



- (51) **International Patent Classification:**
C07K 14/415 (2006.01) *C12N 15/82* (2006.01)
- (21) **International Application Number:**
PCT/NL2012/050403
- (22) **International Filing Date:**
8 June 2012 (08.06.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/495,399 10 June 2011 (10.06.2011) US
- (71) **Applicant (for all designated States except US):** **STICHT-
ING TECHNOLOGISCH TOPINSTITUUT GROENE
GENETICA (FOUNDATION TECHNOLOGICAL
TOP INSTITUTE GREEN GENETICS)** [NL/NL]; 68,
Vossenburchkade, NL-2805 PC Gouda (NL).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** **SCHUURINK,
Robert Cornelis** [NL/NL]; c/o P.O. Box 216, NL-6700
AE Wageningen (NL). **HARING, Michael Albertus**
[NL/NL]; c/o P.O. Box 216, NL-6700 AE Wageningen
(NL). **SPYROPOULOU, Eleni** [GR/NL]; c/o P.O. Box
216, NL-6700 AE Wageningen (NL).
- (74) **Agent:** **KINGMA, R., L.**; P.O. Box 3241, NL-2280 GE
Rijswijk (NL).

- (81) **Designated States (unless otherwise indicated, for every
kind of national protection available):** AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States (unless otherwise indicated, for every
kind of regional protection available):** ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

Published:

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

(54) **Title:** TRANSCRIPTION FACTOR MODULATING TERPENE BIOSYNTHESIS

(57) **Abstract:** The present invention relates to identification and isolation of zinc finger transcription factor in tomato that specifically expresses in glandular trichomes of *Solanum lycopersicum* cultivar Moneymaker and binds to the promoters of the genes encoding Terpene Synthase 5 (also known as Monoterpene Synthase1) and Terpene Synthase 11 (also known as Sesquiterpene Synthase 1). The invention provides the isolated, recombinant or synthetic polynucleotides encoding the polypeptide sequences of SEQ ID NO:2 and variants and fragments thereof. The invention also provides constructs, vectors, host cells and plants genetically modified to contain the polynucleotides of the invention. The methods for producing plants with altered levels of terpenes, including trans-formed and mutant plants, are also provided.



Title: **Transcription factor modulating terpene biosynthesis**

Field of the invention

The present invention relates to a zinc finger transcription factor that regulates terpene biosynthesis in plants, a nucleic acid molecule that encodes the transcription factor and methods of use of the transcription factor for producing plants with altered terpene content.

Background of the invention

Terpenes constitute a large group of structurally diverse molecules synthesized naturally by organisms as diverse as bacteria, fungi, plants and animals. Much research has been concentrated on the biochemistry and biological functions of terpenes or their derivatives for potential commercial exploitation (Gershenzon and Dudareva 2007 Nat Chem Biol 3:408).

The result of these studies is a variety of terpene-based products ranging from pharmaceuticals, such as anti-cancer drug paclitaxel and anti-malaria drug artemisin, to fragrances and aroma ingredients, such as menthol and patchuol.

In the field of plant breeding, there is an interest in terpene secondary metabolites produced by many plant species to resist pathogens, to repel or kill pests or to attract beneficial organisms, e.g., predators or parasitoids of pest insects or plant pollinators, or other organisms. Wild plant species frequently produce beneficial secondary metabolites lacking in their cultivated relatives, and therefore are an important source of traits for introgression into cultivated varieties. It is known that secondary metabolites, such as volatile terpenoid compounds, can directly influence insect behavior (Bruce et al., 2005 Trends Plant Sci 10:269-274). For instance, methyl ketones and sesquiterpene carboxylic acids identified in *Solanum habrochaites* and acyl-glucose esters from *Solanum pennellii* were found to be toxic to different insect classes, such as *Lepidoptera*, mites, and aphids (Williams et al., 1980 Science 20:888; Goffreda et al., 1990 Plant Cell 2:643; Juvik et al., 1994 J Econ Entomol 87:482; Frelichowski and Juvik, 2001 J Econ Entomol 94:1249). Often mono- and sesquiterpene hydrocarbons, sesquiterpene acids, methylketones and sugar esters are accumulated in plants in specialized organs such as glandular trichomes on stems and leaves. Several studies correlated the density of glandular trichomes with levels of resistance to pest insects, e.g., maize earworm *Heliothis zea* and Colorado potato beetle (Kauffman and Kennedy, 1989 J Chem Ecol 15:1919-1930; Antonius et al., 2001 J Environ Sci Health B 36:835-848; Antonius et al., 2005 J Environ Sci Health B 40:619-631). The methylketones 2-undecanone and 2-tridecanone accumulated in glandular trichomes of *S. habrochaites* were shown to be toxic to larvae of Colorado potato beetle and adult whiteflies

B. tabaci, respectively (Antonius et al., 2005 J Environ Sci Health B 40:619-631). The myrtle oil, including the monoterpene linalool among its essential components, was shown to have an insecticidal effect on bean weevils, *Acanthoscelides obtectus* Say (*Coleoptera: Bruchidae*) (Ayvaz et al., 2010 J Insect Sci 10: 1536-2442). The sesquiterpenes zingiberene and curcumene, and the monoterpenes p-cymene, α -terpinene, and α -phellandrene from wild tomato *S. habrochaites* and *S. pennellii*, respectively, were shown to have insecticidal properties (Bleeker et al., 2009 Plant Physiol. 151:925). Bio-assays have demonstrated that the sesquiterpenes 7-epizingiberene and its derivative R-curcumene repelled adult whiteflies from landing on tomato plants (Bleeker et al., 2011 Phytochemistry 72:68), and that plants with endogenous production of zingiberene showed resistance to *Tuta absoluta*. (De Azavedo et al., 2003 Euphytica 134:247-251).

Genetic inheritance of the genes associated with development of different types of glandular trichomes and production of zingiberene was studied in interspecific crosses between *S. lycopersicum*, a cultivated tomato which does not produce zingiberene, and *S. habrochaites*, a wild species with high zingiberene production. In F2 plants from these crosses, zingiberene content correlated with resistance to *B. tabaci*. This study suggested feasibility of breeding plants with high levels of zingiberene, 2- tridecanone, and/or acylsugars, which would lead to high levels of resistance to whiteflies (Freitas et al., 2002 Euphytica 127: 275-287). However, programs on introgression of useful traits into cultivated varieties are time consuming and costly, therefore production of secondary metabolites in plants lacking them or elevating levels of these metabolites in plants synthesizing them - albeit in insufficient levels - is an attractive goal.

The biosynthesis of terpenes in plants has been extensively studied and many genes coding for the pathways steps from precursors to final products were discovered (Wither and Keeling, 2007 Agro Microbial Biotechnology 73:980-990; Sallaud et al., 2009 Plant Cell 31:301).

Due to the widespread infestation of crop and ornamental plant species with pest insects such as *B. tabaci* and the greenhouse whitefly *Trialeurodes vaporariorum*, resulting in great economic losses, means of regulating plant natural defense molecules to repel pests has received a renewed interest of scientists and plant breeders.

It is known that manipulation of transcription factors can regulate complex pathways in animals and plants involving numerous target genes. This may result in increased expression of useful compounds. Alternatively, blocking transcription factors may lead to decreased or completely suppressed production of undesirable compounds and/or removal of unwanted traits.

Several transcription factors controlling genes involved in plant secondary metabolism were identified, cloned and showed high efficiency in regulating complex metabolic pathways. For

instance, the transcription factor WRKY was shown to regulate δ -Cadenine Synthase A, a sesquiterpene synthase that catalyzes the first step of pathway leading to production of gossypol in cotton (Xu et al., 2004 Plant Physiol 135:507-515).

Moreover, while overexpression of individual genes of the biosynthetic pathways was shown to provide limited success, perhaps, due to poor substrate availability, genome-wide expression of the flavonol-specific transcription factor, AtMYB12, in tobacco not only regulated the phenylpropanoid pathway, but also modulated other metabolic pathways that led to increased flux availability to this pathway, and eventually to an increased resistance against *Spodoptera lituralis* and *Helicoverpa armigera* insects (Misra et al., 2010 Plant Physiol 152: 2258-2268).

MYB transcription factors have been indicated also to activate multiple enzymes required for production of glucosinolates, crucifer-specific secondary metabolites, in *Arabidopsis*. MYB51 was shown to activate the indolic glucosinolates biosynthesis and confer enhanced resistance to the herbivorous pest *Spodoptera exigua* in plants overexpressing it (Gigolashvili et al., 2007 Plant J 50: 886-901). Other MYB transcription factors, such as MYB76, MYB28 and MYB29, are shown to regulate enzymes involved in the production of aliphatic glucosinolates (Gigolashvili et al., 2007 Plant J 51: 247-261; Gigolashvili et al., 2008 New Phytol 177:627-642).

There is a need in the art to provide transcription factors regulating terpene biosynthesis, in particular transcription factors that have an effect on the biosynthesis of mono- or sesquiterpenes in plants or other organisms leading to the production of terpene compounds that repel or attract insects, or other organisms.

Summary of the invention

An embodiment of the invention herein provides an isolated, synthetic or recombinant nucleic acid sequence selected from the group including: a) a nucleic acid sequence of SEQ ID NO: 1; b) a nucleic acid sequence that encodes a polypeptide having an amino acid sequence of SEQ ID NO: 2 or an amino acid sequence that is at least 60% identical to the amino acid sequence of SEQ ID NO:2; c) a nucleic acid sequence that is at least 60% identical to the nucleic acid sequences of (a) or (b), and encodes a transcription factor that regulates terpene biosynthesis; d) a nucleic acid sequence encoding a polypeptide comprising which has been derived, by way of one or more amino acid substitutions, deletions or insertions, from the polypeptide having the amino acid sequence of SEQ ID NO: 2 and which polypeptide is functionally equivalent to the polypeptide consisting of the amino acid sequence of SEQ ID NO: 2; e) a nucleic acid sequence that hybridizes under stringent conditions to the nucleic acid sequences of (a), (b), (c), or (d); and f) a nucleic acid sequence that hybridizes under stringent conditions to the (optionally reverse) complement

of the nucleic acid sequences of (a), (b), (c), or (d) and that encodes the transcription factor that regulates terpene biosynthesis. Further embodiments include a chimeric gene comprising such nucleic acid sequence, a vector comprising such nucleic acid sequence or such chimeric gene, and a host cell comprising such chimeric gene or such vector.

5 A related embodiment of the invention provides a polypeptide with DNA binding activity that regulates terpene biosynthesis in a plant, such polypeptide having an amino acid sequence selected from the group of: (a) an amino acid sequence of SEQ ID NO: 2; (b) the amino acid sequence according to (a) in which at least one amino acid is substituted, deleted, inserted or added and wherein the polypeptide is functionally equivalent to the
10 polypeptide consisting of the amino acid sequence of SEQ ID NO: 2; and (c) an amino acid sequence that is at least 60% identical to the amino acid sequence according to (a). The polypeptide (i.e., transcription factor) of the invention herein is capable of binding a nucleic acid sequence of a promoter that is operably linked to at least one gene involved in terpene biosynthesis in the plant. For example, the gene is selected from the group
15 comprising a *Terpene Synthase 5 (TPS5)* and a *Terpene Synthase 11 (TPS11)*. For example, the promoter comprises a trichome-specific promoter.

An alternative embodiment of the invention provides a method for increasing the production of at least one terpene in a plant, involving up-regulating a transcription factor that positively regulates at least one gene involved in terpene biosynthesis in the plant, such
20 gene preferably being selected from the group comprising *TPS5* and *TPS11*. For example, up-regulating the transcription factor involves modifying the plant to have an increased copy number of a nucleic acid sequence of SEQ ID NO:1, or a sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 1, compared to a non-modified plant of the same genetic background, thereby increasing the level of the at least one terpene in the
25 modified plant.

In a preferred embodiment, a method is provided for increasing the level of at least one terpene in a plant involving: (a) contacting a plant cell or plant protoplast with a composition that includes a vector having a nucleic acid sequence of SEQ ID NO:1 or a fragment thereof with at least 60% identity to the sequence of SEQ ID NO:1; (b) selecting
30 the plant cell or plant protoplast transformed with the vector wherein the plant cell or plant protoplast overexpresses the nucleic acid sequence or the fragment thereof so that overexpression results in an increased level of the at least one terpene in the cell compared to a non-transformed plant cell or plant protoplast; and (c) regenerating the plant from the transformed cell or protoplast wherein the plant has an increased level of the at least one
35 terpene compared to a non-transformed plant of the same genetic background. Within the scope of the invention is also a method for increasing production of at least one terpene in a population of plants by selectively breeding the transformed plant to produce the population

of transformed plants having the increased level of the at least one terpene compared to a non-transformed plants of the same genetic background.

In yet another embodiment, a method is provided for reducing production of at least one terpene in a plant that involves down-regulating a transcription factor that positively
5 regulates at least one gene involved in terpene biosynthesis in the plant, wherein the gene preferably is selected from the group comprising *TPS5* and *TPS11*. For example, down-regulating the transcription factor may involve modifying the plant to have a mutation in a nucleic acid sequence of SEQ ID NO: 1 or a sequence with at least 60% identity to the
10 nucleic acid sequence of SEQ ID NO: 1, or a nucleic acid sequence of SEQ ID NO: 3 or a sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 3, wherein the mutation results in a decrease of the level of the transcription factor or a loss of function of the transcription factor, compared to a non-modified plant of the same genetic background, thereby reducing the level of the at least one terpene in the modified plant. For example, the mutation may include a substitution, a deletion, an insertion or an addition of at
15 least one nucleotide. In a related embodiment of the method, down-regulating the transcription factor may involve modifying the plant to have an increased level of RNA having a nucleic acid sequence at least in portion complementary to a nucleic acid sequence of SEQ ID NO: 1, or a sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 1, or a nucleic acid sequence of SEQ ID NO: 3 or a sequence with
20 at least 60% identity to the nucleic acid sequence of SEQ ID NO: 3, compared to a non-modified plant of the same genetic background, thereby decreasing the level of the at least one terpene in the modified plant.

In a particularly preferred embodiment, a method is provided for reducing terpene levels in a population of plants, comprising steps of: (a) providing at least one plant within a
25 population of plants comprising a mutation in a gene comprising a nucleic acid sequence of SEQ ID NO: 1 or a nucleic acid sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 1, or a nucleic acid sequence of SEQ ID NO: 3 or a nucleic acid sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 3; and (b) selectively breeding the at least one mutant plant of (a) to produce the population of plants
30 having the reduced terpene levels.

In particular, the terpene of the method includes at least one of a monoterpene and a sesquiterpene that repel insects. For example, the monoterpene includes at least one compound selected from the group of: linalool, β -myrcene, para-cymene, γ -terpinene, α -terpinene, and α -phellandrene. For example, the sesquiterpene comprises at least one
35 compound selected from the group of: nerolidol, germacrene, R-curcumene, S-curcumene and 7-epizingiberene.

Alternatively, the terpene of the methods herein includes at least a monoterpene and a sesquiterpene that attract insects. For example, the monoterpene includes at least one compound selected from the group of: β -phellandrene, limonene and 2-carene. For example, the sesquiterpene comprises at least β -caryophyllene.

5 Generally, the invention herein is concerned with attracting or repelling insects including sap-sucking insects and blood-sucking insects. For example, the sap-sucking insects include psyllids, whiteflies, aphids, mealybugs, plant hoppers and scale insects. For example, the blood sucking insects comprise mosquito, ticks and midges. Insects including thrips, mites and leaf hoppers are also within the scope of the invention.

10 Generally, the plant of the methods of the present invention is at least one crop plant selected from the group of: tomato, pepper, eggplant, lettuce, oilseed rape, broccoli, cauliflower, cabbage crops, cucumber, melon, pumpkin, squash, peanut, soybeans, corn, cotton, beans, cassava, potatoes, sweet potatoes and okra. The plant also includes at least one plant selected from a *Solanaceae* family. The methods herein may also be directed at at
15 least one ornamental plant selected from the group of: hibiscus, poinsettia, lily, iris, rose and petunia.

An embodiment of the invention also provides a plant obtainable or obtained by the methods described herein. Additionally, the invention pertains to a plant comprising a chimeric gene as provided herein. Such plant may be a genetically engineered plant
20 comprising a nucleic acid sequence of SEQ ID NO:1 and variants and fragments thereof. For example, the plant may belong to the *Solanaceae* family.

A tissue culture initiated from the plants described herein, e.g., a transformed or modified plant, also is within the scope of the invention. Such tissue culture has enhanced production or secretion of at least one terpene, terpene isomer, or terpene analog.

25 A related embodiment provides a method for producing a terpene, terpene isomer, or terpene analog involving isolating the terpene, terpene isomer or terpene analog from the tissue culture resulting from the transformed plant.

A final embodiment of the invention provides a method for a marker-assisted introgression of a terpene into a plant including the steps of: (a) identifying a difference in a
30 gene encoding an amino acid sequence of SEQ ID NO:2 between a plant from the *Solanaceae* family and a sexually compatible plant wherein the gene with the difference comprises a molecular marker associated with presence of the terpene in the plant; (b) making a cross between the plant having the molecular marker and the sexually compatible plant; (c) screening a progeny resulting from the cross for the presence of the molecular
35 marker; and (d) identifying the plant within the progeny having the molecular marker and thereby identifying the plant producing the terpene.

Brief description of the drawings

Figure 1 is a schematic drawing of a PJVII-GUS-sYFP1, a pMON999 vector with a modified multiple cloning site. The *S. lycopersicum Terpene Synthase 5 (SITPS5)*, also referred to as *Monoterpene Synthase 1 (SIMTS1)*, promoter was cloned between *SacI* and *XbaI* sites and β -glucuronidase (GUS) fused to a yellow fluorescent protein (sYFP1) was cloned between *XbaI* and *BamHI* sites.

Figure 2 is a set of bar graphs showing quantification of GUS expression in isolated trichomes from stems of transgenic 5' *SITPS5* promoter deletion plants and controls.

Average relative GUS activity for each promoter construct (full length (fl), 1045bp, 805bp, 612bp, 408bp, 207bp), empty vector control and untransformed Money maker plant. Trichome-specific activity is lost at 408bp and then partially restored at 207bp. The 207bp fragment was chosen for the Y1H assay.

Figure 3 is set of bar graphs showing results of Quantitative Real Time PCR for a candidate gene 19(6).

Figure 3 panel A shows tissue-specific expression of the gene in tomato Money maker leaves, whole stem including trichomes, bald stem and isolated trichomes. The highest level of expression was observed in the isolated trichomes.

Figure 3 panel B shows trichome expression of the gene in control and jasmonic acid-sprayed plants (JA), (values corrected for actin).

