International Application Published Under the Patent Co-operation Treaty (PCT)

Title: NOVEL GENES AND USES THEREOF FOR SECONDARY METABOLITE PRODUCTION IN PLANTS

Abstract: The invention provides isolated polypeptides, and homologues thereof, and their use for enhancing secondary metabolite production in plants or plant cells. Polynucleotides encoding such polypeptides and chimeric genes are also provided, as well as vectors, host cells and transgenic plants comprising them. The invention also relates to methods for the production or for stimulating the biosynthesis of secondary metabolites in plant or plant cells.
NOVEL GENES AND USES THEREOF FOR SECONDARY METABOLITE PRODUCTION IN PLANTS

FIELD OF THE INVENTION

The current invention relates to the field of secondary metabolite production in plants and plant cell cultures. More specifically, the present invention relates to the identification of novel genes involved in the biosynthesis and/or production of secondary metabolites in plants.

BACKGROUND

The plant kingdom produces tens of thousands of different small compounds with very complex structures that are often genus or family specific. These molecules, usually called secondary metabolites, display an immense variety in structures and biological activities that plants have tapped into over the course of evolution, and that is now harnessed by man for industrial and medical applications. These compounds play for example a role in the resistance against pests and diseases, attraction of pollinators and interaction with symbiotic microorganisms. Besides the importance for the plant itself, secondary metabolites are of great interest because they determine the quality of food (colour, taste, aroma) and ornamental plants (flower colour, smell). A number of secondary metabolites isolated from plants are commercially available as fine chemicals, for example, drugs, dyes, flavours, fragrances and even pesticides. In addition, various health improving effects and disease preventing activities of secondary metabolites have been discovered, such as anti-oxidative and anti-metastatic-lowering properties. For example taxol is a highly substituted, polyoxygenated cyclic diterpenoid characterised by the taxane ring system, which presents an excellent antitumoral activity against a range of cancers.

The cellular and genetic programs that steer the production of secondary metabolites can be launched rapidly when plants perceive particular environmental stimuli. The jasmonate phytohormones (JAs) play a prominent and universal role in mediating these responses as they can induce synthetic pathways of molecules of a wide structural variety, encompassing all major secondary metabolites (Zhao et al. 2005; Pauwels et al. 2009).

Although about 100,000 plant secondary metabolites are already known, only a small percentage of all plants have been studied to some extent for the presence of secondary metabolites. Interest in such metabolites is growing as e.g. plant sources of new and useful drugs are discovered. Some of these valuable phytochemicals are quite expensive because they are only produced at extremely low levels in plants. In fact, very little is known about the biosynthesis of secondary metabolites in plants. However, some recently elucidated biosynthetic pathways of secondary metabolites are long and complicated
requiring multiple enzymatic steps to produce the desired end product. Most often, the alternative of producing these secondary metabolites through chemical synthesis is complicated due to a large number of asymmetric carbons and in most cases chemical synthesis is not economically feasible.

The recovery of valuable secondary metabolites is mostly achieved through extraction and purification (generally at low yields) of imported, sometimes exotic, plant biomasses, whose reproductive agriculture and secure long term supply are often very difficult, if not impossible to guarantee. Often laborious extraction schemes have to be developed for each specific secondary metabolite of interest. The problems of obtaining useful metabolites from natural sources may potentially be circumvented by cell culture. The culture of plant cells has been explored since the 1960's as a viable alternative for the production of complex phytochemicals of industrial interest. Although plant cell cultures might be somewhat sensitive for shear forces, many cultures can be grown in large bioreactors without difficulty. For example, the use of large-scale plant cell cultures in bioreactors for the production of alkaloid has been extensively studied (Verpoorte et al. 1999). Since it has been observed that undifferentiated cultures such as callus and cell suspension cultures produce only very low levels of secondary metabolites one tends to use differentiated plant cell cultures such as root-and hairy root-culture. For example, tropane alkaloids that are only scarcely synthesized in undifferentiated cells are produced at relatively high levels in cultured roots. Despite the promising features and developments, the production of plant-derived pharmaceutical by plant cell cultures has not been fully commercially exploited. The main reasons for this reluctance shown by industry to produce secondary metabolites by means of cell cultures, compared to the conventional extraction of whole plant material, are economical ones based on the slow growth and the low production levels of secondary metabolites by such plant cell cultures. Important causes are the toxicity of such compounds to the plant cell, and the role of catabolism of the secondary metabolites. Another important problem is that secondary metabolites are mostly retained intracellular^ complicating the downstream processing and purification.

