACAACAATATTTCCAGAGGTTTCGGTTGAAGAGATACTTGG
AGATGAGAGGTGCTGATGGAGGTCCCTGGAGGAGGCTGTGTGCCCTGCCAGCTTTCTGGGTGGGTTTATTATATG
ATGATGATAGTCTCCAAGCTATCCTGGATCTGACAGCTGACTGGACTCCAGCAGAAAGAGAGATGCTAAGGAACA
AAGTTCAGTTACTGGCTTAAAGACTCCGTTTAGAGATGGTTTGTAAAGCATGTCGCTGAAGATGTCCTGAAAC
TCGCAAAGGATGGTTTAGAGCGCCGAGGCTACAAGGAAGCTGGTTTCTTGAACGCAGTTGATGAAGTGGTCAGAA
CAGGAGTTACGCCTGCAGAGAAGCTCTTGAGATGTACAATGGAGAATGGGGACAAAGTGTAGATCCCGTGTTCG
AAGAACTGCTATACTAAGAGAATGGGACGTGAACAAAAGGTGTCTATAAACCTTTGGGTGTGAGTTTATGCTCTC
TGAAGAACTCAAGTCTCAGGAATAAGTATTTTTTTTTCTTTCTGTAATCGGATTTTAAAACTG

FIG. 10

SEQ ID No. 9

>GSHA protein Araly [UniProt ID: D7MDW8]

MALLSQAGGSYTVVPLGCSKTGTKSVSGGVRNLDVLRMKEAFVSSYSRSLSTKSMLLRSVKRSKREHQLIVAASP
PTEEAVVATEPLTREDLIAYLASGCKTKDKYRIGTEHEKEFGFEVNTLRPMKYEQIAELLNGIAERFEWEKVMEGD
KIIGLKQKGQISISLEPGGQFELSGAPLETTLHQTCAEVNSHLYQVKAVAEEMGIGFLGIGFQPKWRREDIPIMPKG
RYDIMRNYMPKVGTGLDMLLRTCTVQVNLDSTEDMIKFRAGLALQPIATALFANSPTFTEGKPNGLSMRSH
IWTDTDKDRTGMLPFVFDSDSGFEQYVDYALDVPMYFAYRKNKYIDCTGMTFRQFLAGKLPCLPGELPSYNDWEN
HLTTIFPEVRLKRYLEMARGADGGPWRRICALPAFWVGLLYDDDSLQAILDLTADWTPAEREMLRNKVPVTGLKTP
FRDGLLKHVAEDVLKLAKDGLERRGYKEAGFLNAVDEVVRTGVTPEKLLLEMYNGEWGQSVDPVFEELLY*

FIG. 11

11/37

SEQ ID No. 10

>NM_001203879.1__Arabidopsis_thaliana_glutamate
cysteine_ligase_mRNA,_complete_cds

ACACTATAAAGTTAACCTTTGTTTAAACTACTAATAAACTGGTTATTTAAAAAACTATTAGGGTTAAATTTGTGA
GAAATGGAAAAGTAAGAGGCCAAAAACGAAAGAATCTTTTATTACAATGATTATTAATTTTATCAACTTTCTTTT
ATAGACAGCTCAGCTCCACTGATCGGGCTCATGCTCCACTGATCAACTTTCTTTTCAATTTTTTATCAGTTTCTCAT
TTCTCTCTCAGTCGCATATAAAATTTGAAACGCCGATTCGTTTTTGTCTCTTTAAGCTTTCTTCTTGATTTTCGACGG
TGTCTCTCAATCTCCGTCAAGCTTGACGAATTTCAAGGCTGGAACATAAGCTGTTGTTTCGGGTGGCGTGAGGA
GATCATACACTGTTGTTCTTCTGGAGTTTGTTCAAAGGCTGGAACATAAGCTGTTGTTTCGGGTGGCGTGAGGA
ATTTGGATGTTTTGAGGATGAAAGAAGCTTTTGGTAGCTCCTACTCTAGGAGTCTATCTACCAAATCAATGCTTC
TCCATTCTGTTAAGAGGAGTAAGAGAGGGCATCAATTGATTGTTGCGGCAAGTCTCCAAACGGAAGAGGCTGTAG
TTGCAACTGAGCCGTTGACGAGAGAGGATCTCATTGCCATCTTGGCTCTGGATGCAAAACAAAGGACAAATATA
GAATAGGTACAGAACATGAGAAATTTGGTTTTGAGGTCAATACCTTTGCGCCCTATGAAGTATGATCAAATAGCCG
AGCTTCTTAATGGTATCGCTGAAAGATTTGAATGGGAAAAAGTAATGGAAGGTGACAAGATCATGGTCTGAAGC
AGGGAAAGCAAAGCATTTCACTTGAACCTGGGGGTGAGTTGAGCTTAGTGGTGCACCTCTTGAGACTTTGCATC
AAACTTGTGCTGAAGTCAATTCACATCTTTATCAGGTAAGAGCAGTTGCTGAGGAAATGGGAATTGGTTTCTTAG
GAATTGGCTTCCAGCCCAAATGGCGTCGGGAGGATATACCCATCATGCCAAAGGGGAGATACGACATTATGAGAA
ACTACATGCCGAAAGTTGGTACCCTTGGTCTTGATATGATGCTCCGAACGTGTACTGTTGAGTTAATCTGGATT
TTAGCTCAGAAGCTGATATGATCAGGAAGTTTTCGTGCTGGTCTTGCTTTACAACCTATAGCAACGGCTCTATTTG
CGAATTTCCCTTTTACAGAAGGAAAGCCAAACGGATTTCTCAGCATGAGAAGCCACATATGGACAGACACTGACA
AGGACCGCACAGGAATGCTACCATTTGTTTTCGATGACTCTTTTGGGTTTGAGCAGTATGTTGACTACGCACTCG
ATGTCCCTATGTACTTTGCCACAGAAAGAACAAATACATCGACTGTACTGGAATGACATTTCCGCAATTCTTGG
CTGGAAAACCTTCCCTGTCTCCCTGGTGAACCTGCCTTCATATAATGATTGGGAAAACCATCTGACAACAATATTCC
CAGAGGTTTCGGTTGAAGAGATACTTGGAGATGAGAGGTGCTGATGGAGGTCCCTGGAGGAGGCTGTGTGCCCTGC
CAGCTTTCTGGGTGGGTTTATTATATGATGATGATAGTCTCCAAGCTATCCTGGATCTGACAGCTGACTGGACTC
CAGCAGAGAGAGAGATGCTAAGGAACAAAGTCCCAGTTACTGGCTTAAAGACTCCTTTTAGGGATGGTTTGTAA
AGCATGTGCTGAAGATGTCCTGAAACTCGCAAAGGATGGTTTAGAGCGCAGAGGCTACAAGGAAGCCGGTTTCT
TGAACGCAGTCGATGAAGTGGTCAGAACAGGAGTTACGCCCTGCCGAGAAGCTCTTGGAGATGTACAATGGAGAAT
GGGACAAAAGCGTAGATCCCGTGTTTGAAGAGCTGCTGTACTAAGAAAATGGGACGTGAACAAAAGGTGCTATA
AACCTTTGGGTGTGAGTTTATGCTATCTGAAGAATTCGAGTCTCCGGAATAAGGATTTTTTTTTTTTTTTTTTTGTTT
TTTTTATTGTTTTTTTTTTTTTAAGTTCGAATATGTCGTATAGATCTGTAAAATAATTCAAATGTGGCTTTGTCCCAT
AGATTGGTTTTTGATTATGCAATCATCTATCACAACAGTGGAACATTATGCAATCATTTATTGTCTATCAAATTAA
TTATGGTTATGAAA

FIG. 12

SEQ ID No. 11

>GSHA protein Arath [UniProt ID: F4JMS5]

MALLSQAGGSYTVVPSGVCSKAGTKAVVSGGVRNLDVLRMKEAFGSSYSRSLSTKSMLLSVSKRSKRGHQLIVAA
SPPTEEAVVATEPLTREDLIAYLASGCKTKDKYRIGTEHEKFGEVNTLRPMKYDQIAELLNGIAERFEWEKVME
GDKIIGLKQKQKQISLEPGGQFELSGAPLETLHQTCAEVNSHLYQVKAVAEEMGIGFLGIGFQPKWRREDIPIMP
KGRYDIMRNYMPKVGTGLDMLLRTCTVQVNLDIFSSEADMIRKFRAGLALQPIATALFANSFPTEGKPNGLSMR
SHIWTDTDKDRTGMLPFVFDSSFGFEQYVDYALDVPMYFAYRKNKYIDCTGMTFRQFLAGKLPLPGELPSYNDW
ENHLTTIFPEVRLKRYLEMRGADGGPWRRRLCALPAFWVGLLYDDDSLQAILDLTADWTPAEREMLRNKVPVTGLK
TPFRDGLLKHVAEDVLKLAKVCLYEPS*

FIG. 13

12/37

SEQ ID No. 12

>XM_006439493.1__Citrus_clementina_hypothetical_protein(CICLE_v
10019696mg)_mRNA,_complete_cds

AGAATAAATACGCTTAGAGAATCTTGATAGCGGTGACGTGGCGACAACCAAGAAAGGGGCATCATCAAAAGGCACC
ATAGTGCCGCATTATCGAATCCCCGTCCCCAGCGCCATTTGATAGTTTCAGCAGGTTCTGTATATAATAACACTAA
CGTCTAACAATACTCACAAGAAAACCAAAGTCCCTCTCTCTCTCTCTGCAATTCTGCAAGTGAGTCTTCGCTGTCC
TATATATATATATAACTCAACGAGTGTTTTATCTTCTACGAAATTCGAAACCCTAGTTTTTAAGATTTCCAATTGT
GGTAATTCGTTTCAGCGTTAACCAAGCGGGAGCAGTATGACTCTCATTTTCGCGGGCCAGTCCATCATGTCGTGTAT
TGACTGATGTAACACGATGTACAATCAGACGAAGTGGGGTTTGTGGCTTGGCAAAAAATGCTGAGGCATCTAGGA
TGAAGGAACTTGTGTTAGTTTTCTTTCCTCATCACGCTATTCTGCCAAGATATCACAAGTTTTACATCTAGACA
ATGCAGGGATAGGAATTCGAAGAGGACGTGATCTATTGTTGCAGCTAGTCCCTCCACGGAGGATGCTGTTGTTG
TAACGGAGCCATTGACCAAAGAGGATCTAGTTGGATACCTTGCCCTCTGGGTGCAAGCCTAAGGAAAAAGTGGAGAA
TTGGTACAGAACATGAGAAGTTTGGTTTTGAGTTTGGAACTTTACGTCCTATGAAATACGAACAAATAGCCGAGT
TGCTTAACAGCATTGCTGAGAGGTTTCGATTGGGAGAAAAGTAATGGAAGGAGATTACATCATAGGACTGAAACAAG
GGAAACAAAGCATATCACTTGAACCTGGCGGTGAGTTTTCGAGTTAGTGGTGCACCTCTTGAAACTCTGCATCAAA
CTTGTGCTGAGGTAAATTCACACCTATATCAGGTGAAAGCTGTTGCAGAGGAAATGGGAATTGGATTTTTAGGAA
TTGGCTTCCAGCCTAAGTGGGGACTTAAAGATATACCTGTTATGCCAAAGGGAAGATATGAGATTATGCGGAACT
ACATGCCTAAAGTTGGATCACTCGGACTTGATATGATGTTTAGGACATGCACTGTTTCAGGTTAATTTGGACTTTA
GTTTCGGAAGCCGACATGATCCGCAAATTTTCGTGCTGGTCTTGCTTTCAGCCCAATAGCAACAGCTCTATTTGCAA
ATTCACCATTATACGAAGGAAAAACCAAATGGCTATCTCAGCATGAGAAGCCACATCTGGACAGATACCGATAACA
ATCGTGCTGGCATGCTTCCCTTTGTTTTTATGACTCCTTTGGGTTTGGAGCAATATGTTGATTACGCCCTTGATG
TTCCAATGTATTTTTGTTTTATCGCAAAAAAGATATATTGATTGTGCTGGGATGTCATTTCAGGGACTTCTTGGCTG
GAAAACCTCCGTGCCTTCCCTGGTGAATTGCCAACTCTTAATGATTGGGAAAAATCATCTGACAACAATATTTCCCG
AGGTTAGGTTGAAGAGATACTTGGAGATGAGAGGCGCTGATGGAGGGCCTTGGAGGAGGTTATGTGCTTTGCCTG
CATTTTGGGTAGGTTTGTATATGACGAGGACTCTCTCCAGAATGTTCTAGACATGACAGCAGATTGGACTACAG
GTGAAAGACAAATGTTGAGGAATAAGGTCCCAAAGACTGGTCTAAAGACGCCATTTCCGGATGGTTTGCTAAGGC
ATGTTGCCCAAGATGTTTTAAAGTTGTCAAAGGTAACCTGGAGTTGTCTCTTCTACTGTCCTTTATCTTGTTGAT
ATTGGAAGTAAACATCTATTCTTCTATAGGATGGTTTGAACGGAGAGGCTTCAAAGAGACAGGGTCTTGAAATG
AGGTGGCTGAGGTAGTCAGAACAGGGGTGACACCGCGGAAAAACTATTGGACATGTATCATGGGAAGTGGCGAG
AGTCAGTAGATCCTGTTTTTGGAGAGCTACTTTATTAAGATGAGTTAAGATTATTTTCGACACGCTTCTGTTTGCA
ATTGTTGCTTAAGGTTTTGGGGAGTAATCCTTTTGTTCGGTCCCTACTGTTTTTAATTTATGATCATTAGGAATA
AAAGTATGATGTAAATTCTCAGCATGAGGATCCTCATGCAAGCAACTCTGATTGAATGTAACCTGGATCTGAAACT
TGGGGCAGCTCGACCAAGAATTGAGTAGGCCATGTGACTTGGGACGTAAATCATGTGCACGTACATTGATCATT
GAAGCCCCATCATAGCATAGCTTTAATATTACGATTTCAAGTTTGTTCAAATTATAGGTGTGTGCAGTTTGTCTAT
AGAGGAGACATTTTAACATCATTGGTGGATTAGATTGAGACTTCTGATCCTATCCTTCTCCCATTCCTCTCGGC
GGTGGCCTAGAGTGTGATGAAAACAATGCTCGAGCTAACCACTTTGATGATTACCTCATTCATCGACATCCTA
CTCTCTTTTGTCTATTGATATCAAGTGGACTGGGAACATATAATTTAATACGTTTTGTTGTTTCCACAGCGCTTT
GTGTATGCATGCTTGTGGACTCTGGTAGATTGATTGAGACTAAAAATCATCCAAACGCTAAAGCATACT
TGGGACATTGGATTCCAACAAGTAGCTGGCTGGAGGGGGCCAAAAACAATACTCAATGCTCGTATGGTCATGGAGGA
CAGACTGACACTCAATAATATGAATCTTCTGCTGAGCCAACTAATCTGATTCATAAACTATACATGCAATCTTT
TATTGATTTAGGTGCTGTAACCACCTGATCTGCATCGACATGTCCAAATTATCCTTTTAAATGTCTTATTATCTAA
ATTGATCATGGTTTACTAACTTGTCCAAAGAAAGTATCAAATAACACGCAGCTTTTACTTATCTTAGGCATCATA
AACTAATC

FIG. 14

SEQ ID No. 13

>GSHA (GSH1) protein Cicle [UniProt ID: V4TSB7]

MTLISRASPSRCVLTDVTRCTIRRSVCGGLAKNAEASRMKETCVSFSSSSRYSAKISQVLHLDNAGIGIRRGRRS
IVAASPTEDAVVVTEPLTKEDLVGYLASGCKPKEKWRIETEHEKFGFEFGTLRPMKYEQIAELLNSIAERFDWE
KVMEGDYIIGLKQKQKQISLEPGGQFELSGAPLETLHQTCAEVNSHLYQVKAAEEMGIGFLGIGFQPKWGLKDI
PVMKPGRYEIMRNYMPKVGLSLGLDMMFRTCTVQVNLDFSSEADMIRKFRAGLALQPIATALFANSFPTEGKPNY
LSMRSHIWTDTDNRRAGMLPFVFDGSGFGEQYVDYALDVPYFVYRKKKYIDCAGMSFRDFLAGKLPCLPGELPT
LNDWENHLTTIFPEVRLKRYLEMGRADGGPWRRCLALPAFWVGLLYDEDSLQNVLDMTADWTTGERQMLRNKVPK
TGLKTPFRDGLLRHVAQDVLKLSKVTWSCLFYCPLSC*