Figure 4 is set of bar graphs showing average relative GUS activity in *N. benthamiana* plants. Five week old *N. benthamiana* leaves were co-infiltrated with *A. tumefaciens* GV3101 cultures carrying various promoter:GUS reporter and the 35S:19(6) effector construct. Average relative GUS activity normalized for 35S:luciferase activity is shown. Activities of these promoter:GUS reporter constructs in *N. benthamiana* with a control 35S:RFP effector construct were approximately 0.05. Enzymatic GUS activity of the crude extract was determined spectrophotometrically using 4-methylumbelliferyl β -D-glucuronide (MUG) as a substrate.

Figure 5 shows a genomic nucleic acid sequence of SEQ ID NO: 3 encoding the transcription factor TF 19(6). The underlined capital letters herein denote exons and the underlined small letters indicate introns. The capital letters in bold show start and stop codons. The capital letters which are not underlined indicate the putative TF19(6) 4 kb promoter region upstream of the start codon and 2 kb 3' UTR region downstream of the stop

codon. The position of the sequences encoding the putative zinc finger motifs are indicated in a light shade of gray.

Detailed description of the invention

5 The present invention relates to a transcription factor, referred to herein as TF19(6), that regulates terpene biosynthesis in plants. The TF19(6) cDNA comprises a nucleic acid sequence of 1056 bp (SEQ ID NO:1) with a single open reading frame encoding a polypeptide of 351 amino acids (SEQ ID NO:2).

Accordingly, the skilled artisan would understand that by expressing the sequences of the
10 present invention in a plant, one may change the expression of one or more genes involved in terpene biosynthesis. By affecting the expression of the genes, one may alter a plant phenotype to include plants with an improved resistance to insects or other pests and pathogens, or plants attracting beneficial organisms, e.g., predators or parasitoids of pest insects or plant pollinators, or other organisms.

15 The sequences of the present invention may originate from any species, especially from plant species, or from any other sources including recombinant or synthetic.

The transcription factors of the invention may also include a DNA regulatory sequence that regulates expression of one or more genes in a plant when a transcription factor having one or more domain binding sites binds to the regulatory sequence of one or more genes in the
20 pathway.

The present invention relates also to methods for modifying a plant phenotype by employing one or more polynucleotides or polypeptides of the invention for altering the expression of one or more genes of the terpene biosynthesis pathway. Alternatively, the polynucleotides and peptides of the invention have a variety of additional uses including, without limitation,
25 use as substrates for further reactions such as inducing mutations, performing PCR reactions, use as substrates for cloning including digestion and ligation reactions, identifying exogenous or endogenous modulators of the transcription factors, use in the production of recombinant protein, use as diagnostic probes for the presence of complementary or partially complementary nucleic acids, or the like.

30 The terpene biosynthesis pathway refers to the pathways leading to the formation of various terpene molecules. The terms "terpenes" and "terpenoids" are used herein interchangeably, and refer to hydrocarbons having a carbon skeleton derived from isoprene units (C_5H_8). Terpenes are subdivided into groups based on their carbon number and may be cyclic or acyclic molecules. The five-, ten-, fifteen-, twenty- and thirty-carbon terpenes are referred to
35 as hemi-, mono-, sesqui-, di-, and triterpenes, respectively.

For example, the term "monoterpenes" refers to a class of terpenes that consists of two isoprene units and have the molecular formula ($C_{10}H_{16}$). Monoterpenes include but are not

limited to cyclic monoterpenes, including myrcene, (*Z*)- and (*E*)-ocimene, linalool, geraniol, nerol, citronellol, myrcenol, geranial, citral a, neral, citral b, citronellal; monocyclic monoterpenes, including limonene, α - and γ -terpinene, α - and β phellandrene, terpinolene, menthol, carveol; bicyclic monoterpenes including α -pinene, β -pinene, myrtenol, myrtenal, verbanol, verbanon, pinocarveol; and tricyclic monoterpenes, including tricyclene.

As used herein, the term “sesquiterpenes” refers to a class of terpenes that consists of three isoprene units and has the molecular formula $C_{15}H_{24}$. Sesquiterpenes include but are not limited to cyclic sesquiterpenes, including farnesene; monocyclic, including zingiberene and humulene; bicyclic, including caryophyllene, vetivazulene, guaizulene; tricyclic, including longifolene, copaene, patchoulol.

As used herein the term “diterpenes” refers to terpenes consisting of four isoprene units and have the molecular formula $C_{20}H_{32}$. Known diterpenes include, for instance, taxol.

As used herein, the term terpenes refers also to terpene analogs, such as alcohols, aldehydes, ketons and esters, and isomers, including stereoisomers and tautomers.

Reference to specific isomers herein such as α - and/or β - isomers does not preclude the skilled artisan to use other isomers and appreciate that the other isomers or mixtures of isomers can substitute for the isomer specifically mentioned, as long as these are functional. For example, it was observed that transformed potato and *Arabidopsis* plants overexpressing the *Nerolidol Synthase 1* gene from strawberry emitted not only linalool but also linalool derivatives including *E*-8-hydroxy linalool, *Z*-8-hydroxy-linalool and *E*-8-hydroxy-6,7-dihydrolinalool (Aharoni et al., 2006 *Phytochemistry Review* 5:49-58). Metabolic engineering of terpenoids in plants by overexpressing enzymes catalyzing steps in the terpene biosynthesis pathway was shown to be successful to generate substantial levels of terpenoids (Lewinson et al., 2001 *Plant Physiol* 127:1256-1264; Aharoni et al., 2003 *Plant Cell* 15:2866-2884; Lucker et al., 2004 *Plant J* 39: 135-145; Lucker et al., 2004 *Plant Physiol* 134: 510-519; Aharoni et al., 2006 *Phytochemistry Review* 5:49-58). However, manipulation of the expression of the individual genes of the pathways sometimes shows limited success due to lack or poor availability of essential precursors. Regulation of the expression of genes involved in metabolic pathways using transcription factors was shown to modulate additional pathways and lead to availability of precursors (Misra et al., 2010 *Plant Physiol* 152:2258-2268).

Terpenes are synthesized from the common precursor isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) through two distinct biosynthesis pathways: the mevalonate pathway found in the plants cytosol and in eukaryotes and the deoxyxylulose-5-phosphate (DXP) pathway found in the plant plastids and in prokaryotes. In the mevalonate pathway, the biosynthesis of IPP starts from the conversion of three molecules of acetyl-CoA to mevalonate followed by subsequent sequential phosphorylation

of mevalonate to diphosphomevalonate followed by decarboxylation to produce IPP. In the DXP pathway, the production of DXP starts from one molecule each of pyruvate and glyceraldehyde-3-phosphate catalyzed by 1-deoxy-D-xylulose-5-phosphate synthase and results in the production of IPP and DMAPP in contrast to the mevalonate pathway where
5 IPP is the sole product. Generally, sesquiterpenes are synthesized from the relevant precursors through the mevalonate pathway in the cytosol, and monoterpenes and diterpens are produced through the DXP pathway in plastids. Exchange of precursors between plastids and cytosol was also observed.

In both pathways, the IPP is further isomerized to DMAPP by the IPP isomerase with
10 subsequent formation of the higher molecular weight acyclic polyprenyl pyrophosphate precursors by prenyl transferases to form the acyclic pyrophosphate terpene precursors. For example, these reactions produce ten-, fifteen-, and twenty-carbon precursors geranyl-pyrophosphate (GPP), farnesyl-pyrophosphate (FPP), geranylgeranyl-pyrophosphate (GGPP), respectively. Prenyltransferases are divided into two main classes based on
15 stereochemistry of the double bond formed at each elongation cycle. Prenyltransferases leading to the formation of double bonds in cis-configuration are called cis- or Z-prenyltransferases and prenyltransferases leading to the formation of double bonds in trans-configuration are called trans-, or *E*-prenyltransferases.

Terpene synthases (TPSs) are the enzymes catalyzing the cyclisation of the acyclic
20 precursors in the multi-step reactions producing the carbon skeleton of terpene, monoterpene or sesquiterpene compounds. For example, the initial step of the catalyzed cyclisation may be the ionization of the diphosphate group to form an allylic cation. The substrate then undergoes isomerizations and rearrangements which can be controlled by the active site of an enzyme. The product, for example, may be an acyclic, mono-, di or
25 tricyclic terpene.

It is known in the art that GPP and neryl diphosphate (NPP), the cis-isomer of GPP, are the substrates for monoterpene biosynthesis, and that FPP and GGPP are the respective substrates for sesquiterpene synthases and diterpene synthases (Chen et al., 2011 Plant J
66:212-229; Schillmiller et al., 2009 Proc Natl Acad Sci 106:10865-10870; Tholl 2006 Curr
30 Opin Plant Biol 9:297-304; Wang and Ohnuma, 2000 Biochim Biophys Acta 1529:33-48). Some TPSs produce a single product, but many produce multiple products from the same precursor, or can produce multiple compounds depending on the precursor supplied (Van Schie et al., 2007 Plant Mol Biol 64:251-263).

Induced terpene biosynthesis was observed to correlate with induced expression of
35 terpene synthases (Navia-Gine et al., 2009 Plant Phys Biochem 47: 416-425; Herde et al., 2008 Plant Cell 20: 1152-1168).

Plants emit volatiles consisting of a mixture of mono- and sesquiterpenes to repel pest insects or to attract beneficial insects, e.g., predators or parasitoids of pest insects or plant pollinators. Two terpene synthases *TPS5* (formerly *Monoterpene Synthase 1*, *MTS1*) and *TPS4* (formerly *Monoterpene Synthase 2*, *MTS2*) were identified in tomato (Van Schie et al., 2007 Plant Mol Biol 64:251-263; Falahara et al., manuscript in preparation; Bleeker et al., manuscript in preparation). The study shows that *TPS5* is expressed in trichomes and that *TPS5* protein catalyzes formation of linalool from GPP *in vitro*. *TPS5* expression in tomato trichomes was observed to be low under normal condition and elevated after induction with jasmonic acid (Van Schie et al., 2007 Plant Mol Biol 64:251-263; Kharel and Koyama, 2003 Nat Prod Rep 20:11-118). Van Schie's study also found that *TPS4* is expressed in stems, roots and in trichomes and that this gene encodes an enzyme leading to the formation of β -myrcene, β -phellandrene and sabinene from GPP (Van Schie et al., 2007 Plant Mol Biol 64:251-263).

Although monoterpenes appear to dominate terpenes identified in *S. lycopersicum*, majority of TPS genes mined from tomato genome are sesquiterpene synthases (Falahara et al., manuscript in preparation). Sequencing of cDNAs derived from trichomes of *S. lycopersicum* and *S. habrochaites* resulted in identification of multiple TPS sequences having similarities to known sesquiterpene synthases (Bleeker et al., manuscript in preparation). For instance, in *S. lycopersicum* database, transcripts for TPS9 (formerly germacrene C synthase; van Der Hoeven et al., 2000 Plant Cell 12: 2283-2294), TPS12 (formerly β -caryophyllene/ α -humulene synthase; van Der Hoeven et al., 2000 Plant Cell 12: 2283-2294), TPS15, TPS16, TPS17 and TPS31 (formerly LeVS2 because of similarity to vetispiradiene synthase from potato) were identified. In *S. habrochaites*, cDNA sequences were identified to have similarities to the TPS9, TPS12, TPS14, TPS15 and TPS17.

The compounds which can be regulated by the transcription factor of the present invention include, without limitation, terpene compounds that attract insects such as β -phellandrene, limonene and 2-carene, referred to herein as "attractants". The compounds regulated by the transcription factor of the present invention also include, but are not limited, to terpenes that repel insects such as R-curcumene, S-curcumene, β -myrcene, para-cymene, γ -terpinene, zingiberene, 7-epizingiberene, α -terpinene and α -phellandrene, referred to herein as "repellents". These compounds are described in the patent application WO 2010/099985 which is incorporated herein by reference.

Additionally, sesquiterpenes present in trichomes of *S. lycopersicum* cv. Moneymaker and *S. habrochaites* PI127826 can also be regulated by the transcription factor herein (See Table below; Bleeker et al., manuscript in preparation).

Table. Sesquiterpenes present in trichomes of *S. lycopersicum* cv. Moneymaker and *S. habrochaites* PI127826.

sesquiterpene	<i>S.</i> <i>lycopersicum</i>	<i>S. habrochaites</i>
Azulene	x	
α -copaene	x	
β -elemene		x
caryophyllene	x	x
γ -elemene	x	x
Alpha humulene	x	x
β -farnesene		x
β -acoradiene		x
Curcumene		x
germacrene D	x	x
Zingiberene		x
Cuparene	x	
β -bisabolene		x
β -sesquiphellandrene		x
Valencene		x
germacrene C	x	
selina-3.7(11)-diene		x
germacrene B		x

Trichome-specific promoters were identified in tomato that regulate expression of the *TPS5* and *TPS11* genes. (The *TPS5* promoter was referred to as the *MTS1* promoter and the *TPS11* promoter was referred to as the *STS1* promoter in the international patent application WO 09/082208 incorporated herein by reference).

As used herein, the term "promoter" refers to a nucleic acid sequence that is capable of initiating transcription of a nucleic acid sequence to which it is operably linked. The promoter controls transcription of one or more genes, located upstream of the transcription site of the gene. Structurally the promoter is characterized by the presence of a binding site for DNA-dependent RNA polymerase, transcription initiation sites and any other DNA domains (cis regulatory elements), including, but not limited to transcription factor binding sites, repressor and activator protein binding sites, and any other sequences of nucleotides known to one of skill in the art to act directly or indirectly to regulate the rate of transcription from the

promoter. Examples of eukaryotic cis regulatory elements include the TATA box located approximately 25 base pairs upstream of the transcription site, the CAAT box located 75-80 base pairs upstream of the initial transcription site, enhancer or silencer elements. The promoters of the invention herein include constitutive promoters that are active in all tissues and organs of the organism but preferably tissue-specific promoters that are active mainly in specific tissues such as trichomes. Especially included are the promoters of *Terpene Synthase 5*, i.e., linalool synthase, or *Monoterpene Synthase 1*, and *Terpene Synthase 11* found in plants belonging to the genus *Solanum* or other plant species.

Polynucleotides of the invention

The present invention provides an isolated, recombinant, or synthetic polynucleotide encoding the polypeptide or variant polypeptides of the transcription factor provided herein. An embodiment of the invention provides an isolated, recombinant or synthetic nucleic acid sequence of SEQ ID NO:1, or a variant thereof which is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to a nucleic acid sequence of SEQ ID NO:1 and which encodes a transcription factor having an amino acid sequence as shown in SEQ ID NO:2, or a fragment or variant thereof that is capable of regulating terpene biosynthesis.

In a further aspect of this embodiment, the nucleic acid sequence encodes a peptide portion of TF19(6) polypeptide having DNA binding activity. Such DNA binding domains binds to a specific target DNA sequence, and have an amino acid sequence that is different from those of known zinc finger domain DNA binding proteins.

In another embodiment, a full length genomic nucleic acid sequence of SEQ ID:NO: 3 is provided that contains coding and non-coding nucleic acid sequences.

The terms "nucleic acid sequence," "nucleotide sequence", "nucleic acid," and "polynucleotides" are used interchangeably meaning a sequence of nucleotides. A nucleic acid sequence may be a single-stranded or double-stranded deoxyribonucleotide, or ribonucleotide of any length, and include coding and non-coding sequences of a gene, exons, introns, sense and anti-sense complimentary sequences, genomic DNA, cDNA, miRNA, siRNA, mRNA, rRNA, tRNA, recombinant nucleic acid sequences, isolated and purified naturally occurring DNA and/or RNA sequences, synthetic DNA and RNA sequences, fragments, primers and nucleic acid probes. Due to the degeneracy of the genetic code various nucleic acid sequences may encode the same amino acid sequence. Any nucleic acid sequence encoding TF19(6) or variants thereof is referred herein as TF19(6) encoding sequence.

"Isolated" refers to a nucleic acid sequence that is removed from its natural environment and which is substantially free from other nucleic acid sequences, and the nucleic acid

sequence does not contain portions of unrelated sequences such as functional genes or polypeptide coding regions. An isolated molecule may be obtained by any methods or combination of methods including molecular biology, biochemical and synthetic techniques. This limitation does not pertain to nucleic acid sequences encoding genes or coding regions artificially added to the nucleic acid sequence after isolation.

"Recombinant nucleic acid sequence" refers to a combination of nucleic acid sequences that are joined together using recombinant DNA technology.

"Recombinant DNA technology" refers to molecular biology procedures to join together nucleic acid sequences as described for instance in Laboratory manuals edited by Weigel and Glazebrook, 2002 Cold Spring Harbor Lab Press; and Sambrook et al., 1989 Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

A fragment of a polynucleotide of SEQ ID NO:1 refers to a nucleic acid sequence comprising contiguous nucleotides of the polynucleotide of the invention herein that is preferably at least 15 bp, at least 30 bp, at least 40 bp, at least 50 bp and/or at least 60 bp in length.

Preferably the fragment of a polynucleotide comprises at least 25, at least 50, at least, 75, at least 75, at least 100, at least 150, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100 contiguous nucleotides of the polynucleotide of the invention. Without being limited, the fragment of the polynucleotides herein may be used as a PCR primer, and/or as a probe, or for anti-sense gene silencing.

The "primer" refers to a short nucleic acid sequence that is hybridized to a template nucleic acid sequence and is used for polymerization of a nucleic acid sequence complimentary to the template.

In a related embodiment of the invention, PCR primers and/or probes for detecting nucleic acid sequences encoding TF19(6) are provided. The skilled artisan would be aware of methods to synthesize degenerate or specific PCR primer pairs to amplify a nucleic acid sequence encoding TF19(6) or fragments thereof, based on SEQ ID NO:1 (see Dieffenbach and Dveksler, 1995 PCR Primer: A Laboratory Manual, Cold Spring Harbor Laboratory Press; McPherson et al., 2000 PCR Basics: From Background to Bench, 1st ed., Springer Verlag, Germany). A detection kit for nucleic acid sequences encoding TF19(6) may include primers and/or probes specific for nucleic acid sequences encoding TF19(6), and an associated protocol to use the primers and/or probes to detect nucleic acid sequences encoding TF19(6) in a sample. Such detection kits may be used to determine whether a plant has been modified, i.e., transformed with a sequence encoding TF19(6).

It is clear to the skilled artisan that mutations, deletions, insertions, and/or substitutions of one or more nucleotides can be introduced into the DNA sequence of SEQ ID NO:1 or

shorter fragments thereof. Generally, a mutation is a change in the DNA sequence of a gene that can alter the amino acid sequence of the protein encoded by the gene.