Therefore a need exists for enhancing the production of secondary metabolites in transgenic organisms and cell lines derived thereof.

**SUMMARY OF THE INVENTION**

The present invention provides a regulatory gene encoding a peptide which can be used to induce or enhance the production and/or secretion of secondary metabolites in plants or cells derived thereof. More specifically, the regulatory gene of the present invention codes for a plant peptide that synergistically interacts with elicitors of secondary metabolism.
According to one aspect, the invention relates to an isolated polypeptide selected from the group consisting of:

(a) a polypeptide encoded by a polynucleotide comprising SEQ ID NO: 1;
(b) a polypeptide comprising a polypeptide sequence having a least 75% identity to the polypeptide encoded by a polynucleotide sequence having SEQ ID NO: 1;
(c) a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 2;
(d) a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 3;
(e) a polypeptide comprising an amino acid sequence with at least 75% identity to SEQ ID NO: 2;
(f) a polypeptide comprising an amino acid sequence with at least 75% identity to SEQ ID NO: 3;
(g) fragments and/or variants of the polypeptides according to (a), (b), (c), (d), (e) or (f).

In one embodiment, the invention relates to any of the above described polypeptides wherein said polypeptide sequence is consisting of an amino acid sequence as set forth in SEQ ID NO: 2 or 3 and polypeptide sequences having at least 75% identity to SEQ ID NO: 2 or 3.

According to another aspect, the invention relates to an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising a polynucleotide sequence having the sequence SEQ ID NO: 1;
(b) a polynucleotide comprising a polynucleotide sequence having at least 70% identity to the sequence having SEQ ID NO: 1;
(c) a polynucleotide which encodes the polypeptide sequence as set forth in SEQ ID NO: 2;
(d) a polynucleotide which encodes the polypeptide sequence as set forth in SEQ ID NO: 3;
(e) fragments and variants of the polynucleotides according to (a), (b), (c) or (d).

In another embodiment, the invention relates to a chimeric gene comprising the following operably-linked sequences: a) a promoter region capable of directing expression in a plant or plant cell; b) a DNA region encoding a polypeptide as defined above; c) a 3' polyadenylation and transcript termination region.

A vector comprising a polynucleotide sequence or a chimeric gene as defined above also forms part of the present invention, as well as a host cell comprising a polynucleotide sequence or a chimeric gene or a vector as defined above.

A further aspect of the present invention relates to a transgenic plant or a cell derived thereof that is transformed with a vector or that comprises a chimeric gene as defined above. Said transgenic plant or
cell derived thereof can be from the group comprising *Taxus* spp., *Medicago* spp., *Catharanthus* spp., *Nicotiana* spp., *Artemisia* spp., or any other plant species.

According to the invention, the above described polynucleotide sequences (and encoded proteins) can be used for the enhanced production and/or stimulation of biosynthesis of secondary metabolites in plants or plant cells. Said secondary metabolites comprise alkaloid compounds, phenylpropanoid compounds or terpenoid compounds.

**BRIEF DESCRIPTION OF THE DRAWINGS**

**FIGURE 1** Average linkage hierarchical clustering of MeJA-modulated gene tags identified by cDNA-AFLP of elicited *Taxus baccata* cells. The time points (h) are indicated at the top. Red and green boxes reflect transcriptional activation and repression, respectively, relative to the average expression level in control (DMSO) cells. Gray boxes correspond to missing time points.