FIG. 15

13/37

SEQ ID No. 14
>XM_003525349.2
Glycine_max_gamma-glutamylcysteine_synthetase_precursor(GSH1),_
transcript_variant_X1,_mRNA
AGAATGAGAAATTGCGTTCTGCGTGTGTTCCCTTGGTAATCGTGAGTGGTACTATTCCCGAATAGCAGCAAGCAA
GGCAAGGCGTGTGTCCGAGCAGAGCAGGAGTGAAGGAGAAAAGGACCCCTTGTCCTCCAGTCCTTATTCTCTTTT
CTCCTCCCTCCCTCACGCAACCGCCAATTCTGAACCCTAAGCTAGCGCCATGGCTGTCGTTTCGCGAAGTGCGA
CGACCTATACGCGCCACTACTTAATACGACACGAGTTTGATAGGAAAACGAAAACCTGCGTTGCCAATAATAGTT
TGTGTTACTCTGCTAAGAAGGCTCCTCCACCGCAGAGGATTGTTGGTGGCCGTAGAGTGATTGTTGCTGCAAGCC
CTCCCACCGAAGACGCTGTAGTTGCCACTGACCCTCTCACGAAGCAGGATCTCGTCGATTATCTTGCCCTCCGGTT
GCAAGCCCAAGGATAAATGGAGAATAGGTACTGAACATGAGAAGTTTGGTTTTGAGATTGGAAGCTTGCGTCCCTA
TGAAGTATGACCAAATAGCAGAATTGCTGAATGGCATTGCTGAGAGGTTTGACTGGGATAAAGTAATGGAAGGTG
ATAAAATTATTGGACTCAAACAGGGGAAGCAGAGCATATCATTGGAGCCTGGTGGTCAGTTTGAACCTTAGTGGAG
CTCCTCTTGAAACCTTGTCATCAGACTTGTGCTGAAGTTAATTCCACCTTTATCAGGTTAAAGCTGTTGCTGAGG
AAATGGGAATTGGATTTTTGGGGATTGGTTTCCAGCCAAAAGTGGGGAATCAAAGACATACCTATAATGCCAAAGG
GAAGATACGACATCATGAGGAACACATGCCTAAAGTTGGCTCTCTGGGCTTGACATGATGTTCCAGGACATGCA
CTGTGCAGGTCAATCTGGACTTTAGTTCTGAAGCTGACATGATCAAGAAATTTGTCGAGGCCCTTGCTTTGCAGC
CGATAGCAACGGCTCTTTTTGCAAATTCACCCCTTAAAGAGGGAAAAGCCAAATGGTTTTGTGAGTATGAGAAGCC
ATATTTGGACTGATACTGATAAGGACCGCACAGGCATGCTGCCTTTTGTGTTTGTGATGACTCTTTTGGGTTTGGAGC
AATATGTTGATTATGCTCTTGATGTTCCCTATGTATTTTGTCTATCGGAAAAACAGATATATCGACTGCCTGGAA
AGACCTTCAGGGACTTTTTGGCTGGAAGACTTCCTTGTATCCCTGGTGAATTACCAACTCTCAATGATTGGGAAA
ATCACTTGACAACATATATTTCTGAGGTGAGGCTGAAGAGGTATTTGGAGATGAGAGGTGCTGATGGAGGGCCCTT
GGAGAAGATTGTGTGCTTTACCAGCATTTTGGGTAGGGTTATTGTACGATGAACTTTCTCTAAAAAGTGTTTTGG
ATATGACAGCTGATTGGACTCCAGAAGAAAGACAAATGTTAAGGAATAAGGTTCCCTGTAACCTGGTCTGAAGACAC
CATTCCGAGACGGTTTGTGTAAGCATGTTGCTGAAGATGTTCTAAAGTTGGCAAAGGATGGCTTGGAGAGAAGAG
GCTTCAAGGAATCGGGATTTTTGAATGAGGTTGCCGAGGTGGTTAGAACAGGTGTCCTCCAGCTGAGAGGCTTT
TGGAATTGTATCATGGAAAGTGGGAGCAATCCGTAGATCATGTGTTGAGGAATTGCTTTATTAAGGTAGTATTG
TCTTTCAAATGTCTGTGGAAGATTAATTGTGTAATCCTTTGGTTATAGTTCTGGTTATCTCTCATTGAGCTTCA
TTTAGATATAGGAAATAATATAAATGTAATTTTTGCCTTGAGGCTTCTCATATACCTTTCTTGGAACATGATC
TGATCAAGCTGACATTCAATGTCACAAATTTGAAAATCCATCAATATAATCATTTTAATAAGAGTGTCAATAGAT
TAAATGTTTGTGGTA

FIG. 16

SEQ ID No. 15
>GSHA protein Glyma [UniProt ID: I1K6J1]
MAVVSRSATTYTRHYLIRHEFDRKTKTCVANNSLCYSAKKAPPPQRIVGGRRVIVAASPPTEDAVVATDPLTKQD
LVDYLASGCKPKDKWIRIGTEHEKFGEIISLRPMKYDQIAELLNGIAERFDWDKVMEGDKIIGLKQGKQSSISLEP
GGQFELSGAPLETLHQTCAEVNSHLYQVKAVAEEMGIGFLGIGFQPKWGIKDIPIMPGRYDIMRNYMPKVGSLG
LDMMFRTCTVQVNLDFSSEADMIKKFRAGLALQPIATALFANSFPKEGKPNFVSMRSHIWTDDTKDRTGMLPFV
FDDSFGEFYVDYALDVPYFVYRKNRYIDCTGKTRFDRLAGRLPCIPGELPTLNDWENHLTTIFPEVRLKRYLE
MRGADGGPWRRLCALPAFWVGLLYDELSLKSVDLMDTADWTPEERQMLRNKVPVTGLKTPFRDGLLKHVAEDVLKL
AKDGLERRGFKESGFLNEVAEVVRTGVTPAERLLELYHGKWEQSVDHVFEELLY*

FIG. 17

14/37

SEQ ID No. 16

>XM_003630464.1_Medicago_truncatula_(Medtr)

Glutamate-cysteine_ligase_(MTR_8g098350)_mRNA,_complete_cds

TTATTCTCTTTTTCAATTTTCAATTTTCAATTTCTAATTTCTAATTTGTAATTTCAATACTATCATGACTACAATAT
 TTAGACTCGCTTCCTCATCTCTCCCTCTCTCCGCCACGACGCAACACCGCATAACTTCCATATTCGGAAAACCT
 CCATTTCCAACACCTTCTCTTTCTCTTCAAAGAATTCTTTGTCGTTCAAGAGAATTCTTACTTCCGGTGGTAGTC
 GTAGATTCATCGTTGCTGCTAGTCCTCCAACCTGAAGATGCTGTTGTTGCTACTGAACCTCTTACAAAACAGGATC
 TCATTGATTATCTCGCTTCTGGCTGCAAAACCAAGGATAAATGGAGAATAGGTACTGAACATGAAAAGTTTGGTT
 TTGAGCTTGGAAGCTTGCGTCTTATGAAGTATGAGCAAAATATCAGAATTGCTGAATGGCATTGCTGAAAGATTTG
 ACTGGGATAAAGTAATGGAAGGTGATAACATTATTGGACTCAAACAGGGGAAGCAGAGCATATCATTAGAGCCAG
 GTGGTCAGTTTGAACCTTAGTGGAGCCCCCTCTGGAACTCTGCATCAGACTTGTGCTGAAGTTAATTGCGACCTCT
 ATCAAGTTAAAGCTGTTGCTGAGGAGATGGGAATTGGATTTTTGGGGATTGGTTTCCAGCCAAAGTGGGAACGCA
 AAGACATACCTATGATGCCAAAGGGAAGATATGAGATTATGAAAAAATACATGCCTAAAGTTGGTTCTCTTGGGC
 TTGACATGATGTTTCAAGGACATGCACTGTACAGGTAAATCTGGACTTTAGTTCTGAAGCTGACATGATCAGGAAAT
 TTCGTGCTGGCCTAGCTTTCAGCCAATAGCAACAGCTCTTTTTGCAAATTCACCCTTCACAGATGGAAGCCCA
 ATGGGTTTGTAGTATGAGAAGCCATATTTGGACTGATACCGATAAGGACCGCACAGGCATGCTGCCTTTTGT
 TTGATGACTCTTTCGGGTTTGAAGCAGTACGTTGATTTTGTCTTGTATGTTCCAATGTATTTGTCTATCGGAAA
 AGAAATATATCGACTGTACTGGAATGACCTTCAGGGACTTCTTGGCAGGAAAGCTTCCTTGTATTCCTGGTGAAT
 TACCGACCTTAATGATTGGGAGAATCATTTGACCACTATATTTCCCGAGGTTAGGCTAAAGAGGTATTTGGAGA
 TGAGAGGTGCTGATGGAGGGCCTTGGAGAAGGTTATGTGCTTTACCAGCATTTTGGGTAGGGATATTGTATGATG
 AGGTTTCTTTGCAACGTGTTTTGGATATGACAGCAGATTGGACTCTAGAAGAAAGAGAAATGCTAAGGAATAAGG
 TCACCGTAACTGGTCTAAAGACACCATTTTCGAGATGGTTTGTGCTGAAGCATGTTGCTGAAGAGGTTCTAGAGCTGG
 CAAAGGATGGCTTGGAAGAAGAGGCTTTAAGGAATCAGGATTTTTAAATGCGGTAGCTGAGGTGGTTAGAACAG
 GTGTCACTCCAGCTGAGAGGCTTTTGAATGTATCATGGAAAGTGGGAGCAATCTGTAGATCATGTATTTGACG
 AATTACTATACTAAGGTGTATTCGGCGTCTTGCTTTTCTTTCACATTTCTGTGGGAGGTTAATTGTATAAGCCAC
 CTTTTTGGGTAGAATCGTGTGTAAACTCTGGCTATATAGTCTGTTATCTCTCTTGTGAGCTGCTTTTAGAT
 ATAGAAAATAATATCTGATGTATTTTTCGCTTGGAAACATTATCAGATCAAGCTAGTATTTAATTATCATGTCGTA
 AGTTTGATAATAACATCAGCTTAATCAGCTTGTCAAAAAAATAAACATCAGCTTAATCATCTTCACCAAAGTT
 GTTTGAAGGAACCTTTCTTTTCTCTAATTCAGATTGGAGACTAATTCAA

FIG. 18

SEQ ID No. 17

>GSH1 protein Medtr [UniProt ID: G7LHC5]

MTTIFRLASSSSPSLRHDATPHNFHIRKTSISNTFSFSSKNSLSFKRILTSGGSRRFIVAASPTEDAVVATEPL
 TKQDLIDYLASGCKTKDKWRIGTEHEKFGFELGSLRPMKYEQISELLNGIAERFDWDKVMEGDNIIGLKQKQSI
 SLEPGGQFELSGAPLETLHQTCAEVNSHLYQVKAVAEEMGIGFLGIGFQPKWERKDI PMMPKGRYEIMKKYMPKV
 GSLGLDMMFRTCTVQVNLDFSSEADMIRKFRAGLALQPIATALFANSFPFDGKPNFGFVSMRSHIWTDDDKDRTGM
 LPFVFDSDSFGFEQYVDFALDVPMYFVYRKKYIDCTGMTFRDFLAGKLPCIPGELPTLNDWENHLTTIFPEVRLK
 RYLEMRGADGGPWRRICALPAFWVGILYDEVSLQVRVLDMTADWTLEREMLRNKVTVTGLKTPFRDGLLKHVAEE
 VLELAKDGLERRGFKESGFLNAVAEVVRTGVTPEERLLELYHGKWEQSVDHVDFDELLY*

FIG. 19

15/37

SEQ ID No. 18

>NM_001247081.1__Solanum_lycopersicum (Solly)_gamma-glutamylcysteine_synthetase_(GSH1),_mRNA

AGATTCTTTTTCATTCAATCTGCTAGCAGTTGTCTCTCTTTATAGCTAAACAAAAATTAAAGAAATGGGAGTGAT
 AAAGCTCTAGCCGCCTCAGCACACCAAAATCCTAATCATTTTGTCTGATACCAATTGTCTCTATATATATATTGG
 AAGGTTTTCTTTCAATTTTTCAGTCAATTCTATAGAACAAACCTACAAAGTCTTAATTTTTCAGTTTACTAGAGC
 CAAGGCACAAAGCATGGCCTTGATGTCTCAGGCAGGTTCTTCACATTGCATTTACTCTGAAAAGGTAAGATGTAT
 CAGTGGACATAGAAGCATCATTACAATATGGATATGTTTCAAGATGCGGGAGATATGCTTTGGCGTGGATATTTT
 TTCACGCAATGCCTCAAGACGAGTGCAAGGGAATTATTTGAATCACATAGGAGTGGGAAGTAGACGTGGAGACCT
 GACAATTGTAGCTGCAAGTCTTCCAACAGAGGATGCTGTTGTTGCAGCAGAGCCGTTAACAAAAGAAGATCTTGT
 AGGATATCTTGCTTCTGGATGCAAATCCAAAGAAAAGTGGAGGATAGGCACTGAACATGAAAAGTTTGGTTTTCGA
 GTTTGGAAACCCTGCGACCCATGAAGTATGATCAAATAGCTGACTTGCTAAATGGTATTGCTGAGCGGTTTGATTG
 GGAAAAGTAATGGAGGGTGACAAGATTATTGGCCTGAAACAGGGGAAAGCAAAGCATATCATTAGAACCCTGGTGG
 TCAGTTTGGAGCTTAGTGGTGCACCACTTGAAACACTGCATCAAACCTTGTCAGAGGTTAATTCACATCTTTACCA
 GGTAAAGCTGTTGCAGAAGAGATGGGAATTGGATTCTTAGGAACTGGATTCCAGCCAAAGTGGGGGCTGAAAGA
 TATACCAATAATGCCGAAGGGGAGATATGAGATTATTAGAAATTACATGCCTAAAGTTGGCTCACTTGGGCTGGA
 TATGATGTTTAGAACATGCACTGTTTCAAGTAAATCTGGACTTCAGTTCTGAAGCTGACATGATCAGAAAGTTCCG
 TGCTGGTCTTGCTTGCAGCCTATTGCTACAGCTCTATTTGCAAATTCACCTTTCAGTGAAGGAAAACCTAATGG
 TTATCTCAGCAAGAGAAGCCACATTTGGACCGATACAGATAATAACCGCGCTGGGATGCTTCCTTTCTGCTTTGA
 TGACTCTTTTGGGTTTGGAGCAATATGTAGATTATGCACTTGATGTCCCAATGTATTTTGTCTATCGGAAGAAGAA
 GTATGTTGATTGTACTGGATTGTCTTTCCGGGACTTCATGAATGGAAAACCTCCGCCCATTCGCCGCGAATACCC
 TACTCTTAATGATTGGGAGAATCATCTCACAACAATATTTCTGAGGTCAGACTCAAAAGATATCTGGAAATGAG
 AGGTGCTGATGGAGGGCCTTGGAGACGGTTATGCGCATTGCCTGCATTCTGGGTGGGTATACTCTACGATGAGGG
 GTCTTTGCAAAGCGTTTGGACATGACATTTGATTGGACTGCAGAAGAAAGAGACATGTTAAGGAATAAGGTGCC
 AAAAAGTGGTCTGAAGACACCATTTTCAGATGGATTGCTTATGCATGTTGCTCAAGATGTTGTCAAGTTGGCAA
 GGAAGGCTTGGAAAGAAGAGGCTTTAAGGAAACAGGATTTCTGAACGAAGTAGCCGAGGTTGTTAAACAGGTGT
 AACACCAGCTGAGAAGCTACTGGAATTGTACCATGGGAAGTGGGGACAAAGTGTGGATCCTATCTTTGAGGAGCT
 TCTCTACTGAAGTTGTTTGGAGAGTGCAATTTCAATTCTGTCTTTTCTGTGTTGTAAATGATTTCTCAATTTCA
 TTGAGAGTTATATCAGCCGATTTTTCCATCGTTAGACAAATAATCTAGCTAGCTAGATTAAGATATGTTACATGT
 ATCTATGGCTCGAAAACGTCATATGTAGCTCTCTCCCTTGGGATGTGATCAATCTGGTTCCAGATAGCATTTGT
 ATACATTTTCCCCAGAAAAGTAACATAGCAATGCTAAAGTAATAAATGAATCTATTGGCTGACCAAAAAAAAAA

FIG. 20

SEQ ID No. 19

>GSHA protein Solly [UniProt ID: O22493]

MALMSQAGSSHCIYSEKVRCSISGHRSIINNMDMFRMREICFGVDISSRNASRRVQGNLYLNHIGVGSRRGDLTIVA
 ASPPTEDAVVAAEPLTKEDLVGYLASGCKSKEKWRIgteHEKFGFEFGTLRPMKYDQIADLLNGIAERFDWEKVM
 EGDKIIIGLKQKQSIISLEPGGFELSGAPLETlHQTCAEVNSHLYQVKAVAEEMGIGFLGTGFQPKWGLKDIPIM
 PKGRYEIIIRNYMPKVGSLGLDMMFRTCTVQVNLDFSSEADMIRKFRAGLALQPIATALFANSPTTEGKPNGLSK
 RSHIWTDTDNRRAGMLPFVFDSSFGFEQYVDYALDVPMyFVYRKKKYVDCTGLSFRDFMNGKLPPIPGEYPTLND
 WENHLTTIFPEVRLKRYLEMrgADGGPWRRLCALPAFWVGILYDEGSLQSVLDMTFDWTAEERDMLRNKVPKSL
 KTFPRDGLLMHVAQDVVKLAKEGLERRGFKETGFLNEVAEVVKTGVTpAEKLELYHGKWGQSVDPiFEELLY*

FIG. 21

16/37

SEQ ID No. 20

>XM_006343925.1__Solanum_tuberosum (Soltu)_glutamate-cysteine_ligase,_chloroplastic-like_(LOC102600891),_mRNA