To test a function of a variant DNA sequences according to the invention, the sequence of interest is operably linked to a selectable or screenable marker gene and expression of the

5 reporter gene is tested in transient expression assays with protoplasts or in stably transformed plants. It is known to the skilled artisan that DNA sequences capable of driving expression are built as modules. Accordingly, expression levels from shorter DNA fragments may be different than the one from the longest fragment and may be different from each other. Embraced by the present invention are also functional equivalents of the nucleic acid
10 sequence coding the transcription factor of the present invention, i.e., nucleotide sequences that hybridize under stringent conditions to the nucleic acid sequence of SEQ ID NO:1.

A stringent hybridization is performed at a temperature 65°C and most preferably at 55°C in double strength (2x) citrate buffered saline (SSC) containing 0.1% SDS followed by rinsing of the support at the same temperature but with a buffer having reduced SSC concentration.

15 Such reduced concentration buffers are typically one tenth strength SSC (0.1xSSC) containing 0.1% SDS, preferably 0.2xSSC containing 0.1% SSC and most preferably half strength SSC (0.5xSSC) containing 0.1% SDS. Functional equivalents of the transcription factor from other organisms can be found by hybridizing a nucleic acid sequence with SEQ ID NO:1 with genomic DNA isolated from other organisms. The skilled artisan knows
20 methods to identify homologous sequences in other organisms. Such newly identified DNA molecules then can be sequenced and the sequence can be compared with the nucleic acid sequence of SEQ ID NO:1 and tested for functional equivalence. Within the scope of the present invention are DNA molecules having at least 60%, 70%, or 75%, preferably 80%, more preferably 90% and most preferably 95%, 96%, 97%, 98%, 99% or more sequence
25 identity to the nucleotide sequence of SEQ ID NO:1.

The percentage of sequence identity between two sequences is determined using computer programs that are based on standard alignment algorithms. Sequences are substantially identical when they share at least a certain minimal percentage of sequence identity as identified by standard computer programs. Computer programs that are preferred within the
30 scope of the present invention include without limitation the CGC program package (Devereux et al., 1984 Nucleic Acid Research 12:387), BestFit, BLASTP, BLASTN, and FASTA (Altschul et al., 1990 J Mol Biol 215: 403), the algorithm of Meyers et al., 1988 Comput Appl Biosci 4 :11, or the algorithm of Needleman et al., 1970 J Mol Biol 48:443. Preferably, the sequence identity refers to the sequence identity over the entire length of the
35 sequence.

A related embodiment of the invention provides a nucleic acid sequence which is complementary or reverse complementary to the nucleic acid sequence according to SEQ ID

NO:1, such as inhibitory RNAs, or a nucleic acid sequence which hybridizes under stringent conditions to at least part of the sequence according to SEQ ID NO:1 or the reverse complementary sequence to SEQ ID NO:1 (e.g., the non-coding DNA strand).

The polynucleotides of the invention may be overexpressed in plant cells and the changes in the expression levels of a number of genes and/or proteins of the plant cells may be observed. Therefore, polynucleotides and polypeptides of the invention may be employed to change the expression of the genes and/or protein in plants, especially the genes and/or proteins of the terpene biosynthesis. Alternatively, polynucleotides or polypeptides may be employed in knockout plants that lead to changes in the expression levels of one or more genes to improve characteristics or traits of the plants especially traits associated with insect resistance.

The term "gene" means a DNA sequence comprising a region, which is transcribed into a RNA molecule, e.g., a mRNA in a cell, operably linked to suitable regulatory regions, e.g., a promoter. A gene may thus comprise several operably linked sequences, such as a promoter, a 5' leader sequence comprising e.g. sequences involved in translation initiation, a coding region of cDNA or genomic DNA, introns, and a 3' non-translated sequence comprising, e.g., transcription termination sites.

A "chimeric gene" or "recombinant gene" refers to any gene, which is not normally found in nature in a species, in particular, a gene in which one or more parts of the nucleic acid sequence are present that are not associated with each other in nature. For example the promoter is not associated in nature with part or all of the transcribed region or with another regulatory region. The term "chimeric gene" is understood to include expression constructs in which a promoter or transcription regulatory sequence is operably linked to one or more coding sequences or to an antisense, i.e., reverse complement of the sense strand, or inverted repeat sequence (sense and antisense, whereby the RNA transcript forms double stranded RNA upon transcription).

A "3' UTR" or "3' non-translated sequence" (also referred to as 3' untranslated region, or 3' end) refers to the nucleic acid sequence found downstream of the coding sequence of a gene, which comprises for example a transcription termination site and (in most, but not all eukaryotic mRNAs) a polyadenylation signal (such as AAUAAA or variants thereof). After termination of transcription, the mRNA transcript may be cleaved downstream of the polyadenylation signal and a poly(A) tail may be added, which is involved in the transport of the mRNA to the site of translation, e.g., cytoplasm.

As used herein, a molecular marker refers to any morphological, biochemical or nucleic acid based phenotypic difference that reveals a DNA polymorphism. Examples of molecular markers include, but are not limited to, AFLPs (amplification fragment length polymorphisms), RFLPs (restriction fragment length polymorphisms), SNPs (single

nucleotide polymorphisms), SSRs (single sequence repeats), and alike. For instance, the skilled artisan would understand how to detect a genomic polymorphism between *Solanum lycopersicum* and sexually compatible species comprising identifying a difference in a gene encoding an amino acid sequence of SEQ ID NO:2 and fragments thereof between the species and thus identifying a molecular marker. For example, a molecular marker can be identified in at least one of a nucleic acid sequence of SEQ ID NO: 1 and a nucleic acid sequence of SEQ ID NO: 3. The molecular marker so identified can be used in marker-assisted selection of plants having the desired composition of terpenes that repel or, alternatively, attract insects, or other organisms.

Polypeptides of the invention

An embodiment of the present invention provides a transcription factor, transcription factor homologous polypeptides, and variants thereof.

The phrase "transcription factor" refers to a protein that regulates expression of one or more genes involved in terpene biosynthesis in an organism. The transcription factor possesses one or more domains for binding DNA regulatory sequences and at least one conserved domains characteristic of a particular family of transcription factors. Transcription factors encompass transcription factors-activators stimulating expression of one or more genes involved in terpene biosynthesis and transcription factors-suppressors inhibiting transcription of the genes by binding to their regulatory sequences.

"Zinc finger protein transcription factor" refers to an activator or a repressor composed of a zinc finger protein domain and any of a variety of transcription factor effectors domains which effect or modulate expression of nucleic acid sequences in the vicinity of zinc finger protein binding site. Zinc finger domains generally are about 25 to 30 amino acid residues in length, and contain high number of cystein residues in a C-X_n-C-X_n-C-X_n-C-X_n-C – type motif where X denotes a variable amino acid and n indicates the number of X residues. X residues are generally polar and basic, and implicate the region as involving in binding nucleic acids. Zinc ions are essential components of zinc finger domain structure designed to interact and bind nucleotides of a nucleic acid molecule.

A protein is an amino acid sequence of any length linked by covalent peptide bonds, and includes oligopeptide, peptide, polypeptide and full length protein whether naturally occurring or synthetic. A polypeptide means an amino acid sequence of consecutively polymerized amino acid residues, for instance, at least 15 residues, at least 30 residues, at least 50 residues.

The term "isolated" polypeptide refers to an amino acid sequence that is removed from its natural environment by any method or combination of methods known in the art and includes recombinant, biochemical and synthetic methods.

An embodiment of the invention provides an isolated or recombinant polypeptide which has an amino acid sequence set forth in SEQ ID NO:2 or fragments, or variants or derivatives thereof.

The term "fragment" of the polypeptide of SEQ ID NO: 2 refers to a subsequence of the polypeptide of the invention that retains its biological function and capacity to alter transcript levels of the genes of the terpene biosynthesis pathways, and is preferably capable of binding a nucleic acid sequence of a promoter that is operably linked to at least one gene selected from the group comprising a Terpene Synthase 5 (TPS5) and a Terpene Synthase 11 (TPS11). The term may refer to a recombinant polypeptide and/or an aggregate polypeptide such as a dimer or multimer.

As used herein, the terms "variant" or "derivative" of the polypeptide set forth in SEQ ID NO:2 refers to polypeptides with substantial similarity of amino acid sequences to the polypeptide herein. The amino acid sequences of the polypeptide of the invention and variants thereof may differ by one or more deletions, additions, and/or substitutions of amino acids while retaining functional equivalence to the polypeptide. In one embodiment, variants of TF19(6) include, for example, proteins having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98%, 99% or more amino acid sequence identity over the entire length to SEQ ID NO:2. Amino acid sequence identity may be determined by pairwise alignment using the Needleman and Wunsch algorithm and GAP default parameters.

Variants also include proteins capable of binding a promoter sequence operably linked to another nucleic acid sequence preferably of a gene that regulates terpene biosynthesis, such as *TPS5* and/or *TPS11*, which have been derived, by way of one or more amino acid substitutions, deletions or insertions, from the polypeptide having the amino acid sequence of SEQ ID NO:2. Preferably, such proteins comprise from 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more up to about 100, 90, 80, 70, 60, 50, 45, 40, 35, 30, 25, 20, 15 amino acid substitutions, deletions or insertions. For instance, amino acids of the polypeptide of the invention may be modified based on similarity in hydrophobicity, hydrophilicity, solubility, polarity of amino acid residues, as long as the variant polypeptide remains functionally equivalent to the polypeptide of the invention. Alternatively, a variant may differ from the polypeptide of the invention by attachment of modifying groups which are covalently or non-covalently linked to the polypeptide backbone. The variant also includes a polypeptide which differs from the polypeptide of the present invention by introduced N-linked or O-linked glycosylation sites, and/or an addition of cysteine residues. The skilled artisan would understand how to modify an amino acid sequence and preserve biological activity using computer programs such as DNASTAR (Madison, WI, USA).

A variant or derivative of zinc finger polypeptide of the invention retains a capacity to alter transcript levels of the genes of the terpene biosynthesis pathways, including or in particular

the transcript levels of the genes encoding a *Terpene Synthase 5 (TPS5)* and/or *Terpene Synthase 11 (TPS11)*.

The polypeptide of the invention is capable of binding a promoter sequence operably linked to another nucleic acid sequence preferably of a gene that regulates terpene biosynthesis, such as *TPS5* and/or *TPS11*.

A nucleic acid sequence encoding the polypeptide of the invention is "operably linked" to another nucleic acid sequence, typically a coding gene, when the gene so linked is transcribed. Operably linked DNA sequences form contiguous reading frames to produce a "fusion protein," i.e., a protein composed of various protein "domains" or "motifs." The nucleotide sequence associated with the promoter sequence may be of homologous or heterologous origin with respect to the plant to be transformed. The sequence may be also entirely or partially synthetic. Regardless of the origin, the nucleic acid sequence associated with the promoter sequence will be expressed or silenced in accordance with promoter properties to which it is linked after binding to the polypeptide of the invention.

In a preferred embodiment of the invention, the associated nucleic acid may code for a protein that is desired to be expressed or suppressed throughout the plant at all times or, alternatively, in specific cells and tissues. Such nucleotide sequences preferably encode proteins conferring desirable phenotypic traits to the plant modified or transformed therewith. More preferably, the associated nucleotide sequence leads to the production of terpenes in the plant. Preferably, the nucleotide sequence encodes *TPS5* or *TPS11* and terpenes that confer insect resistance, disease resistance, resistances to other pests and/or attraction of beneficial organisms, e.g., predators or parasitoids of pest insects or plant pollinators.

Homologous polynucleotides and polypeptides

Homologs, paralogs, orthologs and any other variants of the sequences herein are expected to function in a similar manner by making plants that repel or resist insects, or alternatively, attract beneficial organisms, e.g., predators or parasitoids of pest insects or plant pollinators. Homologous sequences are sequences that share substantial sequence identity or similarity to the nucleic acid sequence of SEQ ID NO:1. Homologous sequences may be derived from any plants including monocots or dicots, and especially crops including but not limited to tomato, pepper, eggplant, lettuce, sunflower, oilseed rape, broccoli, cauliflower and cabbage crops, cucumber, melon, watermelon, pumpkin, squash, peanut, soybeans, cotton, beans, avocado, onion, endive, leek, roots such as arrowroot, carrot, beet, turnip, radish, yam, cassava, potatoes, sweet potatoes and okra. Homologous sequences may also be derived from crop species including maize, barley, pearl millet, wheat, rye, sorghum, rice, tobacco and forage grasses. Homologous sequences may be derived from tree species and fleshy fruit species such as lemons, tangerines, oranges, grapes, peaches, plums, currant,

cherries, melons, strawberry, and mango, or from ornamental plant species such as hibiscus, poinsettia, lily, iris, rose and petunia, and the like. Additionally, homologous sequences may be derived from plant species that are wild relatives of crop plant species. For example, homologous sequences may be derived from nightshade *Atropa belladonna* which is a wild relative of a cultivated tomato *Solanum lycopersicum*, or teosinte species related to maize.

Homologous sequences include orthologous or paralogous sequences. Methods of identifying orthologs or paralogs including phylogenetic methods, sequence similarity and hybridization methods are known in the art and are described herein.

Paralogs result from gene duplication that gives rise to two or more genes with similar sequences and similar functions. Paralogs typically cluster together and are formed by duplications of genes within related plant species. Paralogs are found in groups of similar genes using pair-wise Blast analysis (Feng and Dollittle, 1987 J Mol Evol: 25:351-360) or during phylogenetic analysis of gene families using programs such as CLUSTAL (Thompson et al., 1994 Nucl. Acid Res 22:4573-4680; Higgins et al., 1996 Methods Enzymol 266:383-402). In paralogs, consensus sequences can be identified characteristic to sequences within related genes and having similar functions of the genes.

Genes encoding regulatory elements and transcription factors are conserved in eukaryotes. For instance, plant species that have common ancestors are known to contain many transcription factors that have similar sequences and functions. These sequences are referred to as orthologs. The skilled artisan can identify orthologous sequences and predict the functions of the orthologs, for example, by constructing a phylogenetic tree for a gene family of one species using CLUSTAL or BLAST programs.

A method for identifying or confirming similar functions among homologous sequences is by comparing of the transcript profiles in plants overexpressing or lacking (in knockouts/knockdowns) related polypeptides. The skilled artisan will understand that genes having a similar transcript profiles, with greater than 50% regulated transcripts in common, or with greater than 70% regulated transcripts in common, or greater than 90% regulated transcripts in common will have similar functions.

Homology refers to a sequence similarity, or identity between a polypeptide or a fragment thereof and a reference sequence. A homology of polypeptide sequences are determined based on the number of amino acid sequences in the positions shared by the polypeptides. Homologous sequences encompass amino acid sequences of polypeptide of the present invention modified by chemical or enzymatic means known in the art. See Ausubel et al. (eds) 2000 Current Protocols Mol Biol, Willey & Sons, New York. .

Polypeptides with "substantial identity" refers to sequences of sufficient similarity to the transcription factor of SEQ ID NO:2 which retains biological function of the transcription

factor when overexpressed, ectopically expressed or knocked out in a plant. Polypeptide sequences that are at least 50% identical to the polypeptide of the present invention are considered sufficiently identical.

5 Regulation of terpene biosynthesis

An embodiment of the invention provides a polypeptide, a polynucleotide, a fragment thereof or a chimeric gene or vector that may be used to up- or down regulate expression of genes involved in terpene biosynthesis, preferably selected from at least one of *Terpene Synthase 5 (TPS5)* and *Terpene Synthase 11 (TPS11)*, and thereby modify downstream products of the pathway.

For example, a polypeptide, a polynucleotide, a fragment thereof or a vector of the invention is used to regulate volatiles emitted by plants to repel pest insects or to attract beneficial insects, e.g., predators or parasitoids of pest insects or plant pollinators, or to regulate the transcription level of the genes involved in the terpene biosynthesis and to alter transcript profiles of such genes. An altered transcript profile of the gene refers to the transcript profile that is substantially different from the transcript profile of the correspondent gene in a reference state. Differences and similarities between expression levels are evaluated by statistical methods known in the art.

An embodiment of the invention provides methods to modify terpene levels in a plant by up- and down regulating expression of *TPS5* and *TPS11* genes and thereby modifying levels of at least one terpene from a group of: linalool, neralidol, germacrene, α -humulene, β -caryophyllene, β -elemene, β -phellandrene, limonene, 2-carene and zingiberene (7S configuration), β -curcumene, β -myrcene, para-cymene, γ -terpinene, 7-epizingiberene, α -terpinene and α -phellandrene.

A related embodiment of the invention provides a method for increasing a terpene in a plant by up-regulating a transcription factor of the invention that positively regulates terpene biosynthesis. For instance, up-regulating may increase the level of terpenes which are part of chemical defense of the plants and thereby repel insects or pathogenic organisms.

In yet another embodiment, the invention provides a method for reducing a terpene in a plant by down-regulating a transcription factor that positively regulates terpene biosynthesis. For instance, down-regulation may change the profile of the volatile terpene compounds emitted by the plants thereby making the plants less attractive to insects or other pests. Generally, methods for up- and down-regulation of expression of transcription factors in plant or animal systems are well known to the skilled artisans. Up-regulation may result from overexpression of a protein or polypeptide in a whole plant, plant cells or specialized plant tissues, such as trichomes. Alternatively, the promoter may be altered to up-regulate expression of transcription factors of the invention.

Down-regulation occurs at DNA level by interfering with transcription of the genes thereby decreasing expression of the genes. Alternatively, down-regulation occurs at RNA level by interfering with protein translation from RNA molecules, or by interfering with RNA splicing to produce mRNA species. Down-regulation at RNA level is achieved through RNA

5 interference (RNAi) approach using double stranded RNAs (dsRNAs), small hairpin RNAs (shRNAs), micro RNAs (miRNAs) or small interfering RNA (siRNAs). Phenomenon of RNA interference is also known in the art as cosuppression, post transcriptional gene silencing, and quelling. See Hamilton and Baulcombe, 1999 Science 286: 950-952; Hammond and Hannon, 2001 Nature Rev Gen 2: 110-119; Baulcombe, 2007 Science 315:199-200.

10 An embodiment of the invention provides a method for reducing terpene levels in a population of plants by providing plants mutagenized by either chemical or physical methods, detecting a mutated plant within a population such that the plant has decreased expression of the transcription factor of the invention that positively regulates terpene biosynthesis and selectively breeding the mutated plant to produce the population of
15 mutated plants thereby reducing terpene levels in the plants. An alternative method is provided for increasing terpene levels in a plant population by detecting and selecting a mutated plant within the population that has decreased expression of the transcription factor that negatively regulates terpene biosynthesis.

For example, methods for the detection of a mutation in a target sequence in a member of a
20 mutagenized population is disclosed in WO 2007/037678. The method involves isolating DNA of mutagenized plants, pooling DNA, amplifying the target sequence, i.e., the nucleic acid sequence of TF19(6) with a pair of primers from the DNA pool, determining the nucleic acid sequences of the amplification fragments using high throughput sequencing, identifying mutations by clustering (aligning) the sequences of the fragments, screening the identified
25 mutations for modified functions of the target sequence and identifying members carrying the mutation.