**FIGURE 2** Transactivation of pTXS-fLUC and pT130H-fLUC in the presence of MeJA and the regulatory TB595 factor. Transactivation is expressed relative to the normalized LUC activity of the mock-treated control (CON-). Error bars indicate the standard error (n=8).

**FIGURE 3** Transactivation of alkaloid synthesis promoters by MeJA and TB595. Transactivation is expressed relative to the normalized LUC activity of the mock-treated control (CON-). Error bars indicate the standard error (n=8). OB = NtORCl+NbbHLH1

**FIGURE 4** Total alkaloid accumulation in transgenic *Nicotiana tabacum* hairy roots after 8 days of elicitation. Values are expressed as µg/g DW and compare hairy roots transformed with TB595 or control vector. EtOH: control; MeJ: Methyl jasmonate treatment. Alkaloids correspond to: green, Nicotine; yellow, Anabasine; orange, Anatabine. Values are mean ±SE of 3-4 independent samples.

**FIGURE 5** The TB595 peptide sequence is highly conserved in the plant kingdom. Multiple sequence alignment generated with ClustalW of the TB595 homologous sequences of *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Picea sitchensis* (Picea), *Physcomitrella patens* (Pp), *Ricinus communis* (Rc) and *Selaginella moellendorfii* (Sm).

**FIGURE 6** Phylogenetic tree of TB595 orthologs in *A.thaliana* and *M. truncatula*. 
FIGURE 7 Signal peptide prediction with Signal P in the TB595 sequence.

FIGURE 8 The TB595 peptide sequence. The signal peptide sequence is underlined, the mature peptide in black and bold.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will be described with respect to particular embodiments and with reference to certain drawings but the invention is not limited thereto but only by the claims. Any reference signs in the claims shall not be construed as limiting the scope. The drawings described are only schematic and are non-limiting. In the drawings, the size of some of the elements may be exaggerated and not drawn on scale for illustrative purposes. Where the term "comprising" is used in the present description and claims, it does not exclude other elements or steps. Where an indefinite or definite article is used when referring to a singular noun e.g. "a" or "an", "the", this includes a plural of that noun unless something else is specifically stated. Furthermore, the terms first, second, third and the like in the description and in the claims, are used for distinguishing between similar elements and not necessarily for describing a sequential or chronological order. It is to be understood that the terms so used are interchangeable under appropriate circumstances and that the embodiments of the invention described herein are capable of operation in other sequences than described or illustrated herein.

Unless otherwise defined herein, scientific and technical terms and phrases used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Generally, nomenclatures used in connection with, and techniques of molecular and cellular biology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. See, for example, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2002).

Many approaches have been developed to overcome the common problem of low product yield of secondary metabolites in plants or plant cell cultures. One approach is the addition of elicitors. Elicitors are compounds capable of inducing defence responses in plants. Other approaches to increase the product yield of secondary metabolites comprise the screening and selection of high-producing cell lines, the optimisation of the growth and product parameters and the use of metabolic engineering
(Verpoorte et al. 2000). However, metabolic engineering implies detailed knowledge of the biosynthetic steps of the secondary metabolite(s) of interest. The present invention provides a regulatory gene which can be used to modulate the production of secondary metabolites in organisms and cells derived thereof. It was surprisingly found that the regulatory gene of the present invention codes for a small plant peptide hormone, including an N-terminal signal peptide, that synergistically interacts with elicitors of secondary metabolism.

Thus, according to a first aspect, the invention relates to an isolated polypeptide selected from the group consisting of:

(a) a polypeptide encoded by a polynucleotide comprising SEQ ID NO: 1;
(b) a polypeptide comprising a polypeptide sequence having a least 75% identity to the polypeptide encoded by a polynucleotide sequence having SEQ ID NO: 1;
(c) a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 2;
(d) a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 3;
(e) a polypeptide comprising an amino acid sequence with at least 75% identity to SEQ ID NO: 2;
(f) a polypeptide comprising an amino acid sequence with at least 75% identity to SEQ ID NO: 3;
(g) fragments and/or variants of the polypeptides according to (a), (b), (c), (d), (e) or (f).