TTCAATCTGCTACAGTACTCTCTCTTTATAGCTTAACAAAATTAAAGAAATGGGAGTGATAAAGCTCTAGCCGCC
 TCAGCACACCAAAATCCAAATCATTTTGGCTGATACCCAATTTTCTCTATATATATATTGGAAGGTTTTCTTTCAA
 TTTCCAGTCAATTCTATAGAACAAACCTACAAAGTTTTTCTTTTTCTAGTTTACTGAGCCAAGGCACAAAGCATG
 GCCTTGATGTCTCAGGCAGGTTCTTCACATTGCATTTACTCTGAAAAGATAAGATGTATCGGTGGACATAGCAAC
 ATCATGAACAATATGGAGATGTTCAAAATGAGGGAGATATGCTTTGGCGTGGATATTTCTTCACGGAATGCCTCA
 AGACGAGTGCAAGGGAATCATTTGAATCTCATAGGAGTGGGAAGTAGACGTGGAGGCCTGACAATTGTTGCTGCA
 AGTCCTCCAACAGAGGATGCTGTTGTTGCAGCAGAGCCGTTAACAAAAGAAGATCTTGAGGATATCTTGCTTCT
 GGATGCAAAACCAAGAAAAATGGAGGATAGGCACCTGAACATGAAAAGTTTGTTTTCGAGTTTGGAACTCTGCCGA
 CCCATGAAGTATGATCAAATAGCTGACTTACTAAATGGTATTGCCGAGCGGTTTGACTGGGAAAAAGTAATGGAG
 GGTGACAAGATTATTGGCCTGAAACAGGGGAAAGCAAAGCATATCATTGGAACCTGGTGGTCAGTTTGAACCTAGT
 GGTGCACCAGTTGAAACACTGCATCAAACCTTGTGCAGAGGTTAATTCACATCTTTACCAGGTTAAAGCTGTTGCA
 GAAGAGATGGGAATTGGATTCTTAGGAACTGGATTCCAGCCAAAAGTGGGGGCTGAAAGATATACCAATAATGCCG
 AAGGGGAGATATGAGATTATGAGAAATTACATGCCTAAAGTTGGCTCACTTGGGCTGGATATGATGTTTAGAACA
 TGCATGTTTCAGGTAAATCTGGACTTTAGTTCTGAAGCTGACATGATCAGAAAGTTTCGTGCTGGTCTTGCCTTG
 CAGCCTATTGCTACAGCTCTATTTGCAAATTCGCCTTTCACTGAAGGAAAACCTAATGGTTATCTCAGCAAGAGA
 AGCCACATTTGGACTGATACAGATAATAACCGCGCTGGGATGCTTCCTTTCGTCTTTGATGACTCTTTTGGGTTT
 GAGCAGTATGTGGATTATGCACTTGACGTCCCTATGTATTTTGTCTATCGGAAGAAGAAGTATGTTGATTGTTCT
 GGATTGTCTTTCCGGGACTTCATGAATGGAAAACCTCCGCCTATTCCCGGCGAATATCCTACTCTTAATGATTGG
 GAGAATCATCTCACAACAATATTTCTGAGGTCAGACTCAAAAGATATCTGGAAATGAGGGGTGCTGATGGAGGG
 CCTTGGAGACGGTTATGTGCATTGCCTGCATTCTGGGTGGGTATACTCTACGATGAGGGGTCTTTGCAAAGTGTT
 TTAGACATGACGTCTGATTGGACTGCGGAAGAAAGAGACATGTTGAGGAATAAGGTGCCAAAAAGTGGTCTGAAG
 ACACCATTTTCAGATGGATTGCTTATGCATGTTGCTCAAGATGTTGTCAAGTTGGCAAAGGAAGGCTTAGAAAGA
 AGAGGCTTTAAGGAAACAGGATTTTTGAATGAAGTAACCGAGGTTGTCAAAACAGGTGTAACACCAGCTGAGAAG
 CTACTGGAATTGTACCATGGGAAGTGGGGACAGAGTGTGGATCCTATCTTTGAGGAGCTTCTCTATTGAAGTTGT
 TTGAGAGTGCCATTCAATTCTGTCTTTTTCTGTGTTGTAAATGATTTCTCAATTTCTTTTTGAGAGTTATATC
 AGCCGATTTTGGCCATCGTTAGACAAATAATCTAGCTAGCTAGATTAGATATGTTACATGTATCTACGTCTCGA
 AAACGTATATGTAGCTCTCCCTTGGATGTGATCAATCTGGTTCCAGATAGCATTTGTATACACTGGGACATAC
 ATTTTCCCCAGAAAAGTAACATAGCAATGCTAAAGTAATAAATGAATCTTGGCTGACCAA

FIG. 22

SEQ ID No. 21

>GSHA protein Soltu [UniProt ID: M1AWB9]

MALMSQAGSSHCYSEKIRICIGGHSNIMNMEMFKMREICFGVDISSRNASRRVQGNHLNLIGVGSRRGGLTIVA
 ASPPTEDAVVAEPLTKEDLVGYLASGCKTKEKWRIgtehekfgefgtlrpmkydqiadllngiaerfdwekvm
 EGDKIIGLKQKGQSISLEPGGQFELSGAPVETLHQTCAEVNSHLYQVKAVAEEMGIGFLGTGFQPKWGLKDIPIIM
 PKGRYEIMRNYPKVGSLGLDMMFRTCTVQVNLDFSSEADMIRKFRAGLALQPIATALFANSPTFEGKPNGYLSK
 RSHIWTDTDNRRAGMLPFVFDSSFGFEQYVDYALDVPMYFVYRKKKYVDCSGLSFRDFMNGKLPPIPGEYPTLND
 WENHLTTIFPEVRLKRYLEMRGADGGPWRRLCALPAFWVGILYDEGSLQSVLDMTSDWTAERDMLRNKVPKSLG
 KTPFRDGLLMHVAQDVVKLAKEGLERRGFKETGFLNEVTEVVKTGVTPAEKLLELYHGKWGQSVDPIFEELLY*

FIG. 23

17/37

SEQ ID No. 22

>XM_003569010.1_Brachypodium_distachyon_glutamate--
cysteine_ligase_A (GSH1),_
chloroplastic-like_(LOC100837682),_mRNA
TCGAGGCCGCGGACGGCGGGCGGGCGTGCCTTGGTTGGTCTTCGTCGGGTCTTGGCGCGAGCGAACCCTAGCGA
GACAGAGGGGTTGGGCGTGATGGCGGTGGCGTTCGCGGCTGGCGGTGGCGCGGGTGGTTCGCGCCGGAGGGAGGCGC
GGGCGGGAGGAGGAAGGGGCGGCCGGGGTTTGCCTTGGTTGGGTCCCCACTGCGGGGAGGAAAGGGAGCAAGCG
CGGCGGGGCTGTGGCCGCTAGTCCCCGACGGAGGAGGCGGCGCAGATGACGGAGCCGCTCACCAAGGACGACCT
CGTCGCCTACCTCGTCTCCGGGTGCAAGCCAAAGGAGAACTGGAGAATTGGCACAGAGCATGAGAAGTTTGGTTT
TGATATTGAAACGTTGCGCCCTATAACGTATGACCAGATTAGTGGCATACTGAATGGGCTGTCTGAGAGATTGGA
GTGGGATAAGATAATGGAAGAAAACCATGTCATTGGTCTCAAGCAGGGAAAGCAAAACATATCTCTAGAGCCTGG
TGGCCAGTTTGAACCTAGTGGTGCTCCTCTTGAACATTACATCAAACCTTGTGCTGAGGTCAATTACATCTTTA
TCAGGTCAAAGCAGTTGGAGAGGAAATGGGAGTTGGATTTCTTGAATTTGGTTTTCAGCCAAAGTGGGCACTCAG
TGATATACCAATAATGCCCAAGGGAAGGTATGAAATTATGAGGAATTACATGCCTAAAGTTGGTTCTCTTGGCCT
TGACATGATGTTCCGCACATGCACTGTACAGGTTAATCTCGACTTCAGTTCAGAGAAAGATATGATAGATAAATT
CCGTGTTGGCCTCGCGCTGCAGCCTATCGCAACAGCAATATTTGCTAACTCTCCCTTCAAAGAAGGAAAACCAA
TGGGTTTCTCAGTTTAAAGAACCATATCTGGACAGATACTGATAACAACCGCTCAGGGATGCTTCCTTTTGT
TGATGACTCGTTCGGATTTGAGCAATACGTGGACTATGCATTAGATGTCCCAATGTATTTTGTATACCGGAATAA
ACAATACATTGACTGTACCGGGATGTCATTTCCGGGATTTTATGGTAGGAAAGCTCCCACAGGTTCCAGGGGAGCT
ACCTACTTTGAATGATTGGGAGAACCACTAACGACAATTTTCTGAGGTTAGGTTGAAGAGATACATAGAGAT
GAGAGGTGCTGATGGTGGTCCATGGAGGAGACTGTGTGCTTTGCCTGCATTTTGGGTGGGGTTGCTATATGATGA
GGAATCACTACAAAGTATTTTGGATATGACTGCTGATTGGACAAACGCAGAAAGAGAGATGTTAAGGCGGAAGGT
ACCAGTGACTGGTTTGAAGACACCGTTCCGGGATGGTTACGTAAGAGATCTAGCTGAAGATGTTCTCAAGTTGGC
TAAGAGTGGACTAGAAAGAAGAGGATACAAGGAGGTTGGATTCCTGAGAGAGGTTGACGAGGTGGTTAGAACAGG
AGTGACGCCTGCCGAGAAGCTTCTCAACCTGTACGAGACGAAGTGGCAGCGCAGTGTGGACCCTGTCTTTCAAGA
GTTGTTATACTGATGCTGAGAAATATATCCTGTATAACTTCACCGATCACTGGATGGGAGTCTTCAGATTGCCCC
ATACAAGAACCCCATCTCCATTTCTATTTCCGAAAGCCAGGAGGGCTAACATGTTTGTCTCTGTTTCAGTCGTT
GGGAGCTTTTACAACCTTTTATTTCCCAATCTCTCCAGGGAGCTGGAACAAACCAAGTGTGTGTACCGACCCTGTT
TCTTCTTTTCTGATGAGTGCTGTATCCAAATAAAATCCATCTGGTGTATGATGGAAGGAACCTGATTTGCAGAA
AATAATTTGAAGTGTTCGCGAGTTAAAGCGCTTTATCGTGCTCCTGGCCGTTTTAT

FIG. 24

SEQ ID No. 23

>GSH1 protein Bradi [UniProt ID: I1HMP3]
MAVASRLAVARVVAPEGGAGGRRKGRPGFAVVAVPTAGRKGSKRGGAVAASPPTEEAQMTEPLTKDDLVAAYLVS
GCKPKENWRIGTEHEKFGFDIETLRPITYDQISGILNGLSERFEWDKIMEENHVI GLKQKQNI SLEPGGQFELS
GAPLET LHQTCAEVNSHLYQVKAVGEEMGVGFLGIGFQPKWALSDIPIMPKGRYEIMRNYMPKVGS LGLDMMFRT
CTVQVNLDFSSEKMDIKFRVGLALQPIATAI FANSPFKEGKPNGLSLRSHIWTDTDNNSRGM LFPVFDDSFGE
EQYVDYALDVP MYFVYRNKQYIDCTGMSFRDFMVGKLPQVP GELPTLNDWENHLTTIFPEVRLKRYIEMRGADGG
PWRRLCALPAFWVGLLYDEESLQSI LDMTADWTNAEREMLRKVPVTGLKTPFRDGYVRDLAEDVLKLAKSGLER
RGYKEVGFLREVDEVVRTGVT PAEKLLNLYETKWQRSVDPVFQELLY*

FIG. 25

18/37

SEQ ID No. 24

>NM_001061076.1__Oryza_sativa_Japonica_Group_
(Oryza)Os05g0129000_(Os05g0129000)_mRNA,_complete_cds
ATCTTCTCGCTCGCCTCTCGCTGTCTCGCACTCGCCGCCGCCGCTCGCCGCGCCTCTCTCGCCGGCGGGGAGG
TGAGGTGTACGGATTTGATGGCGGTGGCGTTCGCGGCTCGCGGTGGCGCGGGTGGCGCCGGACGGGGGAGCGGGCGG
GGAGGAGGAGGAGGAGGGGGCGGCCGGTGGTTCGCGGTTCCGACGGCGGGGAGGGGGCGAGGTGGGGCCGTGGCGG
CGAGCCCACCGACGGAGGAGGCGGTGCAGATGACGGAGCCGCTCACCAAGGAGGACCTCGTGGCCTACCTCGTCT
CCGGGTGCAAGCCCAAGGAGAACTGGAGAATTGGCACAGAACATGAGAAGTTCGGTTTTGAAGTTGATACATTGC
GTCCTATAAAGTACGATCAGATCCGTGACATCCTGAATGGACTTGCTGAGAGGTTTGATTGGGACAAGATAGTTG
AAGAAAATAACGTTATCGGTCTCAAGCAGGGAAAACAAAGCATTTCACTAGAACCTGGCGGTGAGTTTGAACCTTA
GTGGTGCTCCTCTTGAAACATTACATCAAATTGTGCTGAGGTCAATTCACATCTTTACCAGGTCAAAGCAGTTG
GAGAGGAAATGGGGATTGGATTTCTCGGAATTGGTTTTGAGCCAAAGTGGGCACTGAGTGACATACCAATAATGC
CCAAGGGGAAGGTACGAAATTATGAGGAATTACATGCCTAAAGTTGGCTCTCTTGGGCTTGATATGATGTTCCGAA
CATGCACTGTTTCAGGTTAATCTTGACTTCAGTTTCAGAGCAAGATATGATAAGGAAATTCGTAAGTGGTCTTGCA
TACAACCTATTGCAACAGCAATATTTGCTAATTCTCCCTTCAAAGAAGGAAAACCAAATGGGTATCTTAGCTTAA
GAAGCCATATTTGGACTGATACTGACAACAACCGCTCAGGGATGCTTCCTTTTGTGTTTCGATGACTCATTGGAT
TTGAGCGATATGTGGACTACGCATTAGATATCCCGATGTATTTTGTCTATCGTAATAAGAAGTACATTGACTGTA
CTGGAATGTCATTTCCGGATTTTATGGTAGGAAAACCTCCACAAGCTCCGGGGGAGCTGCCTACTTTGAATGACT
GGGAGAACCATCTAACAACAATATTTCTGAGGTTAGGCTGAAGCGATATCTAGAGATGAGAGGTGCTGATGGTG
GCCCATGGAGGAGATTGTGTGCTTTGCCAGTATTTTGGGTTGGGCTGTTGTATGATGAGGAATCACTGCAAAGTA
TTTCGGACATGACCTCTGATTGGACAAATGAGGAAAGAGAAATGCTAAGGCGAAAGGTACCTGTAACCGGTTTGA
AGACACCCTTCCGGGATGGATATGTAAGAGATCTAGCTGAAGAAATCTGCAGTTGTCTAAGAATGGATTAGAAA
GAAGAGGGTACAAGGAAGTTGGTTTCTTAAGAGAGGTTGATGCAGTGATCAGTTTCAGGAGTGACCCCACTGAGA
GGCTGCTCAATTTGTACGAGACGAAGTGGCAACGCAGCGTCGATCCTGTCTTCCAGGAGTTGTTATACTGATGCT
GAGAAACCTCTTACACAGCAAACTGGACAGTCTGTAGATGTACTCCCCATACAGAAAAAACCTCATATCCATT
CTATTTTCATGCCTCGGCAGCTAATGTCCGCCATCCGTTCTTCTTGAGCCTGAAGAGGGGAAAAATAACCGCCGTGC
AGCAATTCAGCTTCACGGACTCAATTGTTTCTGCCTCCTGGAACATTTGAGGTCGGAAGCCTTTGTAGGTGCTGT
CTTTGTTGGGGCCTTTTACAACCTCCGATTTCCCTCGTCAACAGGGAGATGGGACAAACCACTGTTTGTACCAGA
CCCTCTTTTGTTCCTGATGATTTTGTATGAAAAATAAGCCCATCTGGTGCATGATGGAACCTATATATATATAT
GTTCCAGAAAAATGTTGGAATAAGTATTTGACAAGCACGTGGAGTGAGTTCCATCC

FIG. 26

SEQ ID No. 25

>GSH1 protein Oryza [UniProt ID: B7EUG4]
MAVASRLAVARVAPDGGAGRRRRRGRPVVAVPTAGRGRGGAASPTTEEAVQMTEPLTKEDLVAYLVSGCKPK
ENWRIGTEHEKFGFEVDTLRPIKYDQIRDIINGLAERFDWDKIVEENNVIGLKQKQSI SLEPGGQFELSGAPLE
TLHQTCAEVNSHLYQVKAVGEEMGIGFLGIGFQPKWALSDIPIMPKGRYEIMRNYMPKVGS LGLDMMFRTCTVQV
NLDFSSEQDMIRKFRITGLALQPIATAIFANS PFKEGKPNGYLSLRSHIWTDTDNNSRGM LFPVFD DSFGFERYVD
YALDIPMYFVYRNKKYIDCTGMSFRDFMVGKLPQAPGELPTLNDWENHLTTIFPEVRLKRYLEM RGADGGPWRRL
CALPVFWVGLLYDEESLQSI SDMTSDWTNEEREMLRKVPVTGLKTPFRDGYVRDLAE EILQLSKNGLERRGYKE
VGFLREVD AVISSGVTPAERLLNLYETKWQRSVDPVFQELLY*