The sequencing may be conducted by methods known in the art, including the dideoxy chain termination method (Sanger sequencing), and high-throughput sequencing methods, such as the methods disclosed in WO 03/004690, WO 03/054142, WO 2004/069849, WO
30 20041070005, WO 2004/070007, and WO 2005/003375, by Seo et al. (2004) Proc. Natl. Acad. Sci. USA 101:5488-93, and technologies of Helios, Solexa, US Genomics, and the like, which are herein incorporated by reference. It is most preferred that sequencing is performed using the apparatus and/or method disclosed in WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007, and WO 2005/003375, which are
35 herein incorporated by reference.

An alternative embodiment of the invention provides a method to alter gene expression in a plant, plant tissue or plant cell. For instance, the polynucleotide of the invention may be

overexpressed in a plant, cell or tissue. The term "overexpresssion" refers to an increased expression of a gene in a plant, tissue or a plant cell compared to expression in a non-altered or wild type plant, tissue or cell, at any stage of development or location of the gene. Overexpression occurs when gene encoding transcription factor of the invention is under control of a strong constitutive or a tissue specific promoter.

Alteration of expression of a polynucleotide of the present invention also results in "ectopic expression" which is the different expression pattern in a transgenic or mutant plant and in a control or wild-type plant. Alteration of expression occurs from interactions of polypeptide of the invention with exogenous or endogenous modulators, or as a result of chemical modification of polypeptide. The term also refers to an altered expression pattern of the polynucleotide of the invention which is altered below the detection level or completely suppressed activity.

An alternative embodiment of the invention provides a method for increasing production or level of a terpene in a population of plants including the following steps: contacting a plurality of plant cells with a composition which includes a vector incorporating a nucleic acid molecule set forth in SEQ ID NO:1; detecting and selecting a transgenic plant cell within the plurality of the cells such that the cell has an increased level of a transcription factor that positively regulates terpene biosynthesis compared to a control plant cell thereby increasing the terpene in the cell; regenerating the cell into a plant and selectively breeding the plant to produce the population of plants with the increased terpene.

The vectors for inserting transgenes into the genome of host cells are well known in the art. As used herein, the term "vector" refers to a nucleic acid molecule engineered using molecular biology methods and recombinant DNA technology for delivery of foreign or exogenous DNA into a host cell. Typically the vector is a DNA molecule that consists of a transgene insert and a nucleic acid backbone. Vectors include plasmids, viruses, cosmids and artificial chromosomes. Binary or co-integrated vectors into which a chimeric gene is inserted may be used for transforming plants.

The chimeric gene generally includes a promoter sequence operably linked to a nucleic acid sequence of a coding gene to be transcribed in the host cells. Alternatively, the promoter sequence may already be present in a vector so that the nucleic acid sequence which is to be transcribed is inserted into the vector downstream of the promoter sequence. Vectors are typically engineered to have an origin of replication, a multicloning site and a selectable marker.

Examples of selectable markers are described below. The skilled artisan would know that different antibiotic or herbicide selectable marker are applicable to different target species. Selectable markers that are routinely used in plant transformation include the *npt II* gene conferring resistance to kanamycin, paromomycin, geneticin, and related antibiotics (Veira

and Messing, 1982 Gene 19: 259-268; Bevan et al., 1983 Nature 304: 184-187) the bacterial *aadA* gene encoding aminoglycoside 3'-adenyltransferase conferring resistance to antibiotics streptomycin or spectinomycin (Goldschmidt-Clermont, 1991 Nucl Acid Res 19: 4083-4089), the *hph* gene conferring resistance to hygromycin (Blochliger and Diggelmann, 1984 Mol Cell Bio 14: 2929-2931). Other markers that can be used include a mutant EPSP gene conferring resistance to glyphosate (Hinchee et al., 1988 Biotechnology 6: 915-922), a mutant acetolactate synthase (ALS) gene conferring resistance to imidazoline or sulphonylurea herbicides (Lee et al., EMBO Journal 7: 1241-1248), a phosphinothricin acetyltransferase gene which confers resistance to herbicide phosphinothricin (White et al., 1990 Nucl Acid Res 18: 1062; Spencer et al., 1990 Theor Appl Genet 79: 625-631). Selection markers resulting in positive selection such as phosphomannose isomerase gene are also used (see WO 93/05163).

An embodiment of the invention provides recombinant expression vectors comprising a nucleic acid sequence of the invention fused to associated nucleic acid sequences such as, for instance, promoter sequences. The vector is used to transform the host cells and the chimeric gene is preferably inserted in the nuclear genome or into the genome of cell organelles, i.e., mitochondria or plastids, such that the expression of the nucleic acid sequence is driven by the activity of the promoter. See Arabidopsis, A Laboratory Manual Eds. Weigel and Glazebrook, Cold Spring Harbor Laboratory Press (2002) and Maniatis et al., Molecular Cloning, Cold Spring Harbor Laboratory Press (1982).

Methods for obtaining transgenic plant cells and plants are well known in the art and include but are not limited to *Agrobacterium*-mediated transformation of plant explants, particle bombardment of plant explants, transformation of plant explants using whiskers technology, transformation using viral vectors, electroporation of plant protoplasts, direct uptake of DNA by protoplasts using polyethylene glycol, microinjection of plant explants and/or protoplasts. *Agrobacterium*-mediated transformation is a preferred method to introduce the nucleic acid molecule of the invention into plant explants. *Agrobacterium tumefaciens* harbors a natural vector called Ti plasmid which was engineered to make it suitable for introduction of exogenous nucleic acid molecules into plant genomes. For genetic transformation, plant-derived explants are incubated with suspension of *Agrobacterium* cells followed by cultivation of the explants on the medium containing a selective agent that promotes growth and regeneration of the transformed cells only.

Methods for detecting transformed or modified plant include without limitation, Southern Blot Analysis and PCR based methods. Methods for analyzing terpene content in modified plants using gas chromatography-mass spectrometry (GC-MS) are known in the art and are described in Schillmiller et al., 2009 Proc Natl Acad Sci 106:10865-10870 and Adams, 2007 Identification of Essential Oil Components by Gas Chromatography/Mass

Spectrometry, 4th ed., Allured Pub Corp., Carol Stream, IL. The resulting transformed or modified plant may be used in a conventional breeding to produce more transformed or modified plants with altered profile of terpene compounds.

An embodiment of the invention provides antibodies specific for polypeptides of the invention or variants thereof. The skilled artisan would understand that the transcription factor or variants thereof can be expressed and purified in a heterologous expression system, for instance *Escherichia coli*, and used to raise monoclonal or polyclonal antibodies specific for polypeptides of the invention. Antibodies can be also raised against synthetic polypeptides from the amino acid sequences of the transcription factor or variants thereof. Methods of raising antibodies are known in the art and described in Harlow and Lane, 1988 Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies are used to screen expression libraries from the plants and may be used, for example, in a method for identifying plants emitting volatiles that attract beneficial insects or alternatively repel pest insects.

Host cells, plants and tissue cultures of the invention

A phrase "host cell" or "transformed cell" refers to a genetically engineered cell that includes at least one nucleic acid molecule, especially a chimeric gene encoding a desired protein or a nucleic acid sequence which upon transcription yields a terpene useful to repel or attract insect pests. The host cell is preferably a plant cell, but may also be a fungal cell, a yeast cell or a bacterial cell. The host cell preferably includes the chimeric gene integrated into the nuclear or organelle genomes, but may also contain the gene extra-chromosomally.

An embodiment of the invention herein also provides a genetically engineered plant cell that includes a nucleic acid sequence set forth in SEQ ID NO:1 and a plant regenerated from the cell. A genetically engineered plant of the invention includes the plant having an increased level of a terpene compared to a non-genetically engineered plant of the same genetic background. As used herein a genetic background refers to the genotypic base of a breeding line or population of organisms.

A related embodiment of the invention also provides for a tissue culture obtained from the transgenic plant such that the culture has enhanced production or secretion of terpene, terpene isomer or terpene analog, and a method for isolating the same from the tissue culture of the invention.

Suitable cells for expression of a polypeptide of the invention include prokaryotic and eukaryotic cells such as plant cells. Plants suitable for expression of a polypeptide of the invention include, but are not limited to crop species, which are natural hosts of insect pests such as tomato, pepper, eggplant, lettuce, sunflower, oilseed rape, broccoli, cauliflower and cabbage crops, cucumber, melon, watermelon, pumpkin, squash, peanut, soybeans, cotton,

beans, cassava, potatoes, sweet potato and okra. Crop species also include maize, barley, pearl millet, wheat, rye, sorghum, rice and forage grasses. Additionally, plant hosts include tree species, fleshy fruit species such as grapes, peaches, plums, strawberry and mango, and ornamental species such as hibiscus, poinsettia, lilies, iris, rose and petunia.

- 5 Especially preferred are plants belonging to the family *Solanaceae*, including plants that belongs to genera *Solanum*, *Capsicum*, *Nicotiana*, *Petunia*, and the like. In a preferred embodiment, vegetable species especially of genus *Solanum* are included, for example, tomato (*S. lycopersicum*), eggplant (*S. melongena*), pepino (*S. muricatum*), and the like. As used herein, the phrase “crop species” refers to plants cultivated for purposes of
- 10 obtaining food, feed or plant derived products including carbohydrates, oils and medicinal ingredients.

Insects controlled by regulation of terpene levels in plants

- An embodiment of the invention provides a method for increasing resistance to a pest insect
- 15 by down regulating or alternatively up regulating the genes of the terpene biosynthesis which would result in altering profiles of volatile terpenes emitted by plants to attract beneficial insects or repel pest insects.

- As used herein, the phrases “plant insects” or “plant pests” refer to insect species that infest and damage host crop and ornamental plants. An “infestation” refers to presence of a large
- 20 number of pest organisms in a field or greenhouse, on the surface of a host plant or on anything that might contact a host plant, or in the soil. Insect pests include sap-sucking insect pests, such as psyllids, whiteflies, aphids, mealybugs, plant hoppers and scale insects and share a common property, namely the utilization of plant sap as their food source. Insect pests also include thrips, cicada, mites and leaf hoppers.

- 25 The term “whitefly” or “whiteflies” refers to species of the genus *Bemisia*, especially *B. tabaci*, species of the genus *Trialeurodes*, especially greenhouse whitefly *T. vaporariorum* and banded winged whitefly *T. abutino*. All biotypes of *B. tabaci* such as biotype Q and B, are also included as well as any developmental stage, such as eggs, larvae, pupae and adults.

- 30 As used herein, the term “aphids” refers to plant insect pests belonging to the family *Aphididae*, including but not limiting to *Aphis gossypii*, *A. fabae*, *A. glycines*, *A. nerii*, *A. nasturtii*, *Myzus persicae*, *M. cerasi*, *M. ornatus*, *Nasonovia* especially *N. ribisnigri*, *Macrosiphum*, and *Brevicoryne*.

- The term “insect pests” also refers herein to insects of the order *Diptera* including but not
- 35 limiting to blood sucking or biting insects attacking animals, especially mammals. Blood

sucking ticks are also included. Such insects may act as vectors of human and/or mammalian diseases such as malaria.

"Insect vectors" are insects that are capable of carrying and transmitting viruses to plants. In the context of mammalian disease vectors, insect vectors are insects which attack mammals and can potentially transmit diseases to mammals, such as mosquitoes, which are able to transmit the parasite *Plasmodium* to humans or heartworm to canines.

Preferably, the modified plants of the invention develop enhanced resistance to one or more pest insects.

"Insect pest resistance" is an enhanced ability of modified plants of the present invention to withstand attacks of one or more pests compared to wild type or control plants. Methods to assess insect pest resistance in plants are known in the art. For example, disease symptoms may be scored visually at one or more time points after infestation or contact with an insect pest. Alternatively, insect pests may be detected and optionally quantified during infestation of the plants in an assay. A modified plant shows enhanced pest resistance if the number of insect pests detected in the tissue is significantly lower compared to the number of insects detected in control. Preferably, a significant increase in average yield of modified plants of the present invention compared to control (e.g. at least 1%, 2%, 5%, 10% or more) when plants are grown under insect pest pressure provides an indirect measurement of enhanced resistance to pest insects. Statistical analyses are employed to determine existence of significant differences.

The invention now having been fully described it is exemplified in the Examples below and in the claims, which are not to be construed as further limiting. References cited herein are hereby incorporated by reference in their entireties.

Sequences referred to herein:

A nucleotide sequence of SEQ ID NO:1 encoding a transcription factor TF19(6):

ATGGCTAATTTCTTTTCATTAGGTGGGAATCAAGAACAACAACATCAAGAAATTAGCAGC
AGCCAAGCATTAGTACCCACAGAGAGTAATAATTGGTTTTTGTACAGAAATGAACATCAT
CATCATCATCATAATCAAGAAATACCCAACACTTACAAAGGTTTTGAGTTATGGCAAAGT
GGTAACACTCCACAACACCAACACCAACACCAACAACAACAGTTTCGTCATCC
GATTTATCCTTTGCAAGATCTTTATTCCACTGATGTTGGATTAGGGGTTGGGCCAAGCAG
AAGTGGCTTTGATATATCTGCAGGTGATCATGAGGCGTCGAGGTCGGGATTTCGTGATGA
TGAGGAGTGGTGGAGGAGGAATAAGTTGCCAAGATTGTGGGAACCAAGCTAAGAAAGA
TTGTCAACATATGAGGTGTAGGACTTGTTGTAAGAGTAGAGGGTTTCAGTGTCAAACCTCA
TGTGAAAAGTACTTGGGTTCCAGCAGCTAAAAGGAGAGAAAGGCAACAACAACCTTGCTG
CTTTGCAACAACAACAAGGACATAATAATAATAATAATCATAAGAATAAAAGGC
AAAGGGAGGATCCAAGTGCTTCTTCTCTTGTGTCTACTCGTTTGCCTTCAAACACTAATG

GGTTAGAAGTGGGAAAATTTCCATCAAAAGTACGTACAAGTGCTGTATTTTCAGTGTATTC
 AAATGAGTTCAATTGAGGATGATGAAGATCAATTAGCATATCAAGCTGCTGTGAGCATTG
 GTGGACATGTTTTCAAAGGAATTTTATATGATCAAGGTCATGAAAGTCAGTACAATAACA
 TGGTTGCAGCCGGAGGCGATACGTCTTCCGGTGGTAGTGCTGGCGGAGTTCAGCACCA
 5 CCACCATAATTCCGCTGCAGTAGCTACCGCCACCACTACAAGTGGTGGCGATGCTACTG
 CAGCGGGTCCATCGAATTTTCTAGATCCTTCTTTATTTCCAGCTCCCCTTAGCACTTTTAT
 GGTAGCTGGTACGCAATTTTTTCCACCTTCAAGATCTCCTTGA

An amino acid sequence of SEQ ID:2 of the transcription factor TF19(6):

10 MANFFSLGNGNEQQHQEISSSQALVPTESNNWFLYRNEHHHHHHNQEIPNTYKGFELWQS
 GNT PQHQHQHHQQQQQFRHPYPLQDLYSTDVGLGVGPSRSGFDISAGDHEASRSGFVM
 MRSGGGGISCQDCGNQAKKDCQHMR CRTCKSRGFQCQTHVKSTWVPAKR RERQQQL
 AALQQQQQGHNNNNNNHKNKRQREDPSASSLVSTR LPSNTNGLEVGFPSKVRTSAVFQ
 CIQMSSIEDDEDQLAYQAAVSIGGHVFKGILYDQGHESQYNNMVAAGGDTSSGGSAGGVQ
 15 HHHHNSAAVATATTTSGGDATAAGPSNFLDPSLFPAPLSTFMVAGTQFFPPSRSP

A genomic nucleic acid sequence of SEQ ID NO: 3 encoding the transcription factor TF 19(6).

GCAAGTCAAATTGTATATGCTCTTAATTAGGGTTTATGCACCTAGAAGTAGTATTTTAATC
 20 GCATATTAGTTATCGATGTTCTTAAATTAATCGCCTATTAGTTATCGATGTTCTTAAATAT
 AATTGACCTAATTACAAAATTAAGATAGA ACTGATTATATTTTTTCAATTTTATCCTTACAA
 GGAGCAATTCTTTTTGAAAGTATGAACCACTTTGTAAAGTTTTTTTTTAAAAAAATCTTAA
 AGGAGTAAATCAGTAAA ACTACCTTTCATATTTATGATTTTTTAAGAAGCATGTAAAGAAA
 AAATAGAAAATCAATATAGAACGAAAAAAGAATTTTATAAAACCTCATAACTTAATAAAAA
 25 GAATCATATTTATAAGAAATATTTTCTTCCCACATGGAATAATAAATTGCACAACTGTAA
 ATATTCTCTACTACAATATAATGTTATAACACACGTATACCGTTGGTTTTTCAGTATAAATA
 TAATATCCATATTTTAGATATATTAGCTGTAAAAACAATATAATATGGATGGAGACAAGT
 TAAATGTATGTAATTTTACCTTGTAATCCCAATTCTCAATATATATATATATATATATATA
 TATATATTCCTCCAAGACAAAACATTGGATTTTTATTCTAGGAACTTGAATTAAGAATTCA
 30 ATTTACTCGTAAAAATTA AAAAGAAATTTCTTATGATCTTATCAAATATTTAATAGGTGAAT
 AGTTAAATTTGACAAGTTAAATTAAGATGTATGTCCATCACCTCATCATAATTCAAATTATT
 TTAACAAATATCCTTAATCATGATCTTCTTTCTTTTTAGGTTGAAATAATTATCATTAGATT
 TGTACATAGTATAGGAATATATTAAGAGCTAATATATCTAAAATGTGGATAAATAAAACAA
 TTCTTGCTCAAAATTTTAAATAGTTTTAATACTTTTACAATACTTGACACGCGGCATTATAT
 35 AGCCAACAATTTTACGGGCTAAGACATAACATTTATCTTGAAATTCTCTATTATTAATCA
 TTAGCTTAGATTGTCTGAGTTTTTGAAGGTCTTTTATTTAGTTATAGGCAATTTTACCTAG
 TTTTATAGAATTA AAATTATTGTCCGTTGTTATTTAGGTAAGAAAAAAAGTTAATAAATC

AGACAAGAAAATATAAAGAACCGAAATAATTATGTAATGCCTAAAATAGTTGGTTTTATAT
ACATAAAGACTGTTGAAAATTGAAATTAATATTGCGGCTCCTTCATTTATTGGTATTACTG
TTAATTACGTGATTGAAAGGAAAAAAAAAAGTTTTACCAAAAAAAAAAGTATAAAAAATAAGTTT
TGTACTTTTATGTCAACAGTTAGTCATCAATAGTTACTGCTATAATACTAGGTGCCAACACT
5 ATGTATAATTCTGAATGTGATAATAATTTCTGGAAAAAAAAAATTAAAGGATATTTGATTTGA
TATGGTCCTAGATAATGTAGATGATGAAGGGGTGTTAATTAGTCGTTTCAAATTGATAGG
TTATTTTGAAAATTGTGTCATATTAATTGTTTATATTTTCAATGATGTATGATTAATAATTAA
AATTTTTGAATCTGTCTTAATCGTTTTGGTTTCTCTTTGATTTAGGTATAATTCAAATTGAT
GGTTAATTTTTTTAAACGTCATCTAACAACCTATAAAATTTGATAAAAAATATTTAAAATTTA
10 CAATAACATAAAATGAAAATATGTTTTCCAACCTATACCAATTTAGGAGGAGAAACATAGTT
ATTGTTTTACTATTATCGCTAGTATTATGAATGAGATATGAAAATTTATATTAATTTATATT
GGAATCTATAATTGATTTTTATTAATAAAAAAATTAAGTGCGTGTACTTTGACATTTTTTTTTGTT
TTTAACTCGGCATTCAAAGTTCATATTGAAGTTTTAACTAAATTCGAATCGCACTCTTCAG
AGCAATGCAGGGATGGGTCTCCCAACAACATTTTGTGCGATAGTCTATACCCAGAGCTCG
15 AACTTAAGACCTCTGATTAAGAATAAAATACTTCATTTATAAACTGATCATCTTAATATTTT
CAAATTTAAATGTCACATATTTTCTAAGATATCCTCGAAACATAATAATAAGTTGAAATG
TATAATGTTTGATTGAGACTAACTGAGGCGTTTATATATACAATCGTAGAATTAATAATAT
TTAATTGCCATCTGAAAATTAATTTAAATATTTATCTATGTACTATACCTTAATTAATTCTTT
CATGACAAACTTTCTTGGACATTTTTTTCATGAAAAATGCATATAACTTAAACAAGGCCGAT
20 ACCTTACACCCAAATTGGACAGTATATTTAGAAGAGGGGGAATAATGGTAAAGAGGGGCC
GGTATCAGGTTTACACAGAGATGAAAAGTTAGGTGGAGTTTATTTGTTCCGATGGATTTA
TCAGTTTTTTTCGTAGATTTTATATTTATATTAGATTCTTTTTTTTACGTATATATTAAATTAT
AATCCCTAAACAAATTGATTTAGAATCTCAAACCTCATAATCTTAACCTTCGCCCTAACCTT
TATATATATATATATATATAATATTTTTTAATATATCATTAGTTACACATTATTTTTTTATATT
25 ATGTTGTGTATTACTGATGAATAATGATTTATGGAAATACAAAAAGCTCTTATTCAGTAAT
ACATACATTAGTATGATCATCTTTTTTTCACATTCTTCCATCGCGATATATGTTTTTTTTTT
AACTATAACACAATAAACACTGCATTAAAAATAATTGTACATATTTTTTTGTGTCTCAATTT
ATGTGATACCTTTTAGTTTTTTAAGAGCTAAACAATTTAAATTTGAGCGAGAATTTACGCA
TGAAATTTTCGAAAATTCTAAAAAGAAATTTATATATTAATAAAAACTACGTAAAAATACTA
30 TAAGACACAATAATTGACAATTCAAAATATTTAAAAGTCAAAGATATACTTATTTGAATTC
AAAATCTGAAAAGTATCACATAAATAGGAGGAGAGAGTAACGAATATCAATATTAATGAT
ATATTATATTCACCACTAATATTCTTAAAAATAAATATTAACCACTTAATTCGATGT
GAATTATTAGTTTGATCCCTGAACTATTGACAGTATTATAAACACTCCTCTACTGGGTTAG
ATGAACTTAAATACACACTCGATCTTGTCACAATGATGAAATACACCCTAATGAAAATCAT
35 ATTTACTCTTCCTATTTCTAACCATCGGAAAGAGTCATCGTGGCTAGGAACTATACTAG
CGACCTACCCAATTCAATTATAGAAATTTTCGCGATCAATGATTGAAAATTTAGAATGTTTC
CAACACTTTATCTGTCAACTTTTTTATTAAGAGTTTCAAGCTCGTATAAGAATTTGAAATCA

CTTTTAGTATATCATGTAGTAGATCTAAATATATTTAAAATTATTATAAAATTTTTTTTAAAAA
ACTAATAATTCACACTAAATTGACAAATATCTTCAATACTTAGCTTCTCACTTATTTTATAC
GACCTACCAAACAATCGCGAACTTTTTAAGTTACTGCAAAGTGTAGCGGTAAAGAGAG
GGGAGGGGGGGGGGGGTAGTTGTGGTGCTTTTTAGCGTTGGCGGCGTTTGCAGAGCTG
5 TAATATATATAATATACCTTTTTCTATTAATGTACCCTCACTCACTCACTTCCTCTCCATAAT
TCTTTATACAAACAATCATTTTTTCTTAACTTGCTCTATTATAAAATTCACATTTTTTCTTTA
TATATACACATACATATAGAGCAAAAAAGAAGTTCTAATTTTGTAACCCCTTCAAAAAA
GAAAAATAATTTTTTTTGAGATCATAAATGAAGAAATCCAAGGGATACAAACATCATATTT
GTGTTATAAGTTGGTGCACCTTTTGTGGTATGGATTGTGATTAATCACTAATCATAATCAAG
10 ATTAACAACAAGTAATGGCTAATTTCTTTTCATTAGGTGGGAATCAAGAACAACAACATC
AAGAAATTAGCAGCAGCCAAGCATTAGTACCCACAGAGAGTAATAATTGGTTTTTGTACA
GAAATGAACATCATCATCATCATAATCAAGAAATACCCAACACTTACAAAGGTTTTGA
GTTATGGCAAAGTGGTAACACTCCACAACACCAACACCAACACCACCAACAACAACAAC
AGTTTCGTCATCCGATTTATCCTTTGCAAGATCTTTATTCCACTGATGTTGGATTAGGGG
15 TTGGGCCAAGCAGAAGTGGCTTTGATATATCTGCAGGTGATCATCAGAAACAGATTTAG
AATTTAACTTTATCTATTCAGTACTTTCTAAAGTACTTATAGATCTATAATTTAAGTTTGA
TAAATTTAATATTTATGTTCTAAACAATTCACAACATTTTGCAATTAGGGATTTTCGAAACGT
ATTTACTGAAGCATGTTAGAATTCCCAGCCTCGAAAAGGCATGGGAATTTGGTCTATGG
ACTTGGGAAATTCTCCATTCATGAGCTAACTTTTGAGGTTAAATTAGGTTTCATATGTCATA
20 TCTTTACATGATATCAGAGTAAGATTCATCTCAATTCCTTGTTACCAATATTGGCCCCC
ATATTATTGTGTCCACAATCTAGTTAACCTACGCTGGCCCCTCCATATTACAGTGTCCAC
GTTCTAGTTAACGAGATCTGGGCTTGCAGAAGAGTGTAAGAATTCAGAAAAAGGATGA
GTATTTGGTCTCCTTGTATAAACTTGAGCAATCCTTCCTTCATGAGCTAGCTAGTTTTGG
AATTAAGTTAGACTAGATGTCATATCTTTTAATATTTATGTTCTCACTGTAGAACCATATA
25 GCAACGAACTATAGTACTATTTGTTGCACCGCTCTCTCTATATATATCGTGCATATTAAG
TTCAATTGAATCTGTTGCTAAAAGGCGGGATGGGGATTATTATTGTGCAGGTGATCATGA
GGCGTCGAGGTCTGGGATTCGTGATGATGAGGAGTGGTGGAGGAGGAATAAGTTGCCA
AGATTGTGGGAACCAAGCTAAGAAAGATTGTCAACATATGAGGTGTAGGACTTGTTGTA
AGAGTAGAGGGTTTCAGTGTCAAACCTCATGTGAAAAGTACTTGGGTTCCAGCAGCTAAA
30 AGGAGAGAAAGGCAACAACAACCTTGCTGCTTTGCAACAACAACAACAAGGACATAATAA
TAATAATAATAATCATAAGAATAAAAGGCAAAGGGAGGATCCAAGTGCTTCTTCTCTTGT
GTCTACTCGTTTGCCTTCAAACACTAATGGTAAAGTACTTCATGTTTTTCTTACCTTTTCA
TTGCTACGTCTGTTTTAATTTAAAGGTCTTAGTTTGACTGAACATGAATATAAGATGTTGA
AATTGAAAAACGTAGATAAATATTTAAATTGAAACGAGGGAATAATTAATTTTTTTTGTG
35 TCACACAAAGACATAGAGTCTTGAGATCCATCATGTAAAGAAGATTAATTTGATCATTGC
CTAAATGAATTCTATATAAAGTAAGTCTATAGAGAAAAGAGACCCTATAGTAAATTTCGTCA
GCTTTTTCTTTTTCTATTTGTCATTCTCTTCTTCCATCATCACTCTTCTTTTTTATTACTCTA

CAAAAGATTGACAAAAATTCGTAATGAGATATATTCAAATTTTTGAGTTAATTATGAATTTT
TAATTCTAGTTAATAGAAAGTGTGAATAAATTATTTATATGTATTACTAACAAAAATAGCAAA
ACTAAACTTTTACTTGTACCCTTGCGCGTGTGTATGCACAATTTCTTCTCTTAGACCTAC
ACATGATATTTATCTCGACCCTAAAAAGATCACCATTATTCTTAATTTCAATTTTCGTCAAT
5 TTTTTTTTAAGATAATAACTATTATTTGAGTAATAATATATGTGACTTACCCAAAAAACTGT
TAGTGGAGTGAGTATTTGAGAAACCAACTCTCTAATTCATGTATAATAATTGGTGTTATCA
TATATTGTCATTAGTATTGGAATTAACCTATATATCTATTAGTAAATGTACTTTTGAAATAA
TAACTATTATTTGAGTAATAATATATGTTGCTTACCAAAAAAATAACTATTAGTTGAGTGG
CTATTAACCTCTCCAAATATGTATAATAATTGGTGTTATCATTTTCATTAGTATTGGAATTAA
10 CTTATATATATAGTAAATGCACTTGCAATTTCAAATTTTTTACCTGCTTTTCCTTTTAGTTC
GATTAATAATAATTGACTATTTTTCAAGCAAGTGTTTATTCTAACTTTTCAGATGAAATGT
TTAAAAAAACCACAAGATTAAATAGTGTTTTGATACATTTGACATATTTTTAGTTTTAGACC
ATAAAATTCAAATTGCTTTACTAAATTTTCGTGTCAAGTGATACTAGGTAAAAAAAATATT
TATTTGCAATACATTAGTCCAAATAAACCTAATTTTGTATTATGGAATTTTCATGTGTTATTT
15 TTAGGGTTAGAAGTGGGAAAATTTCCATCAAAAGTACGTAAAGTGCTGTATTTCAAGTGTA
TTCAAATGAGTTCAATTGAGGATGATGAAGATCAATTAGCATATCAAGCTGCTGTGAGCA
TTGGTGGACATGTTTTCAAAGGAATTTTATATGATCAAGGTCATGAAAGTCAGTACAATA
ACATGGTTGCAGCCGGAGGCGATACGTCTTCCGGTGGTAGTGCTGGCGGAGTTCAGCA
CCACCACCATAATTCCGCTGCAGTAGCTACCGCCACCCTACAAGTGGTGGCGATGCTA
20 CTGCAGCGGGTCCATCGAATTTTCTAGATCCTTCTTTATTTCCAGCTCCCCTTAGCACTT
TTATGGTAGCTGGTACGCAATTTTTTCCACCTTCAAGATCTCCTTGATCGTCCACATTGAT
AATATTGAGGTGTCTTTTTAATTTTTATGTCAAGAGATTTGTTTTTAATTGAAGTATTGATG
TTGAATTGAGTTGTTTACATTAATTCTCTTTGGATTCTACATGAAGTTGTTTTTTTTCTCT
AGTTCCTTATGGTTAATTATTGGTATCATATAGATTTGCTTTTTTATTTACGTTAAGATGA
25 TAATATAAGATAAGATGATAATATACTTAAATGTATATATGTTTTGGGTTGAGTCTTACGA
TTACTTATTATTAGAATTTTTTGTATGTGTATTCCGGCTCATAATGTGCCAAAAGATAACAA
AAGCAAAATTTAAGAGCATTACATAATATTATAAGTTTGTGATGGACTGTAAGTATATTT
TAGATTTTTTAATTAGAGTTTTTAAATTTAAACCTAAAAGAAATCGTATTTAAAAGAGCAG
TTTACCCTATAAGTGATTTTTTTAAGAATAAATATGGATTAGTCGAACCCAATAGTCGGGC
30 AACAGTTAGAAGCTAAAAAAGATTATAATTTTAAGAAAATACTTACTTTATAAAATTGAGA
TATTTGGTTAAGTTTTTLAGAGGGGGAAAAGAAATGTGCTTTTGAATAATAGCATGAATTA
ATCTTTACAATTAGAAAAAAGAAAATTAATAATACAAAAAGTAATTGTGAAATTAGGTCA
AGCACAACTAAGGTTCTAAACTGATTTTAAAAAAAACTTTTAAATTAATTAATCAACA
CAAAATTATTACTCTCCAAAAATATTTTCTACATAATACTTATCAAATAAATATATTTAGA
35 AAATTTGGCCAACTAACATGACTCTTCTTGATTAAGCACATAAATCAAGTTGTTAATAAA
ACTTTGGCTTTATAGCAATGACTCATTTGCTTTCAAACATAAAAAAATGAACAAACATTA
AATATATATTTAACGGAGTAAGATATATTCCAACTAGGACACTAGAAATGGTGAAAGCT

TAGTACGTTTGGAACATCAATTCAATTAACTCGAATGTCACTGTTTAACTTGTCTTAATA
 TATGTGATAATATTTGATGGATCTTAAATATTATTTCTTTAAAAAATAATTATTCGTTAGAA
 GGACAATAAGTGCTACAATGACTTAAATTTCTAAATTTTCAACTAGGCATAATCCTTCAAA
 ATAACTTTCATCATACTTTTGAATAATTAAATATGATATTATTGAAGTTATGTAAATTTTCAT
 5 GTTTCGGGCTTGTTTCGGGTTTTTTAAATATCAAATCAAATTATTCGTGTAAAATTTTAAA
 ATTATAAATCAGACCAAATTAATAAAATTCAGATTTTTTTCGGGGTTTTTCAACTCTGGGTTG
 ATTCGTATTTTTTCAAGTACCAAACCAAACCATTGTGTGCAATTTTTTAAATTTTTTAATCAAAC
 CAAACTAATAAACTTCGGATTTTTTCCAGATTTTTTAGATTTTTTTCGGGTAAAGTTTGCATACA
 AACATATAATTAACCTTGTGCTCCAAATATTTCTTTAGTCCAACCATAATATAATTATCTAAG
 10 GTATTTCTTGAAAAAATTACACAAAAGATGAGATGAGTATTGATGACACAAAAATATTCAA
 TAAAAAATAACAATAAATCATCATATAAAATAAATATTGTAAAGTCATAATGAAAATAATCA
 TAATTTAAAATTTTTAAATCATGCTAAAATAAGTTTAATAAGTATTAGTTACATTATTAATA
 TTTAAGGAAAACAAAAATTAGATTATGTAAATAAATATAAACTAAAGAACAAATATTCAA
 TATTATTGTCATTTTTAGTGTTGAATTGATTTTTTCTTTTTGCATTAGTATTAATTTAATTTT
 15 AATTTAAGCTTTATTATAATTATCAATCTATGAACTATAATCTATATTGGACCATTCCAAAT
 TCTATATTTTAACTTGAAACAATATATTAAGTTAAAACTATGAAATAGTATAAGAAAT
 ATTTTAAAATAATATCAACGTAAATATTTTATGTATAAAATAATTTTTACACATATAATATA
 AGGATTTTTTTCCCGATTTGATTCAATT

20

Examples

Example 1. Plant Material, Hormone Treatment and RNA Isolation

Tomato plants (*Solanum lycopersicum* cultivar Moneymaker) were grown in soil in a
 greenhouse with day/night temperatures of 23°C to 18°C and a 16/8 hours light/dark regime
 for 4 weeks. Trichomes were collected at the bottom of a 50 ml tube by vortexing stem
 25 pieces frozen in liquid nitrogen. The remaining cleaned stem segments were thoroughly
 brushed to remove all remaining trichome material (bald stem sample). Whole stems
 (including trichomes) and leaves were also frozen in liquid nitrogen, the material was ground
 and total RNA was isolated using Trizol (Invitrogen, Paisley, UK).

30

Jasmonic acid (JA) was applied to plants by spraying 1 mM solution made with tap water
 containing 0.05% SilwetL-77. Control plants were sprayed with tap water containing 0.05%
 SilwetL-77. Trichomes were collected 24 hours later and total RNA was isolated as
 described above. DNA was removed with DNase (Ambion, Huntingdon, UK) according to the
 manufacturer and cDNA was synthesized from 1,5 µg RNA using M-MuLV Reverse
 Transcriptase (Fermentas, St. Leon-Rot, Germany) according to the manufacturer in 20 µl
 35 volume that was diluted to 50 µl prior to using it for PCR.

Example 2. Constructs and Stable Plant Transformations

The full *SITPS5* promoter and a series of 5' deletions of it were generated using PCR. A *SacI* restriction enzyme site was created at the 5' end of each forward primer and an *XbaI* restriction enzyme site at the 3' end of the reverse primer. Fifty ng of plasmid DNA pKG1662adp-SIMTS1p:GUS were used as template with 0.25 units of Phusion Hot Start polymerase (Finnzymes, Espoo, Finland), each primer in a concentration of 0.4 mM and dNTPs in a concentration of 0.2 mM in 25µl reaction volume. MgCl₂ was added to the PCR mix with a final concentration of 0.3mM. The cycling program was set to 1 min 98°C, 30 cycles of 10 sec 98°C, 30 sec 58°C, 60 sec at 72°C, followed by 5 min final extension at 72°C and cooling to 12°C until removed from the thermocycler.