FIG. 27

19/37

SEQ ID No. 26

>NM_001111672.1_Zea mays_gamma-

glutamylcysteine_synthetase1_(gsh1),_mRNA

CCGTCGCGGCCAGCCCTCCACGGAGGAGGCGGTGCAGATGACGGAGCCGCTCACCAAGGAGGACCTCGTCGCCT
 ACCTCGTCTCCGGGTGCAAGCCCAAGGAGAATTGGAGAATTGGGACGGAGCACGAAAAGTTCGGTTTCGAAGTCG
 ACACTTTACGCCCTTTAAAAATATGATCAGATTCTGACATACTGAACGGTCTTGCTGAGAGATTTGATTGGGACA
 AGATAATGGAAAAAACAAATGTTATCGGTCTCAAGCAGGGAAAGCAAAGCATCTCACTAGAACCTGGAGGCCAAT
 TTGAACTTAGTGGCGCTCCTCTCGAAACATTACATCAAACCTTGTGCCGAGGTCAATTCGCATCTTTATCAGGTCA
 AGGCAGTTGGAGAGGAAATGGGAATAGGATTTCTTGGGCTTGGCTTTTCAGCCAAAATGGGCACTGAGTGACATAC
 CAATAATGCCAAAGGGAAGATACGAAATAATGAGGAATTACATGCCTAAAGTTGGTACTCTTGGCCTTGATATGA
 TGTTCCGGACATGTACTGTGCAGGTTAATCTTGACTTCAGTTCAGAACAGGATATGATAAGGAAATTTTCGTGCTG
 GCCTCGCTTTGCAGCCTATTGCAACTGCAATATTTGCCAATTCCTCCGTTCAAAGAAGGAAAACCAAATGGATTTTC
 TCAGCTTAAGGAGCCATATCTGGACAGATACTGATAATAATCGTGCAGGGATGCTCCCTTTTGTCTTTGACGACT
 CATTTGGGTTTGAGCAATATGTGGACTATGCATTAGAAGTCCCCATGTATTTTGTGTACCGAAATAAAAAGTATA
 TTGACTGCACCGGAATGTCGTTTCGGGATTTTATGCAAGGAAAGCTTCCACAGGCTCCTGGGGAGTTGCCTACTC
 TTACCGATTGGGAGAACCATCTAACAACAATTTTCCAGAGGTTAGGCTAAAGAGGTACCTTGAGATGAGAGGTG
 CTGATGGTGGCCCATGGAGGAGATTGTGTGCGTTGCCTGCATTTTGGGTTGGGCTGCTGTACGACGAGGAATCGT
 TACAAAGCATTTTAGACATGACTTTTGATTGGACAAAGGAGGAAAGAGAGATGTTAAGACGGAAGGTACCATCGA
 CTGGTTTGAAGACGCCGTTTTCGTGATGGATATGTAAGAGATTTAGCTGAGGAAGTTCTAAAACCTGGCCAAGAATG
 GACTGGAAAAGAGAGGGTACAAGGAGGTTGGTTTTCCTTAGAGAGGTGACGCAAGTAGTGAGAACAGGAGTGACGC
 CTGCGGAGAGGCTGCTGAGCCCCGTACGAGACCAAGTGGCAACGCAACGTCGACCATGTTTTCGAGCATTTGTTAT
 ACTGATTACTGAGGATGACGCCGCCGAACGACGTTGGACTGAGCTTGTAGATGTACCCCGTACCCTAATACAGAG
 TCTCGTCTTGATTCATTCAATGCCTGAGCCGGAAGGGCTCAGTTTCAGGCACACAGGAGCAGAGCTTTTGGTTGG
 TTGTTGACGGGGGGCAACCGCTGCACAGCCTGGTGTGTGCCTACAAATAACTTCGCCTGGAACCTGGAACGGATC
 ACTCATCACTGTGATTTGCTTTGATTTGATCTGCCTGTCTCTCCAGGGAGATGGAACAACTACTGTTTGTACCA
 AAAAAAAAAAAAAA

FIG. 28

SEQ ID No. 27

>GSHA protein Zeama [UniProt ID: Q8W4W3]

MTEPLTKEDLVAYLVSGCKPKENWRIGTEHEKFGFEVDTLRPLKYDQIRDI LNGLAERFDWDKIMEKNNVIGLKQ
 GKQSI SLEPGGQFELSGAPLET LHQTCAEVNSHLYQVKAVGEEMGIGFLGLGFQPKWALSDI PIMPKGRYEIMRN
 YMPKVGT LGLDMMFRTCTVQVNLDFSSEQDMIRKFRAGLALQPIATAI FANSPFKEGKPNGFLSLRSHIWTDTDN
 NRAGMLPFVFD DSFGFEQYVDYALEVPMYFVYRNKKYIDCTGMSFRDFMQGKLPQAPGELPTLTDWENHLTTIFP
 EVRLKRYLEMRGADGGPWRRRLCALPAFWVGLLYDEESLQSI LDMTFDWTKEEREMLRRKVPSTGLKTPFRDGYVR
 DLAEVVLKLA KNGLERRGYKEVGFLREVDEVVRTGVTPAERLLSPYETKWQRNV DHVFEHLLY*

FIG. 29

20/37

FIGURE 30

SEQ ID No. 28

>XM_001757468.1__Physcomitrella_patens_subsp._patens_predicted_protein_(PHYPADRAFT_70546)_mRNA,_complete_cds
ATGAGGATCACGGGCATAGGGAAGGGGAGGTGATTGGAAGTCCGCCGAGTAACGATGCTGGGAGCTACGGTGGG
GAGCCGCTCACCAGGCAGGACTTGGTTGGGTATTTGGCGTCTGGGTGCAAACCTAAAGAGAAATGGAGGATTGGA
ACGGAGCACGAGAAATTTGGGTTCCAGCTAGACAATCTACAGCCTATGACGTATCCTCAAATTAAGCAGCTTCTG
GAGGGGCTGGCGGATCGTTTTGAGTGGCAAAGAGTGATGGAAGGGGACAACATAATTGGACTGACTCTTGATGGT
CAAGGTGTCTCGTTGGAACCTGGAGGTCAGGTCGAATTAAGCGGTGCTCCTCTGGAGACCCCTCACCAAACATGT
GCAGAGGTCAATTCGCATCTATATCAGGTGAAGGCTGTCGCAGAGGAGCTCGGCCTCGGTTTCGTAGGCATTGGA
TTCCAACCCAAGTGGAGCGTGGCAGATACTCCCATTATGCCAAAGGGCAGGTATGAGATAATGAGGAATTACATG
CCGAAAGTTGGATCGTTCGGCCTGGACATGATGTTCCGCACATGCACGGTTCAGGTCAATTTGGATTTTCGATTCT
GAGGCGGATATGGTTAGAAAAGTTTCGTGTGGGGCTTGCTTTACAGCCTATCGCTACAGCCCTTTTCGAAATTCC
CCTTTCAAGGAAGGAAAACCAAACGGCTTTTTTAAGCTACCGAAGTCACATCTGGACAGACGTGGACAACAATCGG
GCCGGAGATTTACCATTTGTGTTTGAAGATGGCTTTGGGTTTGATAAGTATGTTGAATATGCACTAGGCGTGCCT
ATGTACTTTGTTTATCGGAATGGTCCGTACATTGACTGTTCTGGTATGTCGTTCAAGGATTTTCTGGAAGGGAAG
CTATCTGCTCTGCCTGGAGAGAAGCCAACAATGAACGATTGGGAAAATCATTTAACAACAATCTTTCCCGAGGTG
AGATTGAAACGGTACTTAGAGATGAGGGGAGCTGATGGTGGTCCATGGCGCCGATTGTGTGCTTTGCCTGCATTC
TGGACTGGCTTGTGTATGACGAGGAATCACTTCAAGGTGCCTTGGACATCATTAATGACTGGACCGAAAGGGAG
CGTGCGATGCTTCGCTCTAAGGTTCCGAAGCTGGGTTTGAATGTTCCCTTCAGGGATGGACTTCTGAAACATGTG
GCGCAGGATGTCTTCAAACCTGCCAAGGATGGGTTGACTCGGAGAGGTCTTAACGAGACTGGTTTCTTAAAAGAT
ATCGAGGAGACTGTGAGAACAGGAAAACTCCTGCAGAACATATCTTGGAGCTCTATCATGGGAAGTGGGGTGAA
AATGTTGATCCCGTCTTTGAGAAGCTTCTGTATTAG

FIG. 30

SEQ ID No. 29

>GSHA protein Phypa [UniProt ID: A9RU07]

MRITGIGKGEVIGSPPSNDAGSYGGEPLTRQDLVGYLASGCKPKEKWRIgteHEKFGFQLDNLQPMTYPQIKQLL
EGLADRFEWQRVMEGDNIIGLTLDGQGVSLPPGGQVELSGAPLETLHQTCAEVNSHLYQVKAVAEELGLGFVGIG
FQPKWSVADTPIMPGRYEIMRNYMPKVGSGFLDMMFRTCTVQVNLDFDSEADMVRKFRVGLALQPIATALFANS
PFKEGKPNGFLSYRSHIWTVDVNNRAGDLPFVFEDGFGFDKYVEYALGVPMYFVYRNGRYIDCSGMSFKDFLEGK
LSALPGEKPTMNDWENHLTTIFPEVRLKRYLEMrgADGGPWRLCALPAFWTGLLYDEESLQGALDIINDWTERE
RAMLRSKVPKLGlnVFPFRDGLLKHVAQDVFKLAKDGLTRRGLNETGFLKDIEETVRTGKTPAEHILELYHGKWGE
NVDPVFEEKLLY*

FIG. 31

21/37

SEQ ID No. 30

>XM_001701595.1__Chlamydomonas_reinhardtii_gamma-glutamylcysteine_synthetase_(GSH1)_mRNA,_complete_cds
 GAGTGTCACTTACTACTTGGATAAGAATCACCAGCACTTGGGCGACCACAGCGAGCGAACGATTATTCTTCACTGTT
 ACGTTGCACTACATCTTGGCCCTGTTCCGATCAATCCCTAGGCTGCCGCACTAGTGCGATGGCTCTCGCCTCAGG
 CGTTGGCCGTCGCCAGCATGTGTCGGCCTCGCCCTCGCGCAGTCGGGGTGTGCCAAGCCACGCTTGAGCCCTGT
 CCACGCGAACCGCGCTGCGGTGCGGAGCGTCGCACAGAGCCCCCTCTTAAAGCAGGAGCTGGTGGATTACCTGAA
 ATCGGGGTGTAGGCCTCGCAGCGCGTGGCGAATCGGCACCGAGCACGAGAAGCTGGGTTTCAACCTGGCAGACAA
 CAGCCGCATGAACTACGACCAGATTGCACAGGTGCTACGCAAGCTGGAGGCCCGGTTTGGTTGGGAGCCCATCAT
 GGAGGAGGGCCGTATCATCGGCGTGCACTTGGATGGTCAGAGTGTGACGCTGGAGCCCGGCGGCCAGTTTGAAC
 GAGCGGGGCGCCGTGGAGACCATTACAAGACGTGTGCGGAGGTGAACAGCCACCTCTACCAGGTCAAGGCCAT
 CTGCGAGGAGCTTCAGACAGGATTCTTGGGCGTGGGCTTTGACCCCAAGTGGGCCATCAGCGACGTTCCCATGAT
 GGGCAAGGGCCGCTACAAGCTGATGAAGTCGTACATGCCCACGGTGGGCTCCATGGGCTGGACATGATGTTCCG
 CACATGCACCGTGCAGGTGAACCTGGACTTTGAGAGCGAGCAGGACATGGTGGAGAAGTTCGCGATCGGCCTGGC
 GCTGCAGCCCATCGCCAACGCGCTCTTCGCCAGCTCGCCATTCAAGGAGGGCAAGCCACCGGGTACCTGAGCAC
 CCGCGGTACAGTGTGGACGGACGTGGACGCCCTCGCGCACCGGCAACCTGCCGTTCTGTGTCGAGAAGGACATGTG
 CTTTCGAGAGCTACGTGGACTACGCCATGGCGGTGCCCATGTACTTCGTGTACCGCAACGGGCAGTACATCAACGC
 GCTGGGCATGAGCTGGAAGGACTTCATGGCCGGCAAGCTGCCGCGCTGCCGGGCGAATACCCACCATCGCCGA
 CTGGGCCAACCACCTGACCACCATCTCCCCGAGGTGCGGCTCAAGAAGTTCCTGGAGATGCGCGGCGCGGACGG
 CGGCCCTGGCGCATGCTGTGCGCGCTGCCGGCGCTGTGGGTGGGGCTCATATACGATCCGGAGGCGCAGCGCCA
 GGCCCTGGCGCTGATTGAGGATTGGACGCCCGCGGAGCGCGACTACCTGCGCACCGAGGTGACCCGCTTCGGCCT
 GCGCACGCCCTTCCGCGCCGGCACCTGCGAGGACGTGGCCAAGCAGGTGGTGTCCATCGCGCACGGCGGCCTGGA
 GCGGCGAGGCTACGACGAGACGTCCTTCTCAAGCGCCTGGAGGTTCATCGCGGAGACTGGCCTCACACAGGCCGA
 CCACCTGCTTGAGCTGTACGAGACCAAGTGGCAGCGCTCGGTGGACCCGCTGTACAAGGAGTTCATGTACTGAGC
 AGGAGGAGGGGCGAGCGGCAGCAGCAGCAGGAGGTGGCGGCGGAGCAGGAGGAAGTGTATCGGTATGTCGAAAGA
 CGACGTTAAGTGACTTGAAAAGTGATGTTTGCAGAGGCTGATGCCTGTGCCGTTTGTATGTGAGTGTGAGTACTGC
 GCCAGTATTCGTGTTTCGCAACGGCGTTACCAACATATTGCGGTGGGATGCATAGTTGATCAGTTTGGTTGTGCC
 AACCAACATTTCTCGCTTGGGCTGATGCATGGGCGAAGTGGCGAATGCCGTGGGTGGCGCTGGGGCTCAGCAACA
 TGAAGCTTGTGTGTGTACAGCCAGACCAGCTTGAAGGAAGACAGCTTGAAGGAATGAGCTCGGCCTCGAGCGT
 TGTGACAATCGTGCTGTGCTGCGGACAACATATTATGTTGGTGTGCCCTCCGATGTTATGAAGTTGCTTCGCGTG
 GGATGCGAACGCAGGTACAGGTAGAGACCTCAAACGGTGACAGGGACGGGTCTGGGGCGGAAACCTATGGGTATC
 AGGCGATCTGATCAAGAGGTGGATCCGCTTGGGACATAGCACGCGTGCCTGCTACGATGAGGCAATCGGGCCTGT
 GCCCCGGTTCTCAGGCAGTCAGGCGCCGGGTAGTGCGGCGATGAAGTACGGTATAGTGCCTACTCAGTGCATC
 GGGCGGCGGTGCGGATATGGGCTTCACCCGTGCGTTGCGTGAAGCTTTTGGGGTCTGTGGCTTCGCTGTGCGGC
 GTGCCAGACGAGGCTCCGAGATGGTGCTGAACTGGTTGTGAAGGTGGCCGGTTCGGACTTGTAAGGCCCTCGCTC
 AGG

FIG. 32

SEQ ID No. 31

>GSHA protein Chlre [UniProt ID: A8IA77]
 MALASGVGRRQHVSASPSRSRGVPSRLSPVHANAPAVAERRTEPLLKQELVDYLKSGCRPRSAWRIGTEHEKLG
 FNLDNSRMNYDQIAQVLRKLEARFGWEPIMEEGRIIGVQLDQSVTLEPGGQFELSAPVETIHKTCAEVNSHL
 YQVKAICEELQGTGLGVGFDPKWAISDVPMPKGRYKLMKSYMPTVGSMLDMMFRTCTVQVNLDSESEQDMVEK
 FRIGLALQPIANALFASSPFKEGKPTGYLSTRGHVWTDVDASRTGNLPPVFKEKDMCFESYVDYAMAVPMYFVYRN
 GQYINALGMSWKDFMAGKLPALPGEYPTIADWANHLTTIFPEVRLKKFLEMARGADGPPWRMLCALPALWVGLIYD
 PEAQRQALALIEDWTPAERDYLRTVTRFGLRTPFRAGTVQDVAKQVVSIAHGGGLERRGYDETSFLKRLEVIAET
 GLTQADHLLLELYETKWQRSVDPLYKEFMY*