Primers used (5' -> 3' sequence, numbers indicate the position of the 5' nucleotide of each primer. The A of the start codon ATG is assigned to +1):

SI_TPS5p -18 R GCTCTAGATTTATTTGTTCTGCTCAA (SEQ ID NO: 4)

SI_TPS5p -1253 F2 CGAGCTCGTTTCATTCAAAGTAGTGG (SEQ ID NO: 5)

SI_TPS5p -1045 F CGAGCTCAGCTGAACCAAATCCCAA (SEQ ID NO: 6)

SI_TPS5p -805 F2 CGAGCTCGTCCTATTTTCCATATTG (SEQ ID NO: 7)

SI_TPS5p -612 F2 CGAGCTCATCAACAGTATTAAATGTGCTTC (SEQ ID NO: 8)

SI_TPS5p -408 F2 CGAGCTCAGTAATAATGAAAATCGCATCG (SEQ ID NO: 9)

SI_TPS5p -207 F2 CGAGCTCACATGTGCTATTTTATGCTA (SEQ ID NO: 10)

The 6 PCR products were purified using an Invitex (Palm Springs, CA, USA) column according to the manufacturer's protocol. They were subsequently double digested with *SacI* and *XbaI* and ligated upstream of the ATG start codon of β-glucuronidase fused to yellow fluorescent protein (sYFP1) in the *SacI* and *XbaI* sites of the vector pJVII, replacing CaMV 35S promoter (Figure 1). The PCR products were verified by sequencing and the expression cassettes (promoter fragment + GUSsYFP1 + NOS terminator) were transferred to the binary vector pBINplus (van Engelen et al., 1995 Transgenic Res 4: 288-290) by digesting with restriction enzymes *SacI* and *SmaI* and ligating in the multiple cloning site of pBINplus at the same restriction sites. These 6 constructs and an empty pBINplus vector were introduced to *Agrobacterium tumefaciens* strain EHA105 and used to create transgenic plants using explants derived from cotyledons of sterile seedlings of *Solanum lycopersicum* cultivar Moneymaker (MM), as previously reported (Cortina and Culianez-Macia, 2004, Plant Cell Tissue and Organ Culture 76: 269-275).

Example 3. Analysis of Transgenic Plants

One transgenic line was obtained from the empty pBINplus vector, four independent transgenic lines from the full length *SITPS5* promoter construct, three from the 1045bp

S/TPS5 promoter construct, five for the 805bp and 612bp *S/TPS5* promoter constructs, eight for the 408bp *S/TPS5* promoter construct and nine for the 207bp *S/TPS5* promoter construct. Insertion of the transgene was verified by PCR on genomic DNA isolated from leaves of the different T0 lines.

5 Primers used (5' -> 3' sequence):

pJVII_1182GUS_R CCACCAACGCTGATCAATTC (SEQ ID NO: 11)

pJVII_6936_F ATGTGCTGCAAGGCGATTAAG (SEQ ID NO:12)

The PCR was performed with Taq DNA Polymerase (Fermentas, St. Leon-Rot, Germany) in 25µl volume according to the manufacturer and the cycling program used was set to 2min
10 95°C, 30 cycles of 30sec 95°C, 30sec 58°C, 90sec at 72°C, followed by 5min final extension at 72°C and cooling to 12°C until removed from the thermocycler.

Initial YFP expression of the T0 plants was estimated under a fluorescence stereoscope and it was determined to be specific for the "head" of the type VI trichomes of the plants (data not shown). Trichomes were isolated from T1 plants as mentioned above, crude extracts
15 were prepared according to Jefferson R.A. *et al.*, 1987, EMBO J 6: 3901-3907 and the enzymatic GUS activity was determined spectrophotometrically using 4-methylumbelliferyl β-D-glucuronide (MUG) as a substrate (Figure 2).

Example 4. Yeast One Hybrid and Identification of Clone 19(6)

20 The 207bp *S/TPS5* promoter fragment showed trichome specific activity, although less strong than that of the full length promoter (Figure 2), and therefore this fragment was used for the yeast one hybrid (Y1H) assay.

An *EcoRI* restriction enzyme site was created at the 5' end of the forward primer and the reverse primer SI_TPS5p -18 R with an *XbaI* restriction enzyme site at the 3' end were used
25 in a PCR to generate the 207bp fragment for cloning. PCR was performed with Phusion Hot Start polymerase (Finnzymes, Espoo, Finland) as mentioned above, except extension time was 30 sec at 72°C.

Primer used (5' -> 3' sequence):

SI_TPS5pEcoRI_207F CGGAATTCACATGTGCTATTTTATGCTA(SEQ ID NO: 13)

30 The PCR fragment was purified using a Roche (Almere, Netherlands) column according to the manufacturer's protocol. Then it was double digested with *EcoRI* and *XbaI* and ligated in the multiple cloning site of pHISi vector (Clontech, Mountain View, CA, USA) at the same sites. After verifying the sequence, the construct was integrated in the yeast pj69-4a genome according to the Clontech MATCHMAKER One-Hybrid System manual. A cDNA
35 library created with mRNA from *Solanum lycopersicum* cultivar Moneymaker trichomes was screened 3 times according to the manufacturer's protocol (Clontech, Mountain View, CA, USA). Three Y1H screens yielded 76 clones, among which one putative transcription factor

19(6), appearing 32 times. The clone was sequenced using primers that fit on the library vector (pAD-GAL4-2.1, Stratagene, Santa Clara, CA, USA) and specific primers designed on the obtained sequence to get the full length clone.

Primers used (5' -> 3' sequence):

- 5 pActF TAATACCACTACAATGGATG (SEQ ID NO: 14)
- pAct_seqR CAACTGTGCATCGTGCAC (SEQ ID NO: 15)
- 19(6)_seqF TTATGGCAAAGTGGTAACA (SEQ ID NO: 16)
- 19(6)_seqF2 TCAGTGTCAAACCTCATGTG (SEQ ID NO: 17)
- 19(6)_seqF3 AAGTACGTACAAGTGCTG (SEQ ID NO: 18)

- 10 The candidate transcription factor was checked for JA-inducibility and trichome specificity by quantitative real-time PCR (Figure 3). The different tissue- and control and JA treated trichome- cDNA was obtained as described above.

Example 5. Real Time Quantitative PCR

- 15 RT-Q-PCRs were performed in the ABI 7500 Real Time PCR System (Applied Biosystems, Carlsbad, CA, USA) using the Platinum SYBR Green qPCR SuperMix-UDG kit (Invitrogen, Paisley, UK). Each 20µl reaction contained 0.25 µM of each primer, 0.1µl ROX reference dye and 1µl template cDNA. The cycling program was set to 2 min 50°C, 7 min 95°C, 45 cycles of 15 sec at 95°C and 1 min at 60°C and a melting curve analysis. Primer pairs were
- 20 tested for specificity and for linearity with a standard cDNA dilution curve. The tomato Actin gene (*ACT*) was used as constitutively expressed control gene.

Primers used (5' -> 3' sequence):

- ACT_QF TTAGCACCTTCCAGCAGATGT (SEQ ID NO: 19)
- ACT_QR2 AACAGACAGGACACTCGCACT (SEQ ID NO: 20)
- 25 19(6)_QF TACAAGTGGTGGCGATGCTAC (SEQ ID NO: 21)
- 19(6)_QR ACCTCAATATTATCAATGTGGACAATC (SEQ ID NO: 22)

Example 6. Transactivation Assay

- DNA binding activity was confirmed in transactivation assays. A *Nco*I restriction enzyme site
- 30 was created at the 5' end of a forward primer and a *Sac*I restriction enzyme site at the 3' end of a reverse primer and a full length cDNA 19(6) was generated in a PCR performed with Phusion Hot Start polymerase (Finnzymes, Espoo, Finland) as mentioned above. Fifty ng of plasmid DNA pAD-GAL4-2.1_clone19(6) were used as template.

Primers used (5' -> 3' sequence):

- 35 *Nco*I_19(6)F catgccATGGCTAATTTCTTTTCATTAGG (SEQ ID NO: 23)
- Sac*I_19(6)R cgagctcTCAAGGAGATCTTGAAGGTG (SEQ ID NO: 24)

The PCR fragment was purified using a Roche (Almere, Netherlands) column according to the manufacturer's protocol. Then it was double digested with *NcoI* and *SacI* and ligated downstream of 35S promoter in the same sites of the vector pKG1662-35S:GUS, replacing β -glucuronidase. The PCR product was verified by sequencing and the expression cassette

5 which included 35S promoter, transcription factor 19(6) and nos terminator, was transferred to the binary vector pBINplus (van Engelen et al., 1995 Transgenic Res 4: 288-290) by digesting with restriction enzymes *HindIII* and *EheI* and ligating in the multiple cloning sites of pBINplus digested with *HindIII* and *SmaI*. The construct was introduced to *Agrobacterium tumefaciens* strain GV3101.

10 Five week old *Nicotiana benthamiana* leaves were co-infiltrated with *A. tumefaciens* GV3101 cultures carrying various promoter:GUS reporter and the 35S:19(6) effector constructs. Specifically, the promoter constructs used were pBINplus-ShMKS1p:GUS, pBINplus-SITPS11p:GUS and pBINplus-SITPS5p:GUS. These constructs were made by cloning into a PJVII-GUSSYFP1 vector each of *SITPS5*, *SITPS11* or *Solanum habrochaites* methylketone

15 synthase 1 (*ShMkS1*; Fridman et al., 2005 Plant Cell 17:1252-1267, published on line March 16, 20005) promoter sequences between *SacI* and *XbaI* sites, and nucleic acid sequences encoding β -glucuronidase (GUS) fused to a yellow fluorescent protein (sYFP1) between *XbaI* and *BamHI* sites (Schematic drawing of PJVII-GUSSYFP1 is shown in Figure 1). A control pGreen-35S:RFP effector construct was also used (pGreen; Hellens et al.,

20 2000, Plant Molecular Biology 42: 819-832). In order to correct for infiltration errors *A. tumefaciens* GV3101 culture carrying pBINplus-35S:LUC was also added in each culture mix in a ratio of 5 (promoter:GUS): 5(35S:TF): 1(35S:LUC). Two days later leaf disks from the infiltrated areas were collected, frozen in liquid nitrogen and crude extracts were prepared in extraction buffer containing 25mM Tris phosphate pH 7.8, 2mM DTT, 2mM

25 CDTA pH 7.8, 10% glycerol and 1% Triton X-100. The enzymatic GUS activity was determined spectrophotometrically using 4-methylumbelliferyl β -D-glucuronide (MUG) as a substrate according to Jefferson R.A. et al., 1987, EMBO J 6: 3901-3907. The luciferase assay was performed using the same extraction buffer according to van Leeuwen et al., 2000, Plant Molecular Biology Reporter 18: 143a- 143t. Enzymatic GUS activity was

30 normalized for luciferase activity for each sample and results are presented in Figure 4. As shown in Figure 4, the 35S:19(6) effector construct activated GUS activity in pBINplus-SITPS5p:GUS more than 10-fold over the control pBINplus-ShMKS1p:GUS construct. Additionally, the effector construct was also able to activate GUS in pBINplus-SITPS11p:GUS 7-fold over the control.

These data confirmed TF 19(6) capability to activate *SITPS5* and *SITPS11* promoters and shows that TF19(6) may be used in regulating expression levels of the *TPS5* and *TPS11*, and alter terpene content in a plant, tissue or cell.

5 Example 7. TF 19(6) sequence identity

Polypeptide sequence identity was determined using BLAST algorithm described in Althaus et al. 1990 J Mol Biol 215:403-410. BLAST program is publicly available through the National Center for Biotechnology Information (NCBI) at the web site of the National Institute of Health, USA.

- 10 A BLAST homology search identified that amino acid sequence of TF 19(6) has 40.68 % identity over the length of the entire protein as compared to the Lateral Root Primordium (LRP1) protein, a member of *Arabidopsis thaliana* SRS (short internode related sequences)-type transcription factors with zinc finger motifs that are induced by auxins. Two zinc finger-type domains were found within TF 19(6): a zinc finger domain (amino acids 128-170) in the
- 15 N-terminal part and a small conserved LRP1-type domain in the C terminus (amino acids 224-272). The polypeptide of the invention also possess conserved DUF702 domain of unknown function characteristic of SRS-type transcription factors. The BLAST search also identified a tomato gene encoding a protein 45% homologous to TF19(6) and containing one zinc finger-type domains and DUF702 domain. The function of this protein is unknown.

20

Example 8: Insect Bioassays

Insect bioassays are performed under controlled conditions in the greenhouse.

Plants are modified using the methods of the present invention. For example: *Solanum lycopersicum* is modified by means of *Agrobacterium*-mediated transformation with pBIN

5 35S-19(6) (encoding the protein having the amino acid sequence of SEQ ID NO:2).

Alternatively, *Solanum lycopersicum* mutants are identified within a mutagenized population so that the mutants carry a mutation or mutations in a nucleic acid sequence of SEQ ID NO:

1 or fragments thereof encoding the transcription factor of the present invention. Insect pest

resistance of modified plants is compared to that of non-modified control plants in choice-

10 and no-choice tests as described in Bleeker et al., 2011 Phytochemistry 72: 8-73; and patent application WO 2010/099,985. Resistance to the following insect classes is determined:

Lepidoptera; Coleoptera; Diptera; Hemiptera; Acari; Thysanoptera.

Insect preference test. A choice test is performed in which insects at different life stages,

e.g., larvae or adults, are allowed to choose between plants that produce terpenes such as

15 linalool or nerolidol (through expression of a nucleic acid sequence encoding the amino acid sequence of SEQ ID NO:2) and control plants. The test determines the repellent activity of the terpenes produced because of the activating effect of the protein of SEQ ID NO:2

Insect performance test (no-choice test). A no-choice test is performed to determine the toxic effects of the terpenes produced by activating expression of a nucleic acid sequence

20 encoding the amino acid sequence of SEQ ID NO:2. In these experiments, insect pest species are forced to eat from (transgenic) plants that have modified terpenes production through expression of a nucleic acid sequence encoding the amino acid sequence of SEQ ID NO:2 and control (or empty vector) plants. Subsequently, insect performance, e.g., growth, development or fitness, is determined as a measure of toxicity.

25

CLAIMS

1. An isolated, synthetic or recombinant nucleic acid sequence selected from the group comprising:
 - 5 a) a nucleic acid sequence of SEQ ID NO: 1;
 - b) a nucleic acid sequence that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 2 or an amino acid sequence that is at least 60% identical to the amino acid sequence of SEQ ID NO:2;
 - c) a nucleic acid sequence that is at least 60% identical to the nucleic acid sequences of (a)
 - 10 or (b), and encodes a transcription factor that regulates terpene biosynthesis;
 - d) a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence which has been derived, by way of one or more amino acid substitutions, deletions or insertions, from the polypeptide having the amino acid sequence of SEQ ID NO:2 and which polypeptide is functionally equivalent to the polypeptide consisting of the amino acid
 - 15 sequence of SEQ ID NO: 2; and
 - e) a nucleic acid sequence that hybridizes under stringent conditions to the nucleic acid sequences of (a), (b), (c), or (d), or
 - f) a nucleic acid sequence that hybridizes under stringent conditions to the complement of the nucleic acid sequences of (a), (b), (c), or (d) and encodes a polypeptide capable of
 - 20 regulating terpene biosynthesis in a plant.
2. A chimeric gene comprising a nucleic acid sequence according to claim 1.
3. A vector comprising a nucleic acid sequence according to claim 1 or a chimeric gene
- 25 according to claim 2.
4. A host cell comprising a chimeric gene according to claim 2 or a vector according to claim 3.
- 30 5. A polypeptide with DNA binding activity that regulates terpene biosynthesis in a plant having an amino acid sequence selected from the group of:
 - (a) an amino acid sequence of SEQ ID NO: 2;
 - (b) an amino acid sequence which has been derived, by way of one or more amino acid substitutions, deletions or insertions, from the polypeptide having the amino acid sequence
 - 35 of SEQ ID NO:2 and which polypeptide is functionally equivalent to the polypeptide consisting of the amino acid sequence of SEQ ID NO: 2; and

(c) an amino acid sequence that is at least 60% identical to the amino acid sequence according to (a).

6. The polypeptide according to claim 5 capable of binding a nucleic acid sequence of a promoter that is operably linked to at least one gene involved in terpene biosynthesis in the plant.

7. The polypeptide according to claim 6 wherein the gene is selected from the group comprising a Terpene Synthase 5 (TPS5) and a Terpene Synthase 11 (TPS11).

8. The polypeptide according to claim 6 or 7 wherein the promoter comprises a trichome-specific promoter.

9. A method for increasing production of at least one terpene in a plant, comprising modifying the plant to have an increased copy number of a nucleic acid sequence of SEQ ID NO:1, or a nucleic acid sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 1, compared to a non-modified plant of the same genetic background, thereby increasing the production of the at least one terpene in the modified plant

10. A method for increasing production of at least one terpene in a plant comprising:
(a) contacting a plant cell with a composition comprising a vector comprising a nucleic acid sequence having at least 60% identity to SEQ ID NO:1 or a functional fragment thereof;

(b) selecting the plant cell transformed with the vector wherein the plant cell overexpresses the nucleic acid sequence or the fragment wherein overexpression results in an increased level of the at least one terpene in the cell compared to a non-transformed plant cell; and

(c) regenerating the plant from the transformed cell of (b) wherein the plant has the increased level of the at least one terpene compared to a non-transformed plant of the same genetic background.

11. A method for increasing production of at least one terpene in a population of plants by selectively breeding the plant of claim 8 to produce the population having the increased level of the at least one terpene.

12. A method for reducing production of at least one terpene in a plant, comprising the step of impairing expression of functional TF19(6) protein in a plant via genetic

modification, and/or the step of reducing binding of functional TF19(6) protein to the promoter sequence of the nucleic acid sequence coding for Terpene Synthase 5 (TPS5) and/or Terpene Synthase 11 (TPS11).

5 13. A method according to claim 12, said method comprising the step of modifying the plant to have at least one mutation in a nucleic acid sequence of SEQ ID NO: 1 or a sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 1, or a nucleic acid sequence of SEQ ID NO: 3 or a sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 3, wherein the mutation results in a decrease of
10 expression of the amino acid sequence of SEQ ID NO:2 or a variant thereof having at least 60% sequence identity to the amino acid sequence of SEQ ID NO:2, and/or a loss of function of the amino acid sequence of SEQ ID NO:2 or a variant thereof having at least 60% sequence identity to the amino acid sequence of SEQ ID NO:2 compared to a non-modified plant of the same genetic background, thereby reducing the level of the at least
15 one terpene in the modified plant.

14. The method according to claim 13 wherein the mutation comprises a substitution, a deletion, an insertion or an addition of at least one nucleotide.

20 15. The method according to claim 12 or 13 wherein the step of impairing expression of functional TF19(6) protein in a plant via genetic modification comprises modifying the plant to have an increased level of RNA having a nucleic acid sequence at least in portion complementary to a nucleic acid sequence of SEQ ID NO: 1, or a nucleic acid sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 1, or a nucleic acid
25 sequence of SEQ ID NO: 3 or a nucleic acid sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 3, compared to a non-modified plant of the same genetic background, thereby decreasing production of the at least one terpene in the modified plant.