FIG. 33

22/37

SEQ ID No. 32

>XM_001418275.1__Ostreococcus_lucimarinus (Ostlu)CCE9901_
gamma-glutamate_cysteine_synthase_(GSC2)_mRNA,_complete_cds
ATGACGGCGACGACGGCGCGCGGGCGACGCGCTCCGCGTCGGCACCGCGACGCGGGCGGGGAGACGCGCGACGCGA
CGCGCGACGACGACGACGATTTCGCGCGGGCGGGCGACGGCGCCGGCGCGCGCGGTCACGCGCGAGGAGTGCCTGGCG
TTCGTCAGGCGGGGGGTGTAAACCGAGAGAGGCGTTTCAGGATCGGGACGGAGCACGAAAAGTTTGGGTACGACGAG
GCGACCAAGCGAAGGATGGATTATCCAGTCGTGCGGCACGTCTTGAGCTCGCTCGTCGAGCGACACGGGTGGACG
CCGATAAACGAAGCGGATAACATCATCGGGTGCACGAAGGACAAGCAAAGCGTGACGCTCGAGCCGGGTGGACAG
TTTGAACCTCTCGGGCGCGCCGCTGGAGAACCTGACGCGACGGACGAGGAGGTGCGATGGCACATCGAGACCGTG
AACGCGCTGGCGCGGGAAAAAGGGGCACAGATTCTTGGGAATCGGGTTTGACCCGAAGTGGTGGTGGCCGAGGTG
CCGATGATGCCGAAAGGGAGGTATAAGATTATGCGCGCGTACATGCCGAAGAAGGGCACTCGCGGACACGACATG
ATGTTTCGCACGTGCACGATTCAAGTCAACTTGGATTTCGAGAGCGAGCAAGACATGGTGCACAAGTTTCAGGACG
TCCCTCGCGCTGCAACCTCTCGCGACGGCGATGTTTCGCCAACTCCGCGTTCGTGGACGGCGCGGACACCGGGTAT
CAATCCATGCGCTCCGACGTGTGGCGGGACACGGACGACGATCGCACGGGCACGCTCGCGTGGGTTTTTCGACGAC
GATTTTCGGGTTTTGAAAAGTACTGCGACTACGTCATGAACGTACCGATGTACTTTGTGTACAGAAACGGCACGTAC
GTGGACGTTTCGGGGGAATCCTGGCTCGATTTTATGGAGGGAAAGCTTCCTCAACTTCCGGGCGACAAGCCGACG
ATCGACGACTGGGAGCAACACCTCACGACTGTGTTTTCCGGAGGTTTCGTCTCAAGAAGTACATGGAGATGCGAGGT
GCCGACGGTGGCTCTTACGAAGCCATCATGGCCCTTCCCGCGCTTTGGGTTGGATTGTTGTATTCCGACAAAGCC
TTGGACGCGGCCGAGGCGATGGTGAGCGACTGGACGCAAGTCGAGCGCGACGCGCTTCGCGTCGACGTCACCAAG
GATGGATTGTGCGCCAAGTTTAGAGGCGGCACGGCGAAGCATATCGCCAAGAAAATGGTCAACTTTCTGTACAG
GGCCTGCAAGAGCGCGGTTTAGGCGAAGAAGTGTACTTGCCTTACTTGCAAGACATCGTCGACGACGGTAAGAGC
GTCGCGGGCGCGCATGTCCGAGATGAACGCGCGCGAGTGGAACGGTGATCTCGAAAAAGTTTACGAGTACGCCACC
TTCCAGACACGCTTCCGAAGATCGTTTTGA

FIG. 34

SEQ ID No. 33

>GSHA protein Ostlu [UniProt ID: A4RZD5]
MTATTARGDALRVGTATRRGRRATRRATTTTIRAAATAPARAVTREECVAFVRRGCKPREAFRIGTEHEKFGYDE
ATKRMDYPVVRHVLSSILVERHGWTPINEADNIIGCTKDKQSVTLEPGGQFELSGAPLENLTQTDEEVRWHIETV
NALAREKGHRFLGIGFDPKWSVAEVPMPKGRYKIMRAYMPKKGTRGHDMFRTCTIQVNLDFESEQDMVRKFRT
SLALQPLATAMFANSFAVDGADTYQSMRSDVWRDTHDDRTGTLAWVFDDDFGFEKYCDYVMNVPMYFVYRNGTY
VDVSGESWLDMEGKLPQLPGDKPTIDDWEQHLTTVFPEVRLKKYMEMRGADGGSYEAIMALPALWVGLLYSDKA
LDAAEAMVSDWTQVERDALRVDVTKDGLSAKFRGGTANDIAKKMVELSVQGLQERGLGEEVYLRYLQDIVDDGKS
VAARMSEMNAREWNGDLEKVYFYATFPDTLPKIV*

FIG. 35

23/37

SEQ ID No. 34

>XM_002872228.1__Arabidopsis_lyrata_subsp._lyrata_glutathione_synthetase (GSH2),_mRNA

AAATCTTAGCGATCTTTGTTTGACTGCGAGCGAATGGGCAGTGGGGGCGGCAGTGTCTGCTCTTCTCTCTTAC
TCTTCTTCTTCCACTTGCTTTGCTTCTCTTTCTCTATTTGTTCTCCTTCATCTTCTTCGCTGAAACTCAATCCA
AGACGCTTCTTTTCCATAACCTTACAAATCTGAGAAAGCAGGCACCTTTGAGGTGTGGGAGATCCCCGGAGATG
GAATCAAAGAAACCCATTTTCGATGTGGAACGTTTGATGATGACTTTGTTTCAGAAACTGGTTTATGATGCTCTC
GTTTGGTCTTCTCTTACGACTCGTCGTGCGTGACAAAACCTTACCAGAAATCAGGAAATGTTCCAGGGGTTGGA
TTGATGCACGCACCGATTTCATTGCTACCAACTGCATTCCCTGAAATTTATTGGAAGCAAGCTTGTGATGTTACT
CCTATTTTCAATGAATTGATTGATCGTGTTAGCTTGGATGGGAAGTTCTTACAGGATAGTCTCTCTAGAACGAAA
AAAGCTGACGTCTTTACATCTAGACTTCTCGACATTCACTCCAAGATGCTTGAAAGAAACAAGAAAGAGGACATT
CGTCTGGGTTTGCATCGGTTGGATTATATGCTTGACGAAGAAACAAATTCACCTTCTTCAGATTGAGATGAACACT
ATCTCTAGTTCGTTTCCAGGCCTTAGTCGTCTTGTAGCCAGCTACACCAGTCATTGCTTCGATCTTATGGGGAT
CAGATTGGCATAGACTCTGAACGTGTACCTAGAAATACATCCACAATCCAATTTGCTGATGCAATAGCTAAAGCT
TGGTTGGAGTACAATAACCCAAGAGCGGTTGTCTGTTAACTGTACAGCCAGAAGAACGCAACATGTACGATCAA
CATTTGCTGAGCAGGAACACGCACAATATAGTAGCCATCAGGAAGACTCTAGCAGAAGTCGAAACTGAAGGGAGT
GTACAAGAGGATGGAACCTTATTGTTGGCGGCCAAGCAGTCGCAGTTGTTTATTTTCAGATCTGGCTATACTCCC
AATGATCATCCGTCTGAATCAGAATGGAATGCTAGGCTGCTTATTGAGGAGTCTTCAGCGGTCAAATGCCCGAGT
ATAGCTTACCATTTAACTGGCAGCAAGAAAAATCCAGCAAGAACTGGCAAAACCAGGTGTTCTCGAGAGGTTTCTG
GACAACAAAGAGGACATTGCTAAGCTGAGGAAATGCTTTGCGGGGCTTTGGAGCTTGGATGACCCAGAAATCGTC
AAGCAGGCAATCGAAAAACCCGGACTGTTTGTATGAAGCCTCAGAGAGAAGGCGGAGGAAACAACATCTATGGA
GATGATGTGAGGGAAAATCTTTGAGACTGCAGAAAGAAGGAGAGGACGGAACGCTGCGTATATCCTGATGCAG
AGGATATTTCCAAAGGTCTCAAACATGTTCTTGGTGCGAGAAGGCGTTTACCATAAAAATCAAGTTATATCAGAA
CTCGGTGTCTATGGTGCTTACCTCAGGAACAAAGACAACGTTATAGTTAACGAGCAGAGTGGTTATCTAATGCGC
ACAAAGATCGAGTCATCAGATGAAGGCGGCGTTGCCGCTGGTTTTGTCAGTATTGGACAGCATTTATCTGAATTGA
GATGGATCACATTACAAACCATCTGATCAATTTTTCTGCTGGTTCGTTCCCTTCTTGTCTTCTTCTTCTGCTTGT
AAAGATTTGTGTATTAATTCATGAAGGTCAATAAGACTCTAGGCAAAGCTCTTACAATTCCTTCCATAGGACCGC
AAAGCATATATTGCTTATCGTTATTTTAGACCGTGTGAGACGATCAGTCGTTTAAGATCATGAATGGAATAAGA

FIG. 36

SEQ ID No. 35

>GSH2 protein Araly [UniProt ID: D7M652]

MGSGGGSVCSSLSYSSSSSTCFASLFSICSPSSSSSLKLNPRRFLFHNPTNLRKQAPLRCGRSPEMESKKPIFDVET
FDDDFVQKLVDALVWSSLHGLVVGDKTYQKSGNVPVGLMHAPI SLLPTAFPEIYWKQACDVTPIFNELIDRVS
LDGKFLQDSLRSRTKKADVFTSRLLDIHSKMLERNKKEDIRLGLHRLDYMLDEETNSLLQIEMNTISSSFPLSRL
VSQHLQSLLSYGDQIGIDSERVPRNTSTIQFADAIKAWLEYNNPRAVVMVTVQPEERNMYDQHLLSRNTHNIV
AIRKTLAEVETEGSVQEDGTLIVGGQAVAVVYFRSGYTPNDHPSESEWNARLLIEESSAVKCP SIAYHLTGSKKI
QQELAKPGVLERFLDNKEDIAKLRKCFAGLWSLDDPEIVKQAI EKPLFVMKPQREGGGNNIYGDDVRENLLRLQ
KEGEDGNAAYILMQRI FPKVSNMFLVREGVYHKNQVISELGVYGAYLRNKDNVIVNEQSGYLMRTKIESSDEGGV
AAGFAVLDSIYLN*

FIG. 37

24/37

SEQ ID No. 36

>XM_006433130.1__Citrus_clementina_(CICLE_v10000728mg)_mRNA,_complete_cds GSH2

ATGGGTGTTTCGCTATTCTTCCCTCTCTCTTTGCAGTCCCCTCATCAAAATCACCAATCGTTTTCGTCTGGCCACC
TTTACTTGTGATACAAAGCAAAAATCAATTTTTTTCTTTCTATCCTACTGATAAAGATAGGATCTTTAACACCCAT
TATCGAAGAGTTTCATTTTTTTCCCATAACCCAGTTCTCAATTTGTGGTTTGCAATTCAAGAATTGGAAATTTGAAG
TGTGGCAGAGTGAAAGAGATGGAGACTCGACAGGGAATTGCAACCAAAGAGGCTTTGGATTTGCAGAGTGTTGAT
CAAGATTTGGTTCAAAAGATGGTTTATGACGCTCTCGTGTGGAGCTCCCTTCATGGTCTTGTGTTGGTGACAGG
AGTGTGCAGAGATCAGGGAAAGTTCCTGGTGTGGATTGGTGCATGCCCCATTTGCCTTGTGCTGTGTCATTC
CCTGAAAGTCATTTAGAGAGGATGTGAGTTAGCCCCATTTTCAATGAACCTCGTCGATCGTGTGAGTTTGGAT
GGAAATTTCTACAGGAGTCTCTATCCAGAACAAGAAAGTGGATCCTTTTACCTCTAGGCTCCTGGATATTCAT
TCAAAGATGCTAGAACTCAATAAGAAAGAGGACATACGCTTGGGTTTACATCGTTGAGATTATGCTTGATGAA
CAAACTAAATCGCTCCTCCAAATAGAGGTTAACACAATTTTCATCTTCTCTGGCCTCAGCTGTCTTGTGAGT
GAGCTTCACAGAAGCTTACTCATTTCACTATGGGGAACGCCTCGGATTAGATTCAAAAAGGTCCCTGGCAACTCT
GCTGTCATTCAGTTTGAGAGGCTATGGCTAAAGCTTGGACTGAGTATAATAATGCTAGGGCTTTGGTTCTGGTT
GTGGTTCAGAGTGAAGAACATAATATGTATGACCAACATTGGCTTAGTTCTGTACTGAAGGAAAGACATAATGTT
ACAACTATTCGGAAGACATTGGCTGAAATTGATGCAGAAGGGGAACACTCCCTGATGGAACACTTCTTGTAGGT
GGCCAAGTAATTGCAGTTGTTTATTTTAGAGCTGGGTATGCACCAACTGATTATCCTTCAGAATCAGAATGGAGA
GCTAGATTACTAATGGAGCAGTCTTCTGCTGTCAAGTGTCCATCAATATCTTATCATTTGGCAGGCACCAAGAAA
ATTCAACAAGAAGCTTGCAGAAACCCAATGTACTTGAGAGGTTCCCTGACAATAAAGAGGATATTGCCAAAGTACGT
AAATGCTTCGCAGGATTGTGGAGTTTGGATGATTCTGAGAATGTCAGCAAAGCAATTGAAAACCCAGAGCTATAT
GTCATGAAGCCCCAAAGAGAAGGCGGAGGGAACAATATTTACGGTGAAGATGTGAAGGAAACCTTCTTAGATTG
AAAAAGGATGGGACTGACGAAGATGCTGCTTACATCCTTATGCAAAGGATATTTCTAGTGTATCCCTCACATTT
TTGATGCGTGATGGCATTTGTGATAAGGATCATGCCATATCTGAACTGGGATATATGGTGCTTACTTAAGGATC
AAGGAGAAGGTGATCATGAATGAACAAAGTGGTTATTTGATGCGGACGAAAGTTGCTTCTTCAAATGAAGGAGGG
GTAGCAGCTGGATTTGCGGTCTTGGACAGTTTATATTTAACTTGATTAACTAGCTTCCTCGCTACTCCCACTGA
CCCACATGATTGATGATATTATATCATAACACAACCATTACATGCTCTCTCACAGGATCTAAACGTTGAAAGAT
TCAGTAAGAGCTCCAGCTTGCTCAATTAATTGACATTCAATCTATGGAAGGAAGAGATACACAGAGAAGTTTCT
TCTATAAAGACCATACATTTTCCATTTTACGAAAGCATTGAATTTGTAATGTGGGTATCGTGGCATGCCTTTTG
TAATTTTTTTTTTTGGGTAATAAATGATAGTGCTTTCCAGTGTCTG

FIG. 38

SEQ ID No. 37

>GSHB protein Cicle [UniProt ID: V4SFK5]

MGVRYSSLSLCSPHQNHQSFRLATFTCDTKQKSIFSFYPTDKDRIENFTHYRRVSFFPITQFSICGLHSRIGNLK
CGRVKEMETRQGIATKEALDLQSVDQDLVQKMYDALVWSSLHGLVVGDRSVQSRGKVPVGLVHAPFALLPVSF
PESHFREACE LAPIFNELVDRVSLDGKFLQESLSRTKKVDPFTSRLLDIHSMLELNKKEDIRLGLHRSYMLDE
QTKSLLQIEVNTISSFSGLSCLVSELHRSLLIHYGERLGLDSKKVPGNSAVIQFAEAMAKAWTEYNNARALVLV
VVQSEHNMYDQHWLSSVLKERHNVTTIRKTLAEIDAEGELLPDGTLVGGQVIADVYFRAGYAPTDYPSESEWR
ARLLMEQSSAVKCPISYHLAGTKKIQQELAKPNVLERFLDNKEDIAKVRKCFAGLWSLDDSENVSKAIENPELY
VMKPQREGGNNIYGEDVKETLLRLKKDGTDEDAAYILMQRIFPSVSLTFLMRDGIKHKDHAISELGIYGAYLRI
KEKVIMNEQSGYLMRTKVASSNEGGVAAGFAVLDSLTYLT

FIG. 39

[illegible]

FIG. 40

MAASAFGGGIRFSKTIPLSKYIPPLQPHHNTFFSFFPSSSSSSSSSSSSSINFSTKPLPMSHHKHLTINGDEQ
NIIKPSPLFVDDYHNIDQQLLDNIAVDALVWASLHGLLMGDKSYQKSGMVPVGLMHAPFSLFPTSFPESKWKQA
CDLAPIFNELVDRVSLDGKFLQESLSRTKQVDEFTSRLLDIHSMKLQLNKKEEIRFGLHRSDYMLDEQTKSLQI
ELNTISSSFAGFSTLVTQLHRYIILTCHGKLLGLDSEKVPANNAVSNAAEALAKAWSEYNNPRAVIMTVVQAEERN
MYDQHFLSAVLRDKYPYHSLNTHGITIVRKTMAEVDQEGEILPDGTLVNGQAVAVIYFRAGYTPVDYPSESEWR
ARLLMEQSSAVKCPSSISYHLVGTKKIQQELAKPGVLERFLENKDDIAKMRECFAGLWSLDDSDIVKKAIERPELF
VMKPQREGGGNNIYGDAVRDTLILKQKTGSQEDAAYILMQRIFPNISAAVLMRNGGCHKDHAISELGI FGTYLRN
KDRVMMNNOSGYLMRTKISSSDEGGVAAGFAVIDSVYLT*