30 16. A method for reducing terpene levels in a population of plants, comprising:
(a) providing at least one plants within a population of plants comprising a mutation in a gene having a nucleic acid sequence of SEQ ID NO: 1 or a sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO: 1, or a nucleic acid sequence of SEQ ID NO: 3 or a sequence with at least 60% identity to the nucleic acid sequence of SEQ ID NO:
35 3; and
(b) selectively breeding the at least one mutant plant of (a) to produce the population of plants having the reduced terpene levels.

17. The method according to any of claims 9-16 wherein the terpene comprises at least one of a monoterpene and a sesquiterpene that repel insects.

5 18. The method according to claim 17 wherein the monoterpene comprises at least one compound selected from the group of: linalool, β -myrcene, para-cymene, γ -terpinene, α -terpinene, and α -phellandrene.

10 19. The method according to claim 17 wherein the sesquiterpene comprises at least one compound selected from the group of: neralidol, germacrene, R-curcumine, S-curcumine and 7-epizingiberene.

20. The method according to any of claims 9-16 wherein the terpene comprises at least one of a monoterpene and a sesquiterpene that attract insects.

15

21. The method according to claim 20 wherein the monoterpene comprises at least one compound selected from the group of: β -phellandrene, limonene and 2-carene.

20 22. The method according to claim 20 wherein the sesquiterpene comprises at least β -caryophyllene.

23. The method according to any of claims 17-22 wherein the insects comprise sap-sucking insects and blood-sucking insects.

25 24. The method according to claim 23 wherein the sap-sucking insects comprise psyllids, whiteflies, aphids, mealybugs, plant hoppers and scale insects.

25. The method according to claim 24, wherein the sap-sucking insects further comprise thrips, mites and leaf hoppers.

30

26. The method according to claim 23 wherein the blood sucking insects comprise mosquito, ticks and midges.

35 27. The method according to any of claims 9-26 wherein the plant is at least one crop plant selected from the group of: tomato, pepper, eggplant, lettuce, oilseed rape, broccoli, cauliflower, cabbage crops, cucumber, melon, pumpkin, squash, peanut, soybeans, corn, cotton, beans, cassava, potatoes, sweet potatoes and okra.

28. The method according to claim 27 wherein the plant comprises at least one plant selected from a Solanaceae family.

5 29. The method according to any of claims 9-26 wherein the plant is at least one ornamental plant selected from the group of: hibiscus, poinsettia, lilies, iris, rose and petunia.

30. A plant obtainable by a method according to any of claims 9-29.

10

31. A plant comprising a chimeric gene according to claim 2.

32. The plant according to any one of claims 30 or 31 wherein the plant belongs to the Solanaceae family.

15

33. A tissue culture obtained from the plant of any of claims 30-32 wherein the culture has enhanced production or secretion of at least one terpene, terpene isomer, or terpene derivative.

20 34. A method for producing a terpene, terpene isomer, or terpene analog comprising isolating the terpene, terpene isomer or terpene analog from the tissue culture of claim 33.

35. A method for a marker-assisted introgression of production of at least one terpene into a plant comprising the steps of:

25 (a) identifying a difference in a gene encoding a polypeptide having an amino acid sequence of SEQ ID NO:2 between a plant from the Solanaceae family and a sexually compatible plant wherein the gene with the difference comprises a molecular marker associated with the presence of the at least one terpene in the plant;

30 (b) making a cross between the plant having the molecular marker and the sexually compatible plant of (a);

(c) screening a progeny resulting from the cross for the presence of the molecular marker;

(d) identifying the plant within the progeny having the molecular marker and thereby identifying the plant producing the at least one terpene.

35

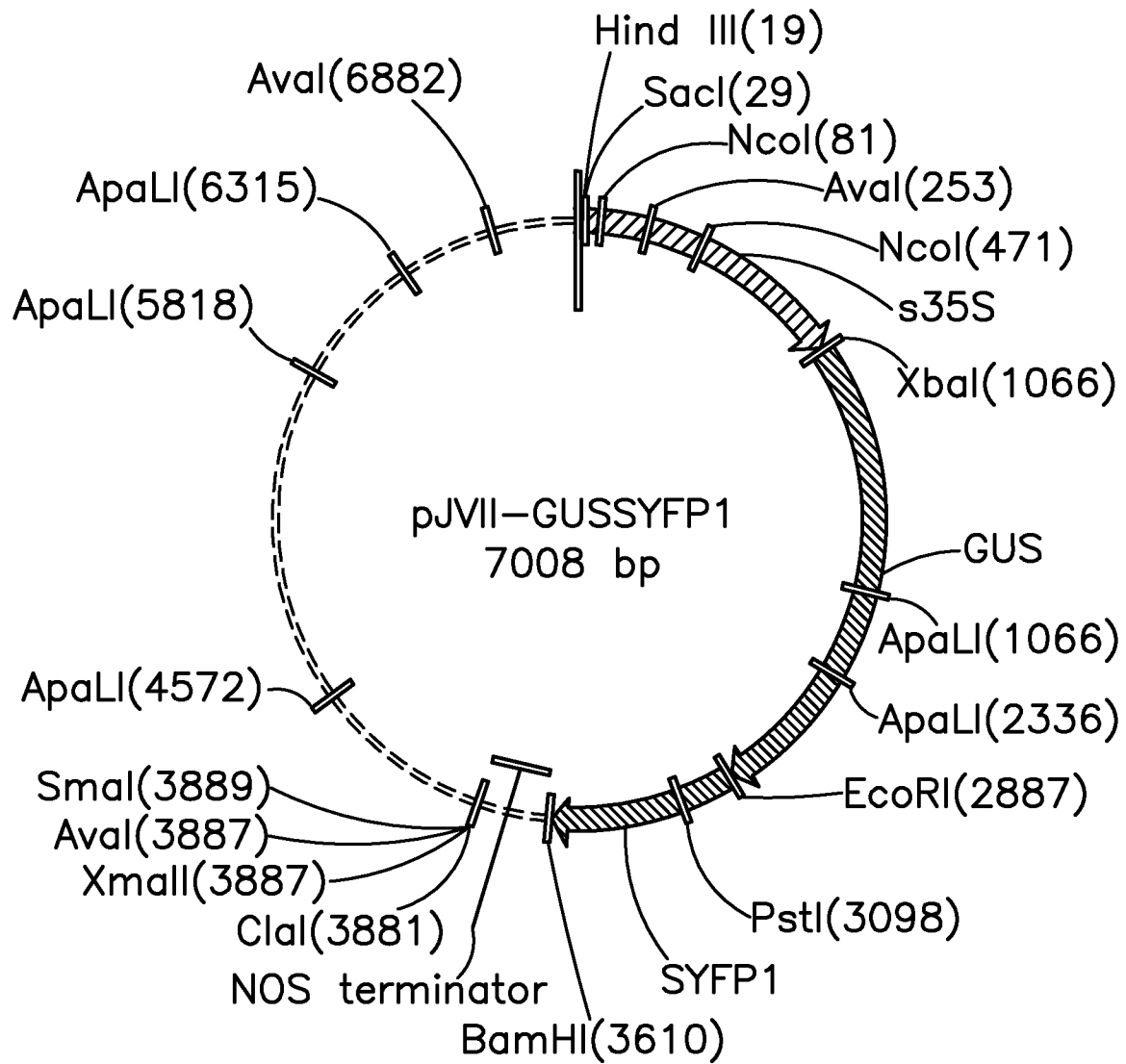


Fig 1

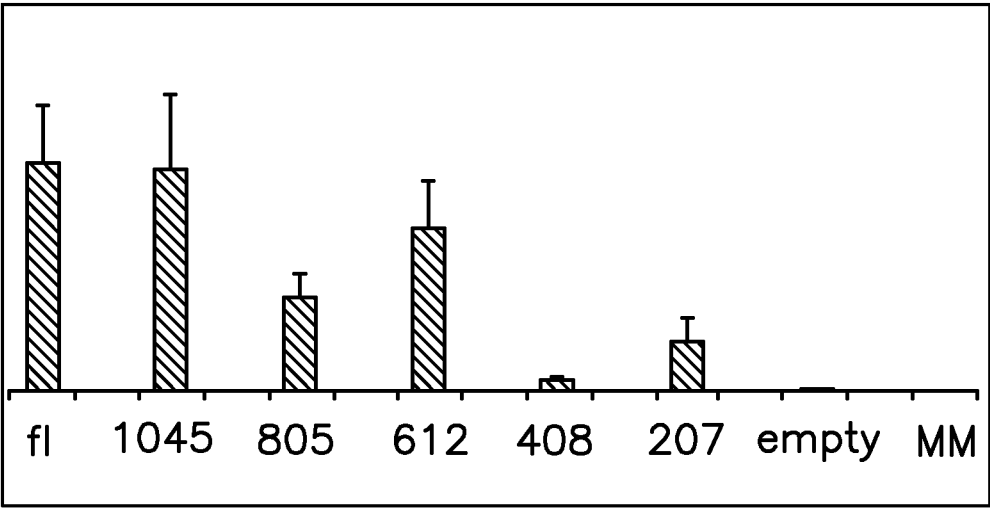


Fig 2

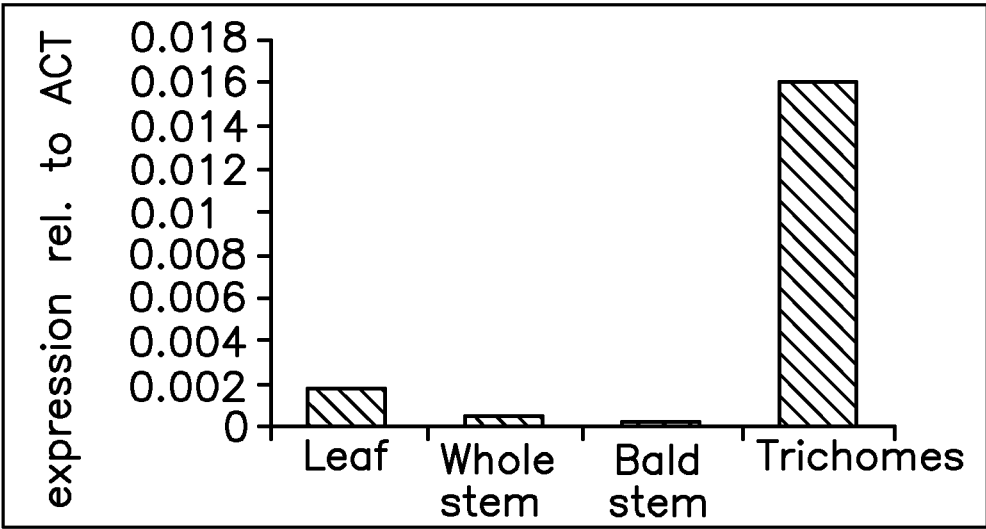


Fig 3 A

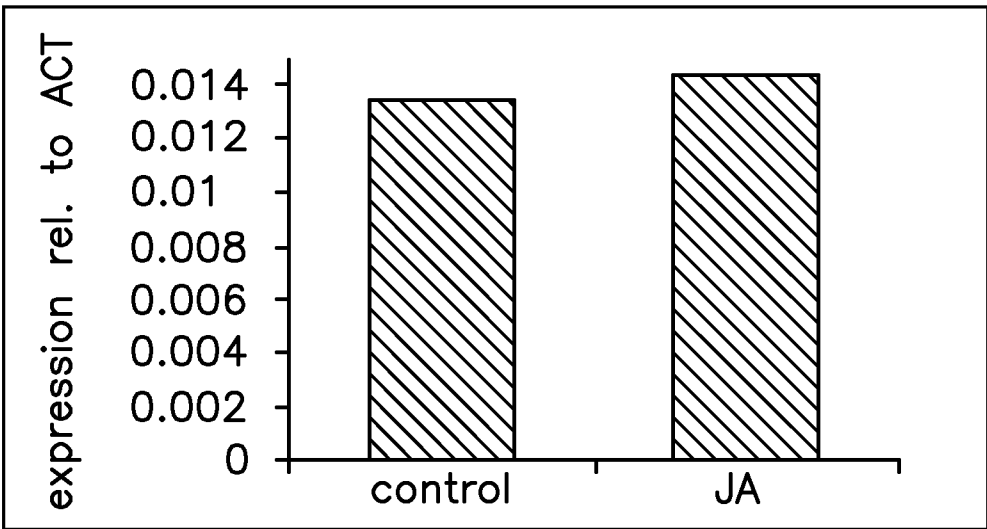


Fig 3 B

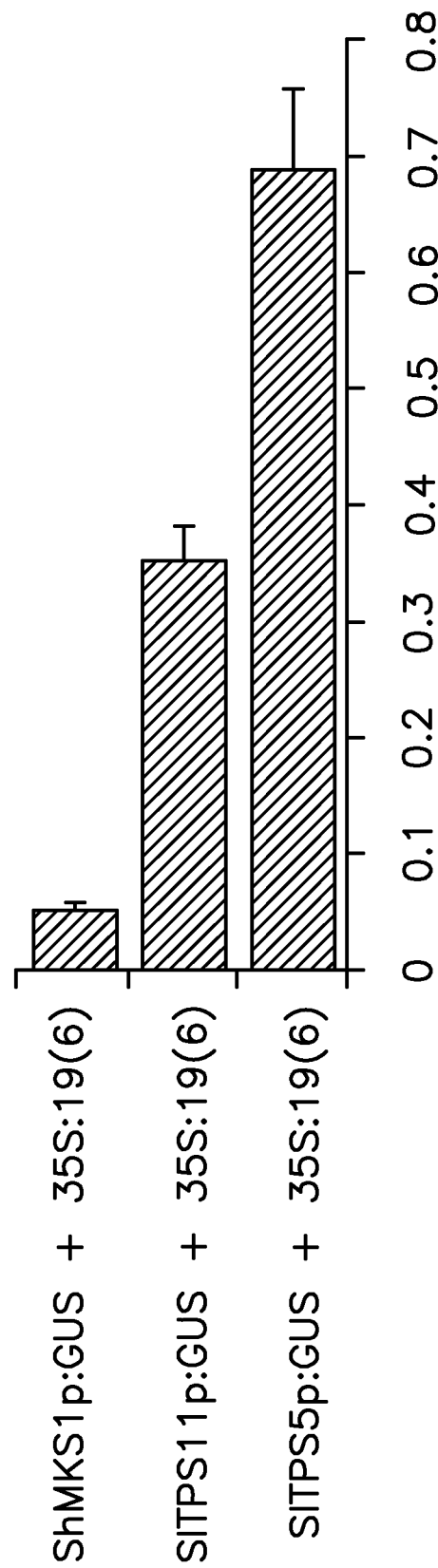


Fig 4

Figure 5

GCAAGTCAAATTGTATATGCTCTTAATTAGGGTTTATGCACCTAGAAAGTAGTA
TTTTAATCGCATATTAGTTATCGATGTTCTTAAATTAATCGCCTATTAGTTATC
GATGTTCTTAAATATAATTGACCTAATTACAAAATTAAGATAGAACTGATTAT
ATTTTTTCAATTTTATCCTTACAAGGAGCAATTCTTTTTTGAAAGTATGAACCA
CTTTGTAAAGTTTTTTTTTTAAAAAAATCTTAAAGGAGTAAATCAGTAAAACT
ACCTTTTCATATTTTATGATTTTTTTAAGAAGCATGTAAAGAAAAAATAGAAAATC
AATATAGAACGAAAAAAGAATTTTATAAAACCTCATAACTTAATAAAAAGAA
TCATATTTATAAGAAATATTTTCTTCCCACATGGAATAATAAATTGCACAAAC
TGTAATATTCTCTACTACAATATAATGTTATAACACACGTATACCGTTGGTT
TTTCAGTATAAATATAATATCCATATTTTAGATATATTAGCTGTAAAAACAA
TATAATATGGATGGAGACAAGTTAAATGTATGTAATTTTACCTTGTAATCCCA
ATTCTCAATATATATATATATATATATATATATATATATATATTCTCCAAGACAAA
ACATTGGATTTTTATTCTAGGAACCTGAATTAAGAATTCAATTTACTCGTAAA
AATTA AAAAGAAATTTCTTATGATCTTATCAAATATTTAATAGGTGAATAGTT
AAATTTGACAAGTTAAATTAAGATGTATGTCCATCACCTCATCATAATTCAAA
TTATTTTAACAAATATCCTTAATCATGATCTTCTTTCTTTTTAGGTTGAAATAA
TTATCATTAGATTTGTACATAGTATAGGAATATATTAAGAGCTAATATATCTA
AAATGTGGATAAATAAAACAATTCTTGCTCAAAATTTTAAATAGTTTAAATAC
TTTTACAATACTTGACACGCGGCATTATATAGCCAACAATTTTACGGGCTAAG
ACATAACATTTATCTTGGAATTTCTCTATTATTAATCATTAGCTTAGATTGTCT
GAGTTTTTTGAAGGTCTTTTATTTAGTTATAGGCAATTTTACCTAGTTTTATAGA
ATTAAAATTATTGTCCGTTGTTATATTTAGGTAAGAAAAAAAGTTAATAAATC
AGACAAGAAAATATAAAGAACCGAAATAATTATGTAATGCCTAAAATAGTTG
GTTTTATATACATAAAGACTGTTGAAAATTGAAATTAATATTGCGGCTCCTTC
ATTTATTGGTATTACTGTTAATTACGTGATTGAAAGGAAAAAAAAAAGTTTTAC
CAAAAAAAAAAGTATAAAAATAAGTTTTGTACTTTATGTCAACAGTTAGTCATC
AATAGTTACTGCTATAATACTAGGTGCCAACACTATGTATAATTCGAATGTGA
TAATAATTTCTGGAAAAAAAAAATTAAAGGATATTTGATTTGATATGGTCCTAG
ATAATGTAGATGATGAAGGGGTGTTAATTAGTCGTTTTCAAATTGATAGGTTAT

TTTGAAAAATTGTGTCATATTAATTGTTTATATTTTCAATGATGTATGATTAAA
ATTAAAATTTTTGAATCTGTCTTAATCGTTTTGGTTTCTCTTTGATTTAGGTAT
AATTCAAATTGATGGTTAATTTTTTTTAAACGTCATCTAACAACATAAAAATTT
GATAAAAAATATTTAAAATTTACAATAACATAAAAATGAAAATATGTTTTCCA
ACTATACCAATTTAGGAGGAGAAACATAGTTATTGTTTTACTATTATCGCTAG
TATTATGAATGAGATATGAAAATTTATATTAATTTATATTGGAATCTATAATT
GATTTTATTAAAAAAAATTAAGTGCGTGTACTTTGACATTTTTTTTTGTTTTTAA
CTCGGCATTCAAAGTTCATATTGAAGTTTTAACTAAATTCGAATCGCACTCTT
CAGAGCAATGCAGGGATGGGTCTCCCAACAACATTTTGTGCGATAGTCTATAC
CCAGAGCTCGAACTTAAGACCTCTGATTAAGAATAAAAATACTTCATTTATAA
ACTGATCATCTTAATATTTTTCAAAATTTAAATGTCACATATTTTCTAAGATATC
CTCGAAACATAATAATAAGTTGAAATGTATAATGTTTGATTGAGACTAAACT
GAGGCGTTTATATATACAATCGTAGAATTAATAATTTAATTGCCATCTGAAA
ATTAATTTAAATATTTATCTATGTACTATACCTTAATTAATTCTTTCATGACAA
ACTTTCTTGGACATTTTTTTCATGAAAAATGCATATAACTTAAACAAGGCCGAT
ACCTTACACCCAAATTGGACAGTATATTTAGAAGAGGGGGAATAATGGTAAA
GAGGGCCGGTATCAGGTTTACACAGAGATGAAAAGTTAGGTGGAGTTTATTT
GTTCCGATGGATTTATCAGTTTTTTTCGTAGATTTTATATTTATATTAGATTCTT
TTTTTTACGTATATATTAATTATAATCCCTAAACAAATTGATTTAGAATCTC
AAACTCATAATCTTAACTTCGCCCCCTAACTTTTATATATATATATATATATA
ATATTTTTAATATATCATTAGTTACACATTATTTTTTTATATTATGTTGTGTATT
ACTGATGAATAATGATTTATGGAAATACAAAAAGCTCTTATTCAGTAATACA
TACATTAGTATGATCATCTTTTTTTCACATTCTTCCATCGCGATATATGTTTTT
TTTTTAAACTATAACACAATAAACACTGCATTAAAAATAATTGTACATATTTT
TTGTGTCTCAATTTATGTGATACCTTTTAGTTTTTTAAGAGCTAAACAATTTAA
ATTTGAGCGAGAATTTACGCATGAAATTTTCGAAAATTCTAAAAAGAAATTT
ATATATTAATAAAAACTACGTAAAAATACTATAAGACACAATAATTGACAAT
TCAAAATATTTAAAAGTCAAAGATATACTTATTTGAATTTCAAAATCTGAAAA
GTATCACATAAATAGGAGGAGAGAGTAACGAATATCAATATTAATGATATAT
TATATTCACCACTAATATTCTTAAAAATAAATATTAAAAAACACCATTAAATTC
GATGTGAATTATTAGTTTGATCCCTGAACTATTGACAGTATTATAAACACTCC

TCTACTGGGTTAGATGAACTTAAATACACACTCGATCTTGTCACAATGATGAA
ATACACCCTAATGAAAATCATATTTACTCTTCCTATTTCTAACCATCGGAAAG
AGTCATCGTGGCTAGGAACTATACTAGCGACCTACCCAATTCATTATAGAA
ATTTTCGCGATCAATGATTGAAAATTTAGAATGTTTCCAACACTTTATCTGTC
AACTTTTTATTAAGAGTTTCAAGCTCGTATAAGAATTTGAAATCACTTTTAGT
ATATCATGTAGTAGATCTAAATATATTTAAAATTATTATAAATTTTTTTTAAA
AACTAATAATTCACACTAAATTGACAAATATCTTCAATACTTAGCTTCTCAC
TTATTTTATACGACCTACCAAACAATCGCGAAACTTTTTAAGTTACTGCAAAC
TGTAGCGGTAAAGAGAGGGGAGGGGGGGGGGGTAGTTGTGGTGCTTTTTAGC
GTTGGCGGCGTTTGCAGAGCTGTAATATATATAATATACCTTTTCTATTAATG
TACCCTCACTCACTCACTTCCTCTCCATAATTCTTTATACAAACAATCATTTTT
TCTTAAACTTGCTCTATTATAAATTCACATTTTTTCTTTATATATACACATACA
TATAGAGCAAAAAAGAAGTTCTAATTTTGTAACCCTTCAAAAAAAGAAAA
ATAATTTTTTTTGAGATCATAAATGAAGAAATCCAAGGGATACAAACATCAT
ATTTGTGTTATAAGTTGGTGCACTTTTGTGGTATGGATTGTGATTAATCACTA
ATCATAATCAAGATTAACAACAAGTAATGGCTAATTTCTTTTCATTAGGTGGG
AATCAAGAACAACAACATCAAGAAATTAGCAGCAGCCAAGCATTAGTACCC
ACAGAGAGTAATAATTGGTTTTGTACAGAAATGAACATCATCATCATCATC
ATAATCAAGAAATACCCAACACTTACAAAGGTTTTGAGTTATGGCAAAGTGG
TAACACTCCACAACACCAACACCAACACCACCAACAACAACAGTTTCGT
CATCCGATTTATCCTTTGCAAGATCTTTATCCACTGATGTTGGATTAGGGGTT
GGGCCAAGCAGAAGTGGCTTTGATATATCTGCAGgtgatcatcagaaacagatttagaatttaa
ctttatctattcagtactttctaagtaacttatagatctataatttaagtttgataaatttaatttatgttctaacaattcacaacattttgc
aattagggaatttcgaaacgtatttactgaagcatgttagaattcccagcctcgaaaaagcatgggaatttggctctatggacttggga
aattctccattcatgagctaacttttgaggtaaaattaggttcatatgtcatatctttacatgatatcagagtaagattcatctcaattcttt
gttcaccaatattggcccccatattattgtgtccacaatctagttaacctacgctggccccctccatattacagtggtccacgttctagtt
aacgagatctgggcttcagaaagagtgtaaagaattcagaaaaaggatgagttttgtctcctgtataaacttgagcaatccttc
cttcagtagctagctagttttggaattaggttagactagatgcatacttttaattttatgttctcactgtagaaccatatagcaacga
aactatagtagtattttgtgaccgctctctctatataatcgtgcataattaagttaattgaatctgttgctaaaaggcgggatgggga
ttattattgtgcagGTGATCATG AGGCGTCGAG GTCGGGATTC
GTGATGATGAGGAGTGGTGGAGGAGGAATAAGTTGCCAAGATTGTGGGAAC

CAAGCTAAGAAAGATTGTCAACATATGAGGTGTAGGACTTGTTGTAAGAGTA
GAGGGTTTCAGTGTCAAACATCATGTGAAAAGTACTTGGGTTCCAGCAGCTAA
AAGGAGAGAAAGGCAACAACAACCTTGCTGCTTTGCAACAACAACAAGG
ACATAATAATAATAATAATCATAAGAATAAAAGGCAAAGGGAGGATCC
AAGTGCTTCTTCTCTTGTGTCTACTCGTTTGCCTTCAAACACTAATGgtaaagtacttc
atgttttcttacctttcattgctacgtctgtttaatttaaaggctcttagttgactgaacatgaatataagatgttgaaattgaaaaacgt
agataaatatttaaattgaaacgagggaataatattaattttttgtatcacacaaagacatagagtcttgagatccatcatgtaaaga
agattaattgatcattgcctaaatgaattctataaaaagtaagctatagagaaaagagaccctatagtaaattcgctcagctttttcttt
tctatttgcattctcttctccatcatcactcttcttttttactctacaaaagattgacaaaaattcgaatgagatatattcaaattttg
agttaattatgaatttttaattctagttaatagaaagtgtgaataaattatttatgtattactaacaataagcaaaaactaaaactttact
tgtacccftgcgctgtgtatgcacaatttcttctcttagacctacacatgatatttatctcgaccctaaaaagatcaccattattcttaa
ttcaatttctgcaatttttttaagataaactattatttgagtaataatataatgtgacttacccaaaaaactgttagtggaagtgaat
ttgagaaaccaactctctaattcatgtataataattgtgttatcatatattgtcattagattggaattaacttatatatctattagtaaagt
tactttgaaataataactattatttgagtaataatataatgttgcttaccaaaaaataactattagttgagtggtctattaactctccaaat
atgtataataattgtgttatcatttcttagtattggaattaacttatatatagtaaagcacttgcaatttcaaattttttacctgctttt
ccttttagtctgattaaaaataaattgactattttcaagcaagtgtttattctaaacttttcagatgaaatgtttaaaaaaaccacaagatt
aaatagtgtttgatacatttgacatatttttagtttagaccataaaattcaaattgctttactaaattcgtgtcaagtatactaggttaa
aaaaaatatttatttgaatacattagtccaaataaacctaattttgtattatggaatttcatgtgtatttttagGGTTAGAAG
TGGGAAAATTTCCATCAAAAGTACGTAAAGTGCTGTATTTCAGTGTATTCAA
TGAGTTCAATTGAGGATGATGAAGATCAATTAGCATATCAAGCTGCTGTGAG
CATTGGTGGACATGTTTTCAAAGGAATTTTATATGATCAAGGTCATGAAAGTC
AGTACAATAACATGGTTGCAGCCGGAGGCGATACGTCTTCCGGTGGTAGTGC
TGGCGGAGTTCAGCACCACCACCATAATTCCGCTGCAGTAGCTACCGCCACC
ACTACAAGTGGTGGCGATGCTACTGCAGCGGGTCCATCGAATTTTCTAGATCC
TTCTTTATTTCCAGCTCCCCTTAGCACTTTTATGGTAGCTGGTACGCAATTTT
TCCACCTTCAAGATCTCCTTGATCGTCCACATTGATAATATTGAGGTGTCTTT
TTAATTTTATGTCAAGAGATTGTTTTTAATTGAAGTATTGATGTTGAATTGA
GTTGTTTACATTAATTCTCTTTGGATTCTACATGAAGTTGTTTTTTTTTCTCTAG
TTCTTATGGTTAATTATTGGTATCATATAGATTGCTTTTTTATTTCACGTTA
AGATGATAATATAAGATAAGATGATAATATACTTAAATGTATATATGTTTTGG
GTTGAGTCTTACGATTACTTATTATTAGAATTTTTTTGTATGTGTATTCGGCTCA

TAATGTGCCAAAAGATAACAAAAGCAAAATTTAAGAGCATTACATAATATT
ATAAGTTTGTGATGGACTGTAAGTATATTTTAGATTTTTTAATTAGAGTTTTTA
AATTTAAACCTAAAAGAAATCGTATTTAAAAAGAGCAGTTTACCCTATAAGT
GATTTTTTTAAGAATAAATATGGATTAGTCGAACCCAATAGTCGGGCAACAG
TTAGAAGCTAAAAAAGATTATAATTTTAAGAAAATACTTACTTTATAAAATTG
AGATATTTGGTTAAGTTTTTAGAGGGGGAAAAGAAATGTGCTTTTGAATAAT
AGCATGAATTAATCTTTACAATTAGAAAAAAAGAAAATTTAAAAATACAAAA
GTAATTGTGAAATTAGGTCAAGCACAACTAAGGTTCTAAAACCTGATTTTAA
AAAAAACTTTTAAATTAATTAATCAACACAAAATTATTACTCTCCAAAAAT
ATTTTCTACATAATACTTATCAAAATAAATATATTTAGAAAATTTGGCCAAAC
TAACATGACTCTTCTTGATTAAGCACATAAATCAAGTTGTTAATAAACTTTG
GCTTTATAGCAATGACTCATTTGCTTTCAAAACATAAAAAAATGAACAAACA
TTAAATATATATTTAACGGAGTAAGATATATTCCAAACTAGGACACTAGAAA
TGGTGAAAGCTTAGTACGTTTGGAACATCAATTCAATTAACTCGAATGTCAC
TGTTTAACTTGTCTTAATATATGTGATAATATTTGATGGATCTTAAATATTATT
TCTTTAAAAAATAATTATTCGTTAGAAGGACAATAAGTGCTACAATGACTTA
AATTTCTAAATTTTCAACTAGGCATAATCCTTCAAAATAACTTTTCATCATACT
TTTGAATAATTAAATATGATATTATTGAAGTTATGTAAATTTTCATGTTTCGG
GCTTGTTTCGGGTTTTTTTAAATATCAAATCAAATTATTCGTGTAAAATTTTTAA
AATTATAAATCAGACCAAATTAATAAAATTCAGATTTTTTTCGGGGTTTTCAAC
TCTGGGTTGATTCGTATTTTTCAAGTACCAAACCAAACCATTGTGTGCAATTT
TTAAATTTTTTAATCAAACCAAATAATAAACTTCGGATTTTTTCCAGATTTTTA
GATTTTTTCGGGTAAAGTTTGCATACAAACATATAATTAACCTTGTGCTCCAAAT
ATTTCTTTAGTCCAACCATAATATAATTATCTAAGGTATTTCTTGAAAAAATT
ACACAAAAGATGAGATGAGTATTGATGACACAAAAATATTCAATAAAAAAT
ACAATAAATCATCATATAAAATAAATATTGTAAAGTCATAATGAAAATAAT
CATAATTTAAAATTTTTTAAATCATGCTAAAATAAGTTTAATAAGTATTAGTTA
CATTATTAAATATTTAAGGAAAACAAAAATTAGATTATGTAAATAAATATAA
AACTAAAGAACAAATATTCAATATTATTGTCATTTTTAGTGTTGAATTGATTT
TTTCTTTTTGCATTAGTATTAATTTAATTTTAATTTAAGCTTTATTATAATTATC
AATCTATGAACATAATCTATATTGGACCATTCCAAATTCTATATTTTAAACT

TGAAACAATATATTTAAAAGTTAAAACTATGAAATAGTATAAGAAATATTTT
AAAATAATATCAACGTAAATATTTTATGTATAAAATAATATTTTACACATATA
ATATAAGGATTTTTTTTCCCGATTTGATTCAATT

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2012/050403

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/415 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)
C07K 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)

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL [Online]</p> <p>19 October 2009 (2009-10-19), "LH_Ea0012N05.r LH_Ea Solanum habrochaetes cDNA, mRNA sequence.", XP002666610, retrieved from EBI accession no. EM_EST:GT170533 Database accession no. GT170533 94% identity to SEQ ID NO:1; abstract; sequence -& MCDOWELL ERIC T ET AL: "Comparative Functional Genomic Analysis of Solanum Glandular Trichome Types", PLANT PHYSIOLOGY (ROCKVILLE), vol. 155, no. 1, January 2011 (2011-01), pages 524-539, XP002666611, ISSN: 0032-0889</p> <p style="text-align: center;">----- -/--</p>	1-35



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

16 August 2012

Date of mailing of the international search report

03/09/2012

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

Griesinger, Irina

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2012/050403

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>US 2009/049566 A1 (ZHANG JAMES Z [US] ET AL) 19 February 2009 (2009-02-19) abstract -& DATABASE Geneseq [Online]</p> <p>9 July 2009 (2009-07-09), "Plant transcription factor sequence, SEQ ID 748.", XP002666612, retrieved from EBI accession no. GSP:AWH30951 Database accession no. AWH30951 sequence -& DATABASE Geneseq [Online]</p> <p>9 July 2009 (2009-07-09), "Transcription factor nucleotide sequence, SEQ ID 7647.", XP002666613, retrieved from EBI accession no. GSN:AWH37850 Database accession no. AWH37850 sequence</p>	1-35
A	<p>-----</p> <p>CHRIS C N VAN SCHIE ET AL: "Tomato linalool synthase is induced in trichomes by jasmonic acid", PLANT MOLECULAR BIOLOGY, KLUWER ACADEMIC PUBLISHERS, DORDRECHT, NL, vol. 64, no. 3, 12 April 2007 (2007-04-12), pages 251-263, XP019507419, ISSN: 1573-5028, DOI: 10.1007/S11103-007-9149-8 cited in the application</p>	1-32
A	<p>-----</p> <p>PIERRE BROUN ET AL: "Importance of transcription factors in the regulation of plant secondary metabolism and their relevance to the control of terpenoid accumulation", PHYTOCHEMISTRY REVIEWS, KLUWER ACADEMIC PUBLISHERS, DO, vol. 5, no. 1, 1 July 2006 (2006-07-01), pages 27-38, XP019405399, ISSN: 1572-980X, DOI: 10.1007/S11101-006-9000-X</p>	1-32
A	<p>-----</p> <p>WO 2009/082208 A2 (KEYGENE NV [NL]; BLEEKER PETRONELLA MARTINA [NL]; VAN SCHIE CHRISTIANU) 2 July 2009 (2009-07-02)</p> <p>-----</p> <p>-/--</p>	1-32

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2012/050403

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>VAN SCHIE C C ET AL: "Regulation of terpenoid and benzenoid production in flowers", CURRENT OPINION IN PLANT BIOLOGY, QUADRANT SUBSCRIPTION SERVICES, GB, vol. 9, no. 2, 1 April 2006 (2006-04-01), pages 203-208, XP028014923, ISSN: 1369-5266 [retrieved on 2006-04-01] page 206, right-hand column, paragraph 1 -----</p>	1-32
A	<p>WO 00/46383 A2 (UNIV LEIDEN [NL]; MEMELINK JOHAN [NL]; FITS CORNELIA THEODORA ELISABE) 10 August 2000 (2000-08-10) -----</p>	1-32
A	<p>GANTET P ET AL: "Transcription factors: tools to engineer the production of pharmacologically active plant metabolites", TRENDS IN PHARMACOLOGICAL SCIENCES, ELSEVIER, HAYWARTH, GB, vol. 23, no. 12, 1 December 2002 (2002-12-01), pages 563-569, XP004394699, ISSN: 0165-6147, DOI: 10.1016/S0165-6147(02)02098-9 page 567 -----</p>	1-32
A	<p>GIGOLASHVILI TAMARA ET AL: "The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana", PLANT JOURNAL, vol. 50, no. 5, June 2007 (2007-06), pages 886-901, XP002666614, ISSN: 0960-7412 cited in the application -----</p>	1-32
A	<p>MISRA PRASHANT ET AL: "Modulation of Transcriptome and Metabolome of Tobacco by Arabidopsis Transcription Factor, AtMYB12, Leads to Insect Resistance", PLANT PHYSIOLOGY (ROCKVILLE), vol. 152, no. 4, April 2010 (2010-04), pages 2258-2268, XP002666615, ISSN: 0032-0889 cited in the application -----</p>	1-32

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/NL2012/050403

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009049566 A1	19-02-2009	NONE	

WO 2009082208 A2	02-07-2009	CN 101952445 A	19-01-2011
		EP 2222859 A2	01-09-2010
		JP 2011507506 A	10-03-2011
		US 2011113512 A1	12-05-2011
		WO 2009082208 A2	02-07-2009

WO 0046383 A2	10-08-2000	AU 779162 B2	06-01-2005
		AU 2580600 A	25-08-2000
		CA 2361678 A1	10-08-2000
		EP 1157119 A1	28-11-2001
		JP 2002535993 A	29-10-2002
		NZ 513356 A	30-01-2004
		WO 0046383 A2	10-08-2000