FIG. 41

26/37

SEQ ID No. 40

>NM_001247085.2__Solanum_lycopersicum_glutathione_synthetase (GS H2), _mRNA

GGCCAACCTCAGAACAAGATTCCCTCGCCTTTAGCAATTAGCAGCCAGCCAAACGTAAAGAGCCACTTTGGACGAT
AATAAGGTAGTAACATAAAAAGATTGGTCCAAAGAATGTCAAAATGGGCAGCGGCTGTTCTTCTCCATCAATCTCA
CTAACTACCATCGCTACCTCTCATTTTCAATCCCAAGAATCACTATCAAACCTCTTTAAATTTCTACTCACCTACT
AGATTTTTTGGAAACCCCATCTCCTTAAATCTTCCAAGATTTTCATACCCAAATCCCCATTGAAGTGTGCTAAAGTG
CCCGAAATGCAGACCCAACTGGAAAGATTCTGCGAAACCCATTGTGATCCTCATGATATCGACTCAAATTTGGTG
CAGAAGCTTGCTAATGATGCACTTGTCTGGTGCCCTCTTCGTGGACTTCTTGTTGGTGACAGAAATTCGGAGAGG
TCAGGAACCTATTCCTGGTGTGATATGGTTCATGCTCCAGTTGCTCTCATACCAATGTCATTTCTGAAAGTCAT
TGGAAGCAAGCTTGTGAAGTTGCTCCTATTTTCAATGAACCTCGTGGATCGCGTGAGTCAGGATGGAGAATTTCTG
CAGCAATCACTGTCTAGGACGAGAAAAGCTGATCCTTTACCTCCAGATTACTAGAAATCCACTCCAAGATGTTA
GAAATCAACAAGCTAGAGGAAATACGCTTGGGGTTGCATCGCTCAGACTATATGCTGGATGAGCAAACGAAATTA
CTTTTGCAAAATAGAGCTTAACACAATCTCATCATCATTTCCAGGGCTAAGTTGTCTGGTCAGCGAGCTTCACAGG
AGCTTGCTTCAACAATACAGAGAAGACATTGCATCAGATCCTAACAGAATTCCTCGAAACAATGCTGTAAATCAA
TTTGCTGAAGCATTGGCTAAAGCTTGGAAATGAATATGGTGATCCAAGGGCTGTGATTATTTTTGTGGTTCAAGCT
GAAGAGCGCAATATGTATGACCAGCATTGGCTTTCTGCATCATTGAGAGAAAGACATCAAGTTACAACCTATCAGA
AAGACGCTGGCTGAAATTGATGCACTTGGCGAGCTTCAACAAGATGGTACCTTGTGTAGATGGTCAAGCTGTG
GCAGTCATTTATTTTAGAGCTGGCTATGCACCAAGTGACTATCACTCGGAATCTGAGTGGAAAGCTAGGCTGCTG
ATGGAGCAATCACGTGCAGTCAAGTGCCCTTCAATTTCTTATCATTTAGCTGGCAGCAAGAAGATTGAGCAAGAG
CTTGCAAAGCCCAATGTACTTGAAAGGTTTCTGGAAAACAAAGATGACATTGCCAAACTACGGAAATGCTTTGCA
GGGTTGTGGAGCTTGGATGAATCCGACATTGTCAAGGATGCAATTGAGAGGCTGAATTGTATGTTATGAAACCA
CAACGAGAAGGAGGAGGAAACAATATCTATGGGGAGGATGTGAGAGGTGCTCTTTTGAAACTGCAAAGGAAGGC
ACTGGAAGTGATGCAGCCTACATACTAATGCAGAGGATTTTCCGAAGATCTCTCACTCGATACTAATGCGAGAA
GGCATCTCTCATAAAGAAGAAACCATATCAGAACTTGGGATATATGGTACCTATTTAAGGAACAAAACGGAAGTT
TTAATCAACCAACAGGCTGGTTACTTGATGCGAACAAGGTCTCTTCATCAGATGAGGGGGGAGTTGCAGCCGGG
TTTGCAGTTTTGGACAGTATATACTTGGTTTTGATGGAATTGTCTTGCGCGTTGACACTATAACAATGGTTGAATG
CAGAAGCTGGACAAGTCATTTCTTACATTATGTTGAGACACCATGCTTTAGCCTGTTTAACCTTTCTTGTTTTA
CATTGTATTTTTCTTGCTTAACATCTTTAATAAATAGAGCTCCACATCTGAAGGAAGTGACCAAAGAATGGTAT
AGTCAATTCAAAGTGTTTGCTTGCAATTTTAACCTATTGAGTGTGTTGACTCTTTGTTACCACAGGAGCTTTTAC
TGTTCTGTTTATTACATGTAATTCTGTTTCACTTTTGTTAGTACTATCAAGATTGCTCACCTGCATTTTTGTTTT
CCCCCTTGTGATCTTATTAGGGGATGTGTGAAATCTGTATTGCTTGAAAAA

FIG. 42

SEQ ID No. 41

>GSHB protein Solly [UniProt ID: 022494]

MGSGCSSPSISLTTIATSHFQSQESLSNSLNFYSPTRFLEPHLLKSSKIFIPKSPLKCAKVPQMOTQLEDSAKPI
VDPHDIDSKLVQKLANDALVWCPLRGLLVGDRNSERSGTIPGVDMVHAPVALIPMSFPESHWKQACEVAPIFNEL
VDRVSQDGEFLQQSLSRTRKADPFTSRLLLEIHSKMLEINKLEEIRLGLHRSYDMLDEQTKLLQLIELNTISSFP
GLSCLVSELHRSLLQQYREDIASDPNRI PANNAVNQFAEALAKAWNEYGDPRAVII FVVQAEERNMYDQHWLSAS
LRERHQVTTIRKTLAEIDALGELQQDGTLVVDGQAVAVIYFRAGYAPSDYHSESEWKARLLMEQSRVVKCPSISY
HLAGSKKIQQELAKPNVLERFLENKDDIAKLKCFAGLWSLDESIVKDAIERPELYVMKPQREGGNNIYGEDV
RGALLKLQKEGTGSDAAYILMQRI FPKISHSILMREGISHKEETISELGIYGTYLNRKTEVLINQQAGYLMRTKV
SSSDEGGVAAGFAVLDSIYLV*

FIG. 43

27/37

SEQ ID No. 42

>XM_003575715.1__

Brachypodium_distachyon_glutathione_synthetase,_chloroplastic-like_(LOC100833508),_mRNA (GSH2)

ATGCCCTCCTCCTGCGTCTCCTCCCCCTCCTCCTCACCGCCACCAATGCGTCGGCCGCTCCTCGCTCCCCC
GTCCCGCGGGCGAGCCTCGCCGCCGTGCGGCCCTCTCTACGCCGAAGCTTCGTGTCCCTGAGGGCGACGAGCACC
GAGGCCGCGGGCGGGCCCGGACGTGGCGCCGACAGGGGATGCTGGGAAGGCGGCGGGCGGTGCGGGCGGAGATGGCG
GAGGCCGCGGGCCGTTTGGTGCGCCTTGACGCGGCTCGTGTGCGGCGACCGCGCCGACCCGAGATCGGGAACAGTC
CCAGGTGTGCGCCTGGTTTCATGCTCCATTTTCGCTACTTCCGGCACGTTTTCCATCATCTTTTTGGAAGCAAGCA
TGCGAACTGGCTCCTATTTTCAATGAGCTCGTGGACCGTGTGAGCTTGGATGGGGAGTTCTTACAAAGTTCTTTG
TCTAGAACAAGGCAGGTTGACGATTTCACTGCTAGGCTGTTGGAAATTCATGCAAAGATGATGGCAGTAAACAAG
AAGGAGGACATCCGCTTAGGGTTGCACCGATCTGACTACATGCTGGATTCTGAAACAAATTCACCTTCTTCAAATT
GAGCTCAACACTATCTCAACCTCATTTCTGCTGCTTGGCTCCCTTGTAGCAAATTCACAGGAGCTTACTTAAT
CGATATGGTGAAGTGTTAGGTCTTGATTCTGAAAGGATTCCTCCCAAATTTGGGCAGCCAGTCAATTTGCTGAAGCA
TTGTACAAAGCATGGGCTGAGTATCATCATGACAGTGCTGTAGTTATGATGGTTGTTCAACCTGAAGAAAGAAAT
ATGTACGATCAATACTGGCTCGTCAATCATTTGAAGGAATCGCATGGTGTGACGGCTATTAGGAAAACCTTTGGCA
CAGGTGGAGGCCGAAGGGCAGGTGCTCCCGGATGGAACGCTTGTGGTAGATGGTCGGCGTGTGCTGTTGTGTAC
TTTAGAGCTGGGTATGCGCCAAACGATTACCCCTTCCAGAAGTGGAATGGAGTGGAAGGCTTTTGATTGAAGAGTCA
TCCGCTATTAAGTGCCCTTCGATATCCTACCACTTAGTGGGAACGAAAAGATCCAGCAAGAACTAGCAAAGCCT
AATGTTCTTGAAAGATTTGTTGACAACAAGACACACATTGCCAAGCTACGGAATGTTTCGCAGGGTTATGGAGC
TTGGACAATGATGAAATAGTGAAATCAGCAATAGAAAAACCTGACTTGTTTGTCTTAAACCTCAGCGAGAAGGT
GGAGGGAACAACTTGATGGTCATGATTTGAGAGAAACGCTGATCAGATTGCAGAAGGAACAGGGAGAGGCCTT
GCGGCCTACATTTTGATGCAGCGGATTTTCCAGAGCTTCTCTTACTCAGCTTGTCAGGGCGGTGTTTGCTTT
GAGGACCTTACAATCTCTGAGCTTGGAATATTTCGGAGCTTACCTGCGGAACAAAGACAATATAGTCATAAATGAT
CAGTGTGGTTACTTAATGCGCACAAAAGTTTCTTCATCAAACGAAGGTGGAGTTGCTGCAGGATTTGCTGTTCTG
GACAGCATTATCCTCACAGATGAGTGATGCAGGAAGGCAGCAGAGTTACAACCTACAAGTGACAGCAGGCACTGT
TCTAGGAATGAATCTGACAGCAAAAAGGTTGCAGTGTGCAAATTTGTGTAGGAGCATCTGTTCTTTTGGCATTGG
AAACAATGTGGATAGCTTAGAAAATAACCAATTGCATGCGCAACTTTGCCACACATGAGATGAGGATGCAACACG
GATTACAATTGGTTTAGCCCTTGGAAGGGAGGATAATTGGATAAATTAGGTTGCATGTGTTACTGTATTACAGCT
ACAAATGGTGGTCAGGTTACAGGATAGAGACCATTGTTTCACTGGAATGATGTAGATGTAAAATAAAAACCTGT
TGGCAGGTGAGCCTGCCAACCTCATTTCTCCGCACTTCAAACATGATCATATGATGAATGTGATGTTCTGCCTG
GTTGTTGGTGACAGGAGTTCAGTACTTCTGTTCTTTA

FIG. 44

SEQ ID No. 43

>GSH2 protein Bradi [UniProt ID: I1IJU8]

MPSSCVSSPSSSHRHQCVRLLAPPVPRASLA AVRPSLRRSFVSLRATSTEAAAGPDVAPTGDAGKAAAVRAEMA
EAAAVWCALHGLVVGDRADPRSGTVPGVGLVHAPFSLPARFPSSFQKQACELAPIFNELVDRVSLDGEFLQSSL
SRTRQVDDFTARLLEIHAKMMAVNKKEDIRLGLHRS DYMLDSETNSLLQIELNTISTSF PGLGSLVSKLHRSLLN
RYGEVLGLDSERIPPNWAASQFAEALYKAWAEYHHD SAVVMVQPEERNMYDQYWLNVNHLKESHGVT AIRKTLA
QVEAEGQVLPDGT LVVDGRRVAVVYFRAGYAPNDY PSEVEWSGRLLIEESSAIKCP SISYHLVGTKKIQQELAKP
NVLERFVDNKTHIAKL RKCFAGLWSLDNDEIVKSAIEKPD L FVLKPQREGGNNLYGHDLRET LIRLQKEQGEAL
AAYILMQRIFFRASLTQLVQGGVCFEDLTISELGI FGAYLRNKDNIVINDQCGYLMRTKVSSSNEGGVAAGFAVL
DSIILTDE*

FIG. 45

28/37

SEQ ID No. 44

>NM_001073055.1__Oryza_sativa_Japonica_Group_Os12g0263000_(Os12g0263000)_mRNA,_complete_cds GSH2

GCTCCATGCCCCATTTGCTCTACTGCCAATGTCATTTTCGAAAGTTTACTGGGATCAGGCGGTTGAATTGGCTCC
CCTTTTCAATGAGCTTGTTGATCGTGTTAGTTTGGATGGGGATTTTTTGCAGGAGACTTTGGCAAGAACGAAAGA
GGTGGATTCATTCACCTGGAAGGCTTCTTGATATCCATGCAAAGATGATGAACTAAACAAGAAAGAGGATGTTAG
GTTGGGTCTTACTAGGTCAGATTACATGATTGATGGGGCCACTGACCAGCTTCTTCAAGTTGAACTCAACACCAT
TTCAACATCATCCAACGGTCTTGCTTGCGGTGTATGTGAACCTCACAGAAACCTGATTAGGCAGCATGAGAGGGA
GCTCGGTTTGGATCCAGAGAGTGTTGTTGGGAACACAGCAATTGCTCAGCATGCTGAAGCATTAGCTGGAGCATG
GGCTGAGTTCAACAATCAAAGTTCAGTAGTTCTGGTGGTTGTTCAACCAGAAGAAAGGTACATGTATGACCAATA
TTGGATAACTGTCGCCTTGAGAGAAATGTATGGGGTGACGACAATTCGCAAAACAATGGCTGCAATAGATGCAGA
AGGAGAGCTCCGTCCTGACGGTACACTAACAATAGATGGATTGCCAGTTGCTGTTGTGTACTTCAGAGCAGGTTA
TACACCAAATGATTACCTTCAGAAGCGGAATGGAGAGCAAGACTCTTGATTGAGTGTTCTCTGCTATTAAGTG
CCCATCAATAGCACACCATCTTGTTGGAACCAAAAAGATTCAACAAGAATTGGCAAAAGAAAATGTGCTCGAAAG
GTTTCTTGACAACAAAGCCGACATAGAAAAAGTTCGGAAATGCTTTGCAGGGCTATGGAGCTTGGAGAATGACAG
TATAGTCATGTCTGCTATTGAATCACCAGAATTGTTTGTCTGAAGCCACAAAGAGAGGGTGGAGGAAACAACAT
TTATGGGGACAATCTGCGTGAAACACTGATAAGTCTGAAGAAGGATGGAAGCAATGAACTTGCAGCCTACATCCT
AATGCAGAGGATCTTTCCACCAGCATCACTCTGCTATCTAGTCCGAGATGGCACATGCATTGGGAAAATGCGGT
TTCTGAATTTGGAATTTTTGGAGCCTACCTGCGGAACAAAGACAGGGTCATTATAAATGACCAGTGTGGTTATTT
GATGAGAACCAAGGCTGCCTCACTGAACGAAGGTGGGGTAGTTGCTGGATACGCATTTTTGAACAGCGTATTCCT
GACATGACATGGTAAGACATGCGAAACATGTGTATCGGTTTGAATATGGATAATCTCCAAGCAGGTTATCAGAG
AATTCATAATTTGTATGCTATTACTTTTAGCAGGAATCATACCAATGTTATTTCCATGACACACTTTTTTTTTATT
TGGCATGATTCGCTATCAGCCGAGATTGAGGTACCATTATTCCCATGTATAAAAATGTTATAGCTATCATATTT
TAATGTATGAGTATCAATTTGCTT

FIG. 46

SEQ ID No. 45

>GSH2 protein Oryza [UniProt ID: Q0IP23]

MSFSKVYWDQAVELAPLFNELVDRVSLDGDFLQETLARTKEVDSFTGRLLDIHAKMMKLNKKEDVRLGLTRSDYM
IDGATDQLLQVELNTISTSSNGLACGVCELHRNLIHQHERELGLDPESVVGNTAIAQHAELAGAWAEFNNQSSV
VLVVVQPEERYMYDQYWTVALREMYGVTTIRKTMAAIDAEGELRPDGTLTIDGLPVAVVYFRAGYTPNDYPSEA
EWRARLLIECSSAIKCPSIAHHLVGTKKIQQELAKENVLERFLDNKADIEKVRKCFAGLWSLENDISIVMSAIESP
ELFVLKPQREGGGNNIYGDNLRETLLSLKKDGSNELAAYILMQRIFFPASLCYLVRDGTICIRENAVSEFGIFGAY
LRNKDRVIINDQCGYLMRTKAASLNEGGVVAGYAFLNSVFLT

FIG. 47

29/37

SEQ ID No. 46

>NM_001148007.1__Zea_mays_ LOC100273590_(LOC100273590),_mRNA
GSH2

GTCCCTCCCCCTGCGTCTCGCACCAACAGCAGGTCCGTCTCCGTCTCCGCGCTGCCAGCCGCCTTCCCGCCCCAGA
CACGGCCTCGTTTCGCGCCGCCCCGCGCCCGGCGGCTTCGGGTGGCGCACGCGGCGATGAGCGCCAAGGCGCCCCCT
CGGCGTGGCGCCGGCGGAGGAGGGGGCCGGGACGCCGGGGCAGCAGGCCGTGCTGGGGCAGATGGTGGAGGATGC
CGCCGTGTGGTGCGCCGTCCACGGGCTCGTCTCGGAGATCGCGCCAACCAGAGATCAGGAACGGTTCCTGGGGT
TGGCCTGGTTACGCTCCATTCTCGCTGCTTCCAGCGCGTTTTCCGGCATCGTTCTTCAATCAAGCATGCGAGCT
GGCTCCTATCTTCAATGAGCTCGTGGATCGTGTGAGCTTGGATGGGGAGTTCTTGCAAGCTGCTTTGTCTAGGCA
AGTTCTGAATTCTATCCACTGGATTCTTTTGGTTAACAGCGGATTATCATTTTCCAATTTCAGAACAAAGCAGGT
TGATGAATTTACTTCAAGGCTGCTAGAAATTCATGATAAGATGATGTCTATAAATAAGAAAGAGGATATACGCTT
AGGATTGCACCGGTCTGACTATATGCTGGATTCTGAAACCAATTCACTACTTCAAATTGAGCTCAATACTATCGC
ATCATCCTTTCTTGGCCTTGGCTCCCTTGTCAGTGAACCTCACAGGACCTTACTCAAACAGTATGGAAAAGTATT
AGGCCTAGATCCTAAAAGGATTCCACAGAATTGGGCGAGCCACTCAATTTGCTGAAGCATTGGGCAGAGCATGGGC
CGAGTATAACAATGACAGTGCTGTTGTTTTGATGGTTGTCCAACCTGAAGAAAGAAATATGTACGACCAGTACTG
TCTTGTCAATCATCTGAAGGAATCATATGGTGTGGCAACTATAAGGAAAACATTGGCACAAGTAGAGGCTCAAGG
GCAGGTTCTTACAGATGGAACACTTGTGGTAGATGGTCGGACAGTGGCTGTTGTGTATTTTACAGAGCCGGGTATGC
ACCAACTGATTACCCTTCGGAAGCGGAGTGGAGAGCGAGGCTTCTGATGGAACAATCATCTGCAGTGAAGTGCCC
TTCGATAGCCTACCATTTGGTAGGAACAAAAAGATCCAGCAAGAACTAGCAAAACCTAATGTTCTTGAAAGATT
TCTTGACAACAAGGACGACATTGCCAACTACGCAAGAGTTTTCGAGGGTTGTGGAGCTTGGATAATGAAGAGAT
TGTGAGATCTGCAATACAGAAACCTAACCTGTTTGTCTGAAGCCTCAGCGAGAAGGCGGAGGGAACAACATATA
TGGTCAGGATTTTGCAGGACACACTGATCAAACCTCAGAAGGAACAGGGGGAGGCGCTTGCGGCGTACATACTGAT
GCAGAGGATTTTTCCAAAAGCATCTCTTACTCCTCTAGTCAGGGGCGGTGATTGCTTTGAGGACCTTACAATCTC
GGAGCTGGGAATATACGGAGCCTACCTGCGGAACAGAGACAAGGTGATCTTGAATAACCAGTCTGGCTACTTGAT
GCGCACCAAAGTTTCTTCTTCAAACGAGGGCGGCGTTGCTGCAGGGTTTGCCGTTTTTGGACAGCGTTCTCCTGAC
TGATGAGTGAGACAGCAGCGAGGTGGCAGAGCTATGGGTGACACCTGACCTGCATTAGTGACAGTGCCACAGTC
AACTCAAACGATGACAATATCAACTGCAATCGTAGGGGAACTATGAAATTGCTTACAGAATTGCGGACAGTCGG
AGAAGATGGCAGTGTTACGTTATGATATAGTTTCATGTTTGTATGCTGGTTGAAAATGTAGGGAACTCTTGTTT
TCACTAGAAAGGAACCCCTGGAGGGGGGAGTAACGGTGTAGAAGGTTTATGTTGTACAGTTGGCATGCATTTCAT
CATATAGGAATAGCTAGGAGATGGCACCGTTATGCCAACACAGTTATAACAAATATCAATTGTAAATTTTATGAT
TGGTTATCGACTAAGCAAAACAATTTTATAAGAAATTGTTGTGCGGC

FIG. 48

SEQ ID No. 47

>GSH2 protein Zeama [UniProt ID: B4FWG0]

MSAKAPLGVAPEEGAGTPGQQAVLGQMVEDAAVWCAVHGLVVGDRANQRS GTVPGVGLVHAPFSLLPARFPASF
FNQACELAPIFNELVDRVSLDGEFLQAALSRQVLNSIHWILLVNSGLSFSNFR TKQVDEFTSRLL EIHDKMMSIN
KKEDIRLGLHRSDYMLDSETNSLLQIELNTIASSFPGLGSLVSELHRTLLKQYGVKVLGLDPKRIPQNWAATQFAE
ALGRAWAEYNDSAVVLMVVQPEERNMYDQYCLVNHLKESYGVATIRKTLAQVEAQGVLT DGTLVVDGRTVAVV
YFRAGYAPTDYPSEA EWRARLLMEQSSAVKCP SIAYHLVGTKKIQQELAKPNVLERFLDNKDDIAKLRSFAGLW
SLDNEEIVRS AIQKPNLFVLKPQREGGNNIYGQDLQDTL IKLQKEQGEALAA YILMQRI FPKASLTPLVRGGDC
FEDLTI SELGIYGAYLRNRDKVILNNQSGYLMRTKVSSSNEGGVAAGFAVLDSVLLTDE*

FIG. 49

30/37

SEQ ID No. 48

>XM_001757544.1__Physcomitrella_patens_subsp._patens_
protein(PHYPADRAFT_119267)_mRNA,_complete_cds GSH2

ATGGAGTTGACGAAGGACCTCACAGAGGATATCGTCCGGGAGGCACTCGTTTGGGCGAGCTTGCATGGAGGAGTT
GTGGGCGACAAAACAAATGGGAAGTCAGGCACTTCGCCAGGCGTGGGGATGGTCCATGTGCCGCTAGCTCTGCTA
CCGACACCCCTTCCAAGACAAGCTTTTGAGCAAGCTCTTGAGCTTGCCACTCACTACAACACCCTCGTGGATCGT
GTTAGTCAAGACAGTCATTTCCCTTCGAGAATGCCTTGCGCATACACGTAAGGCAGACAGTTTTACAGCGAAGCTA
ATGAACATCTACGACCAGATACTTGCAGAAGGCATCGCACAGGACATTCAGCTGGGGCTTCACCGTAGTGATTAC
ATGTTGGACATGGCGACTGGAGCTCTGTTGCAAGTGGAGATGAACACAATCTCATCCTCCTTTGCAGGACTCGGC
GCTTTGACAAGTCAACTGCACAAGTATCTTATCGGCCGGTTTGGGCAAGACTTGGGTCTCAACGTGGAGAATGTA
CCAGAAAACAGTGCTGCAAATGATTTTGCGGAAGGTCTGGGTCAAGCTTGGTTGGAATATGGAAATCCTAATGCA
GCTGTGCTGATGATAGTGACGCAAAGGAGACGAATATGTATGACCAGCACTGGCTCAGTTTCAAGCTGTATGAG
AAGTATGGCGCCAAGGTGTTCCGAAGAACTTTTGCAGAAGTTGCTGCGCAGGCTAAACTTGACGAGAAGCGGACG
TTGCTGATCGGAGAGCAGCCGATAGCATTAGTATACCTTCAGGTCAAGGTACTCTCCAGACGACTACGTCTCCGAC
CGCGATTGGGATGTCAGGCTGCTGATGGAAAGGTCAACCGCCGTGAAGTCGCCAAGCATCTCCTATCACCTCGTC
GGTGCAAAGAAAATCCAACAAGAGCTCACGAAGCCTGGAGTTTTTGGAGCGGTTTTTACACGACAAGGACGCCATT
GCCAAAGTGAGAGAGTCCCTTCGCAGGTCTCTGGAGCCTGGAGAACGATGGCGGCGATCACATCCTTGAAGAGGCT
GTTGCGAAACCCGAAGGGTTCGTGTTGAAGCCGCAGCGAGAGGGAGGAGGAAACAACATTTACGGCAAAGATGTT
GCAACAACGATTAAGAAACTGAATCAAGAGGAATTGGCAGCTTACGTCTCTGATGCAACGCATCTTCCCTACAGTC
CACGACTCTTACCTCGTCCGAGACGGCGTCTGCAGAACATCCAAGATCATCTCTGAATTGGGCATCTTTGGCACT
TACTTGAGAAATGGGGAGAAGGAAATTCTGAACAAACAAGCGGGCCACTTGGTGCGAAGCAAAGTGTCTGACTCT
GATGAAGGTGGTGTAGCTGCCGGCTATGGCGTCTCGACAGCCCATATCTCACCTAG

FIG. 50

SEQ ID No. 49

>GSH2 protein Phypa [UniProt ID: A9RTT8]

MELTKDLTEDIIVREALVWASLHGGVVGDKTNGKSGTSPGVGMVHVPLALLPTPFPRQAFEQALELATHYNTLVDR
VSQDSHFLRECLAHTRKADSFTAKLMNIYDQILAEGIAQDIQLGLHRSYMLDMATGALLQVEMNTISSSFAGLG
ALTSQLHKYLIGRFGQDLGLNVENVPENSAANDFAEGLGQAWLEYGNPNAAVLMIVHAKETNMYDQHWLSFKLYE
KYGAKVFRRTFAEVAAQAKLDEKRTLLIGEQPIALVYFRSGYSPDDYVSDRDWDVRLLMERSTAVKSPSISYHLV
GAKKIQQELTKPGVLERFLHDKDAIAKVRESFAGLWSLENDGGDHIILEEAVAKPEGFVLKPQREGGGNNIYGKDV
ATTIKKLNQEELAAAYVLMQRIFFTVDHSYLVRDGVCRSTSKIISELGIFGTYLNRNGEKEILNKQAGHLVRSKVS
DSEGGVAAGYGVLDSPYLT*

FIG. 51

31/37

SEQ ID No. 50

>XM_001691491.1__Chlamydomonas_reinhardtii_strain_CC-503_cw92_mt+ (GSH2)

ATGGACTTGCTGGGCCAATTCCATACAACGGCGCACCCCAATTACCTGCTGAGCCAGTGCAGCAGTACTGTGGTT
GCGCTGAAGGACCCAGGTTTCGTCAACCCGGACTATGCCGTCATCCACGCCCCCTGTCGCTCACACCGGTGCCG
TATCCACGGGCGCGCTTCGAGCTAGCGCAACAGGTGATGCCGCTCTTTAACACGCTGGTGGACGCCGTGGCAGCG
GATGAGTCGTACCTGGAGGCCACGCTGCGCCCCGCCGCGCGCAGGACGCCCTTCACGGGGCGGGCTGCTGGGCTGC
CTGCTGGCCACGCGGCCGGAGCGGCAGGCGCTACGGGGCTCGGGCCGGGAGGTGGTGTGGGCGTGCACCGCAGT
GACTACATGCTGGATGCACCCAGCGGCAAGTTCCTGCAGGTGGAGCTCAACACGATCGCCTCCTCATTGGGTGC
TTGTGAGCGCTGATGACGCGCCTGCACGGGTACATGGCCAGCGGGCGAATCAGCGGCCTGCTCCCTCATGCCGCA
CTCAAGAAGGGTGCTTCTCACGTCCGCAATGCCATCTCCTCCCGCATTGCTGTTCTTTGCAAACCTCTTAAACCT
TACCCGAACCTTGTGTTCTCTGGCTCATGCCACAACAGGATGCGTCGCGGCTTCGCGCTAACCCAACTCTGGAG
CACATCCCGGATGCCATCGCTGCGGCGGCAACAGCCAGCGGCCACGTGCGAGCGGGCGGCGTGGTGGTGTGGTG
TGGATCCAGCTGCAGCTGTGGGAGCGGCACCGCTGCGCAGCTGCGCCTCACGCTGGCGGACATCGCGGCGGGC
GGCCGACTGGACGCCGCCAGCGGCACCTCACGCTGCCGGTGGGCGAGCCCGCGGCGCTGTTCTACTTCAGGGCC
GGCTACACGCCCCGCCGACTACCCGAGTGAGGCGGAGTGGGCGGCGCGGGAGCTCATTGAGCGCTCCAACGCCGCC
AAGTGCCCCACCGTGCGGTATCAGCTGGCGGGCACCAAGAAGGTGCAACAGGACCTGGCGACACCCGGCGTGCTG
GAGCGCTTCTGCGCCGGCTACCGCCGCGGCGGACGCGGCGCTGCTCCGTGAGTTCTTCGCGGGGCTGTGGTCG
CTGGACCCACGGGACCTGGCGGCGGACGCGAGGCGGCGGCGGCGGCGGCGGTGCGCGACGCCATTGCTCGGCCGGAG
GCGTACGTGCTCAAGCCCCAGCGCGAGGGTGGCGGCAACAACCTGTACGGCGAGGAGTTGCGGGCGGCGCTGCAG
CAGGGCGGCGAGGAGCTGAGCGCCTACATCCTGATGCAGCGCATCCTGCCGCCCGTCAACAGGTCTGTGCTGGTG
CGCAACGGGCAGTGGCAGGAGGCTGAGACGCTGAGTGAAGCTGGGCATCTACGGCACCTTCGTGCGCCACAACAAC
GCCGTGGTGGTGAACCAGCAGAGCGGCCACTTGGTCCGGACCAAGACCGCCAGCTCCAACGAGGGCGGTGTGGCG
GCGGGCTTTGCGGTGCTGGATAGCCCTTACCTGGTGGAGTGAGGCGGCGGCTCACCTGGAGGAGCGCGAGGCGTG
GGGCTGTGCGGAAAGGGTGAGGGCTGGAGGCTGTAGAGGCTGTGGTGGCGTGTGTGAGTTGTGATGCGTGTGT
GTGTGTGTGTAGGTGGGTGTGGATGGGGGTAGGGGTGTTTTTGCCCGAACGCAACGCAATAGGGGGTGGGTGTCC
ATGCAAGGCGGCTTGTGGGGTCAGAGCCCAGAAATGGGGAAGTCGGCACACGGGGTCATATATTGCAGTACATAC
GTGGAGCAGGGAGCTAGAATGTAGGGGATAGGGCTGCATGGGCTGGAAGTGCAGGAGCTTGTGACAGTAGGTGAGT
AGGTGCGTATGAGAAACCTGCACATCCCGTGTGACCCAGT

FIG. 52

SEQ ID No. 51

>GSHB protein Chlre [UniProt ID: A8IR74]

MDLLGQFHTTAHPNYLLSQCSSTVVALKDPRFVNPDIYAVIHAPLSLTPVPYPRARFELAQQVMPLFNTLVDAVAA
DESYLEATLRPAAAQDAFTGRLLGCLLATRPERQALRSGREVLGVHRSDYMLDAPSGKFLQVELNTIASSFGC
LSALMTRLHGVMASGRISGLLPHAALKKGASHVRNAISSRIAVLCKLLKYPNLFVSGSCPQQDASRLPPNPTLE
HIPDAIAAAATASGHVAAGGVVVMWIQLQLWERHRLRTLRLTLADIAAGGRDLAASGTLTLPGGQPAAVFYFRA
GYTPADYPSEAEWAARELIERSNAAKCPTVAYQLAGTKKVQQDLATPGVLERFLPPATAAADAALLREFFAGLWS
LDFRDLAADAGGAAAARDAIARPEAYVLKPQREGGNNLYGEELRAALQQGGEELSAYILMQRILPPVNRSVLV
RNGQWQEAETLSELGIYGTFFVRHNNVAVVNQQSGHLVTRTKTASSNEGGVAAGFAVLDSPLYLVE*

FIG. 53

32/37

SEQ ID No. 52

>XM_001418379.1__Ostreococcus_lucimarinus_CCE9901_glutathione_synthase_(GSH2)_mRNA,_complete_cds

ATGATCGGAGGAGACGAAGCGGGGACGTTACGCACGCGCCGGTGTCTGTCGCGCCGACGACGTTTCCGAGGAGC
GCGTACGAGCTCGCCGAGGCGACGTCGCCGGCGTTTCGCGGAGCTGTACGACGGAGTGTCTGAGGGACGATGGATTCT
CTGAGGGAAACGCTGCGGGGGGCGATTAAAGAACGATGCATTACGCGCCGCGCTGTGGGACGTGTACGAACGGAGC
GGCGGGCAAGAGGGGAGACAGAAGGCGGAGGTTGGCGTGTTCGAAGCGATTACATGCTCGACGCGCCGAGCGGG
TTGCCTTTGCAAGTCGAGGTGAACACGGTGAGCACGTCGTTTCATGGCGCTCGGATCGAAGGTGAGCGCGATGCAT
AGACACTTAGCGGCGTGGTCGGCGAAACCTGGAGAAGTGGTGGATTTTCGACGCGATGCCACAGAACGACGCGCTC
TCGAGCGCGGGCGACGTCCTCGCGGCGGGATGGAAAGCGATGGGGAGTAAAGGGAAAATCATGATGGTGGTTTCAG
CCGAACGAGCGTAATATGTTTCGATCAGCGGCACATCTCGCATTATATATTCGAAACGCACGGAATTCACACCGTC
AGAGCGACTTTAGCCGAAGTAAACGCCGAGGGCGCGTTGCGCGACGGAACGCTCACGTTTCGAGGCGACGACATC
GGCTTGGTGTACTTTTCGAGCCGGATACACCCCGTCGGACTATCCGACCAGTGGCGAATGGTCAGCACGGTTGTTA
ATCGAACAATCGAACGCTTTGAAGAGCCCTTCGGTGGCAATGCACTTGGCGGGATCGAAGAAGGTGCAACAAAAG
CTCGCAGAAGCGGGCGTGCTCGAATCGTTTCATGCCGGAACGCGCCGCGGACATGCGACGTGTCTTTGCAGATTTG
TATTCCTTAGACGGCGACGAAGGCACCGAGGCGATTAAACTCGCACTCGCGAATCCAGCGGGCTACGTCTTAAAG
CCTCAGCGTGAGGGTGGTGGAAATAATTTGTATTTCGGACGATTTGCGCAAACGACTTGAACTCGGCGGAGACTTG
GGCGCGTTCATTCTCATGCAACGCATTCTCCCTCCATCTCATCAAACGCTCACCGTTTCGCGCCGGTAAGGTGCTC
GAAATCGAAACACTGTCCGAGTTAGGTATTTACGGTAGTTATCTCCGCGTAGGCGACGAGATCGTGATCAATAAA
CACGGTGGGAGTTTGTCTCCGGACGAAGGCGGCGACTTCCGACGAAGGCGGCGTTCGCGGCGGGGTACGCCGTCTTA
GACTCTCCTCTTCTCGTTTAG

FIG. 54

SEQ ID No. 53

>GSH2 protein Ostlu [UniProt ID: A4RYM3]

MIGGDEAGTFTHAPVSCAPTTFFPRSAAYELAEATSPAFELYDGVSRDDGFLRETLRGAIKNDAFTAALWDVYERS
GGQEGRQKAIEVGVLRSDYMLDAPSGLPQLQEVNTVSTSFMALGSKVSAMHRHLAAWSAKPGEVVDFDAMPQNDAL
SSAGDVLAAAGWKAMGSKGIMMVVQPNERNMFDQRHISHYIFETHGIHTVRATLAEVNAEGALRDGTLTFRGDDI
GLVYFRAGYTPTSDYPTSGEWSARLLIEQSNALKSPSVAMHLAGSKKVQQLAEAGVLESFMPERAADMRRVFADL
YSLDGDGTEAIKLALANPAGYVLKPQREGGNNLYSDDLKRKRLELGGDLGAFILMQRILPPSHQTLTVRAGKVL
EIETLSELGIYGSYLRVGDEIVINKHGGSLLRTKAATSDEGGVAAGYAVLDSPLL*

FIG. 55

66
 (59) **SVKRSKR**REHQI **IV**AAASPPTE **FA**VVAATEPLT **RED**LIAYLAS **GCK**PKDKYRIGTEHEKFGFEVNTLR
 (61) **SVKRSKR**RGHQI **IV**AAASPPTE **FA**VVAATEPLT **RED**LIAYLAS **GCK**PKDKYRIGTEHEKFGFEVNTLR
 (61) **SVKRSKR**RGHQI **IV**AAASPPTE **FA**VVAATEPLT **RED**LIAYLAS **GCK**PKDKYRIGTEHEKFGFEVNTLR
 (46) RIVGG--RRV **IV**AAASPPTE **DA**VVA**D**PLTK**Q**DLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (47) RILTSGGSRRE **IV**AAASPPTE **DA**VVAATEPLT **TK**QDLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (65) **AGIGIR**RGRRS **IV**AAASPPTE **DA**VVAATEPLT **TK**QDLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (62) **IGVSR**RGDLT **IV**AAASPPTE **DA**VVAATEPLT **TK**QDLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (62) **IGVSR**RGGLT **IV**AAASPPTE **DA**VVAATEPLT **TK**QDLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (37) **AGR**KGSK-RGGAAASPPTE **FA**QOMTEPLT **TK**QDLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (35) **AGR**-G--RGGAAASPPTE **FA**QOMTEPLT **TK**QDLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (1) -----MTEPLT **TK**QDLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (6) **IG**KGE-----**VIG**SPPSN **DA**GSY**G**GEPLT **RO**DLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 (19) **RSR**GVSPRLSPVHNA**FA**VAERR-TEPL**K**QDLVAYLAS **GCR**PPRSAMRIGTEHEKFGFEIGSLR
 (16) **ATR**RCRR-ATTRATT**T**IRAAATAPARAV**RE**ECVAVPR **GCK**PR**EA**FRIGTEHEKFGFEIGSLR
 (66) A K RR **IV**AAASPPTE **DA**VVAATEPLT **TK**QDLVAYLAS **GCK**PKDKWIRIGTEHEKFGFEIGSLR
 Consensus

131
 (124) **PMKY**EQIAELINGIAERF**FE**WEKVM**EG**DKI**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (126) **PMKY**DOIAELINGIAERF**FE**WEKVM**EG**DKI**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (126) **PMKY**DOIAELINGIAERF**FE**WEKVM**EG**DKI**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (108) **PMKY**DOIAELINGIAERF**LD**WKVM**EG**DKI**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (112) **PMKY**EQISELINGIAERF**LD**WKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (130) **PMKY**EQIAELINGIAERF**LD**WKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (127) **PMKY**DOIAELINGIAERF**LD**WKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (127) **PMKY**DOIAELINGIAERF**LD**WKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (101) **PMKY**DOISGILNGIAERF**LD**WKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (96) **PMKY**DOIRDLINGIAERF**LD**WKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (42) **PMKY**DOIRDLINGIAERF**LD**WKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (65) **PMKY**EQIKQILEGLADRF**FE**WEKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (83) **PMKY**DOIAQVLRKLEAR**FE**WEKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (80) **PMKY**PMVRHVLS**SS**VERHGWTP**INE**ADN**IG**CTKD**Q**SVT **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 (131) **PMKY**DOIAELINGIAERF**LD**WKVM**EG**DN**IG**LKQKQKQ**S**IS **LE**PGGQ**FE**LSGAP**LE**TLHQTCAEVN
 Consensus

FIG. 56 (Continued)

196		(189)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWR	REDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Araly*		SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWR	REDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Arath(1)*	(199)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWR	REDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Arath(2)*	(191)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWR	REDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Glyma*	(173)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWG	IKDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Medtr*	(177)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWR	REDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Cicle*	(195)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWG	IKDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Solly*	(192)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWG	IKDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Soltu*	(192)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWG	IKDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Bradi*	(166)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWAL	SDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Orysa*	(161)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWAL	SDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Zeama*	(107)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWAL	SDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Phypa*	(130)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWAL	SDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Chlre*	(148)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWAL	SDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
	Ostlu*	(145)	WHJETVNA	LA	REK	CHRE	FL	GI	GFO	PKWAL	SDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV
	Consensus	(196)	SHLYQVKA	VAEE	MGIG	FL	GI	GFO	PKWAL	SDI	PI	MPKGRY	D	IMRN	YMP	KV	GL	GL	DMML	RTCT	VQV	
325		(254)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Araly*		NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Arath(1)*	(256)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Arath(2)*	(256)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Glyma*	(238)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Medtr*	(242)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Cicle*	(260)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Solly*	(257)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Soltu*	(257)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Bradi*	(231)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Orysa*	(226)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Zeama*	(172)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Phypa*	(195)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Chlre*	(213)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Ostlu*	(210)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	
	Consensus	(261)	NLDF	SE	EA	DMIR	KFRAG	LALOP	I	ATAL	FANS	PF	TEGKPN	GFL	S	MRSHI	WID	TD	KDRIG	ML	PFVFD	

FIG. 56 (Continued)

Araly*	(319)	DSFG	FEQ	VDYAL	DP	PMYFA	YRKNKY	IDCT	GMT	FRQ	FL	AGKT	PC	PGEL	PSYN	DWEN	HLTT	IFPE
Arath(1)*	(321)	DSFG	FEQ	VDYAL	DP	PMYFA	YRKNKY	IDCT	GMT	FRQ	FL	AGKT	PC	PGEL	PSYN	DWEN	HLTT	IFPE
Arath(2)*	(321)	DSFG	FEQ	VDYAL	DP	PMYFA	YRKNKY	IDCT	GMT	FRQ	FL	AGKT	PC	PGEL	PSYN	DWEN	HLTT	IFPE
Glyma*	(303)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKNRY	IDCT	GKT	FRD	FL	AGRT	PC	PGEL	PTLN	DWEN	HLTT	IFPE
Medtr*	(307)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKKKY	IDCT	GMT	FRD	FL	AGKT	PC	PGEL	PTLN	DWEN	HLTT	IFPE
Cicle*	(325)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKKKY	IDCA	GMS	FRD	FL	AGKT	PC	PGEL	PTLN	DWEN	HLTT	IFPE
Solly*	(322)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKKKY	IDCT	GLS	FRD	FM	NGKT	PP	PGFY	PTLN	DWEN	HLTT	IFPE
Soltu*	(322)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKKKY	VDCS	GLS	FRD	FM	NGKT	PP	PGFY	PTLN	DWEN	HLTT	IFPE
Bradi*	(296)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKNQY	IDCT	GMS	FRD	FM	NGKT	POV	PGEL	PTLN	DWEN	HLTT	IFPE
Orysa*	(291)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKNKY	IDCT	GMS	FRD	FM	NGKT	POA	PGEL	PTLN	DWEN	HLTT	IFPE
Zeama*	(237)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKNKY	IDCT	GMS	FRD	FM	NGKT	POA	PGEL	PTLN	DWEN	HLTT	IFPE
Phypa*	(260)	DGFG	FDKY	VEYAL	GV	PMYFV	YRNGRY	IDCS	GMS	FRD	FL	EGKT	SAL	PGEK	PTMN	DWEN	HLTT	IFPE
Chlre*	(278)	KDMC	FFSY	VDYAM	AV	PMYFV	YRNGQY	INAL	GMS	WKD	FM	AGKT	PAI	PGFY	PTIA	DWEN	HLTT	IFPE
Ostlu*	(275)	DpEG	FEQ	VDYAM	AV	PMYFV	YRNGTY	VDMS	GESW	LD	FM	EGKT	POI	PGDK	PTPD	DWEN	HLTT	IFPE
Consensus	(326)	DSFG	FEQ	VDYAL	DP	PMYFV	YRKKKY	IDCT	GMS	FRD	FL	AGKL	P	LGEL	PT	DWEN	HLTT	IFPE

	391	(384)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	D	S	L	Q	A	I	L	D	T	A	D	W	T	P	A	E	R	M	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
Araly*		(386)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	D	S	L	Q	A	I	L	D	T	A	D	W	T	P	A	E	R	M	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
Arath(1)*		(386)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	D	S	L	Q	A	I	L	D	T	A	D	W	T	P	A	E	R	M	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
Arath(2)*		(368)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	L	S	L	K	S	V	L	D	M	T	A	D	W	T	P	E	R	O	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Glyma*		(372)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	V	S	L	C	R	V	L	D	M	T	A	D	W	T	L	E	R	E	M	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
Medtr*		(390)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	D	S	L	Q	N	V	L	D	M	T	A	D	W	T	T	G	E	R	O	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
Cicle*		(387)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	G	S	L	Q	S	V	L	D	M	T	F	D	W	T	A	E	R	D	M	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
Solly*		(387)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	G	S	L	Q	S	V	L	D	M	T	S	D	W	T	A	E	R	D	M	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
Soltu*		(361)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	S	L	Q	S	I	L	D	M	T	A	D	W	T	N	A	E	R	M	L	R	R	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Bradi*		(356)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	S	L	Q	S	I	S	D	M	T	S	D	W	T	N	E	R	E	M	L	R	R	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Orysa*		(302)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	S	L	Q	S	I	L	D	M	T	F	D	W	T	K	E	R	E	M	L	R	R	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Zeama*		(325)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	S	L	Q	A	L	I	N	D	W	T	E	R	A	N	L	R	S	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
Phypa*		(343)	VR	LK	K	F	L	EM	R	G	A	D	G	G	P	W	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	P	A	O	R	O	A	L	I	E	D	W	T	P	A	E	R	D	Y	L	R	T	E	V	T	R	F	G	L	R																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Chltre*		(340)	VR	LK	K	F	L	EM	R	G	A	D	G	G	S	E	A	M	L	P	A	L	M	W	G	L	I	Y	S	D	K	A	L	D	A	E	A	M	S	D	W	T	Q	V	E	R	D	A	L	R	V	D	V	T	K	D	G	L	S																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
Ostlu*		(391)	VR	LK	RY	L	EM	R	G	A	D	G	G	P	W	R	R	L	C	A	L	P	A	F	W	G	L	I	Y	D	E	S	L	Q	S	I	L	D	M	T	A	D	W	T	F	R	E	M	L	R	N	K	V	P	V	T	G	L	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
Consensus																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																

FIG. 56 (continued)

456 520

Araly* (449) TP FRDGLI KHVAEDVLKAKDGLERRGYKEAGFLNANDEVVRTGVTPAEKLLLEMYNCTMGQSVDP

Arath(1)* (451) TP FRDGLI KHVAEDVLKAKVCIYEPS

Arath(2)* (451) TP FRDGLI KHVAEDVLKAKDGLERRGYKEAGFLNANDEVVRTGVTPAEKLLLEMYNCTMGQSVDP

Glyma* (433) TP FRDGLI KHVAEDVLKAKDGLERRGFKESGFLNENAEVVRTGVTPAERLLELYHGKMEQSVTH

Medtr* (437) TP FRDGLI KHVAEDVLKAKDGLERRGFKESGFLNANAEVVRTGVTPAERLLELYHGKMEQSVTH

Cicle* (455) TP FRDGLI RHVAQDVLKLSKVVTWSCLFMCPLSC

Solly* (452) TP FRDGLI MHVAQDVLKAKDGLERRGFKETGFLNENAEVVRTGVTPAEKLLLELYHGKMGQSVDP

Soltu* (452) TP FRDGLI MHVAQDVLKAKDGLERRGFKETGFLNENAEVVRTGVTPAEKLLLELYHGKMGQSVDP

Bradi* (426) TP FRDGLI VVRDLAEDVLKAKSGLERRGYKENGFLREVDDEVVRTGVTPAEKLLNLYETKWRQSVDP

Orysa* (421) TP FRDGLI VVRDLAETIQLSKNGLERRGYKENGFLREVDDEVVRTGVTPAEKLLNLYETKWRQSVDP

Zeama* (367) TP FRDGLI VVRDLAEEVLKAKNGLERRGYKENGFLREVDDEVVRTGVTPAERLLELYETKWRQSVTH

Phypa* (390) VEFRDGLI KHVAQDVLKAKDGLERRGFKETGFLNENAEVVRTGVTPAEKLLLELYHGKMGQSVDP

Chlre* (408) TP FRDGLI VQDVAKQVMSIAHGLERRGYDETSTFKRLLEVAETGLTQADHLELYETKWRQSVDP

Ostlu* (405) AKFRGGTANDIAKKMVEISVQGLQERGLGEEVLYRYLQDIDDKKSVAAARMSEMNARENMGDLFK

Consensus (456) TPF RDGLLKHVAEDVLKAKDGLERRGYKE GFL EVDEVVRTGVTPAEKLLLELY GKW SVDP

521 535

Araly* (514) VFEELLY

Arath(1)* (478) -

Arath(2)* (516) VFEELLY

Glyma* (498) VFEELLY

Medtr* (502) VFEELLY

Cicle* (488) -

Solly* (517) VFEELLY

Soltu* (517) VFEELLY

Bradi* (491) VFEELLY

Orysa* (486) VFEELLY

Zeama* (432) VFEELLY

Phypa* (455) VFEELLY

Chlre* (473) LYKEFM

Ostlu* (470) VFEYATEPDTLPKIV

Consensus (521) VFEELLY

FIG. 56 (Continued)

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/050323A. CLASSIFICATION OF SUBJECT MATTER
INV. A01H5/12 C12N15/82
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)
A01H C12N

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

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

EPO-Internal, BIOSIS, CHEM ABS Data, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ELMAYAN TALINE ET AL: "Synthesis of a bifunctional metallothionein/ beta-glucuronidase fusion protein in transgenic tobacco plants as a means of reducing leaf cadmium levels", PLANT JOURNAL, vol. 6, no. 3, 1994, pages 433-440, XP002756940, ISSN: 0960-7412 abstract ----- -/--	1-47



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

25 April 2016

Date of mailing of the international search report

09/05/2016

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

Marchesini, Patrizia

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2016/050323

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CREISSEN GARY ET AL: "Elevated glutathione biosynthetic capacity in the chloroplasts of transgenic tobacco plants paradoxically causes increased oxidative stress", THE PLANT CELL, AMERICAN SOCIETY OF PLANT BIOLOGISTS, US, vol. 11, no. 7, 1 July 1999 (1999-07-01), pages 1277-1291, XP002199575, ISSN: 1040-4651, DOI: 10.1105/TPC.11.7.1277	18-21, 24-26, 29,35, 36, 40-45,47
Y	material and methods, figure 4 -----	1-47
Y	MARIJKE JOZEFCZAK ET AL: "Glutathione Is a Key Player in Metal-Induced Oxidative Stress Defenses", INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES, vol. 13, no. 12, 7 March 2012 (2012-03-07) , pages 3145-3175, XP055266704, DOI: 10.3390/ijms13033145 point 3. "Glutathione and Metal Homeostasis", pages 3149-3152 -----	1-47
Y	VICTOR KORENKOV ET AL: "Root-selective expression of AtCAX4 and AtCAX2 results in reduced lamina cadmium in field-grown Nicotiana tabacum L.", PLANT BIOTECHNOLOGY JOURNAL, vol. 7, no. 3, 1 April 2009 (2009-04-01), pages 219-226, XP055266757, GB ISSN: 1467-7644, DOI: 10.1111/j.1467-7652.2008.00390.x the whole document material and methods -----	1-47
Y	GRANT C A ET AL: "Cadmium accumulation in crops", CANADIAN JOURNAL OF PLANT SCIENCE, vol. 78, no. 1, January 1998 (1998-01), pages 1-17, XP002756941, ISSN: 0008-4220 page 2 - page 5 -----	1-47
Y	WO 2010/075243 A2 (DU PONT [US]; PIONEER HI BRED INT [US]; ALLEN STEPHEN M [US]; HABBEN J) 1 July 2010 (2010-07-01) claims 1-22; sequences 25, 28, 43 page 30, line 10 - line 15 ----- -/--	1-47

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/050323

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SAM CHERIAN ET AL: "Transgenic Plants in Phytoremediation: Recent Advances and New Possibilities", ENVIRONMENTAL SCIENCE & TECHNOLOGY, vol. 39, no. 24, 1 December 2005 (2005-12-01), pages 9377-9390, XP055267593, US ISSN: 0013-936X, DOI: 10.1021/es0511341 table 2 page 9382, right-hand column - page 9383, left-hand column, paragraph 2 -----	1-47

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2016/050323

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2010075243 A2	01-07-2010	US 2012004114 A1 WO 2010075243 A2	05-01-2012 01-07-2010