In a particular embodiment, the invention relates to any of the above isolated polypeptides wherein said polypeptide sequence is consisting of an amino acid sequence as set forth in SEQ ID NO: 2 or 3 and polypeptide sequences having at least 75% identity to SEQ ID NO: 2 or 3.

According to a second aspect, the invention provides for an isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising a polynucleotide sequence having the sequence SEQ ID NO: 1;
(b) a polynucleotide comprising a polynucleotide sequence having at least 70% identity to the sequence having SEQ ID NO: 1;
(c) a polynucleotide which encodes the polypeptide sequence as set forth in SEQ ID NO: 2;
(d) a polynucleotide which encodes the polypeptide sequence as set forth in SEQ ID NO: 3;
(e) fragments and variants of the polynucleotides according to (a), (b), (c) or (d).

As used herein, the terms "polypeptide", "protein", "peptide" are used interchangeably and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.
As used herein, the terms "nucleic acid", "polynucleotide", "polynucleic acid" are used interchangeably and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. Non-limiting examples of polynucleotides include a gene, a gene fragment, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, control regions, isolated RNA of any sequence, nucleic acid probes, and primers. The polynucleotide molecule may be linear or circular. The polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker or the like. The polynucleotide may comprise single stranded or double stranded DNA or RNA. The polynucleotide may comprise modified bases or a modified backbone. A nucleic acid that is up to about 100 nucleotides in length, is often also referred to as an oligonucleotide.

An "isolated polypeptide" or an "isolated polynucleotide", as used herein, refers to respectively an amino acid sequence or a polynucleotide sequence which is not naturally-occurring or no longer occurring in the natural environment wherein it was originally present.

As used herein, the terms "identical" or percent "identity" in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (e.g., 75% identity over a specified region) when compared and aligned for maximum correspondence over a comparison window or designated region as measured using sequence comparison algorithms or by manual alignment and visual inspection. Preferably, the identity exists over a region that is at least about 25 amino acids or nucleotides in length, or more preferably over a region that is 50-100 amino acids or nucleotides or even more in length. According to preferred embodiments, the invention relates to an isolated polypeptide comprising a polypeptide sequence having at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identity to the polypeptide encoded by a polynucleotide sequence having SEQ ID NO: 1 or having at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identity to a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:2 or 3. In other preferred embodiments, the invention relates to an isolated polynucleotide comprising a polynucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identity to the sequence having SEQ ID NO: 1.
In a particular embodiment, fragments and variants of any of the above polynucleotides or polypeptides also from part of the present invention.

In reference to a nucleotide sequence "a fragment" refers to any sequence of at least 15 consecutive nucleotides, preferably at least 30 consecutive nucleotides, more preferably at least 50, 60, 70, 80, 90, 100, 150, 200 consecutive nucleotides or more, of any of the sequences provided herein. If desired, the fragment may be fused at either terminus to additional base pairs, which may number from 1 to 20, typically 50 to 100, but up to 250 to 500 or more.

A "fragment", as referred to polypeptides, refers to a subsequence of the polypeptide. Fragments may vary in size from as few as 5 amino acids to the length of the intact polypeptide, but are preferably at least 10, 15, 20, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75 amino acids in length. If desired, the fragment may be fused at either terminus to additional amino acids, which may number from 1 to 20, typically 50 to 100, but up to 250 to 500 or more. A "functional fragment" means a polypeptide fragment possessing the biological property able to modulate the production of secondary metabolites (as described further herein) in an organism or cell derived thereof. In a particular embodiment said functional fragment is able to modulate the production of a secondary metabolite in a plant or plant cell derived thereof. A fragment may comprise, for example, an activation domain or a domain for protein-protein interactions. In a particular embodiment, a fragment may comprise a receptor binding domain.

A "variant" as used herein refers to homologs, orthologs and paralogs and include, but are not limited to, homologs, orthologs and paralogs of SEQ. ID NOs: 1, 2 or 3. Homologs of a protein encompass peptides, oligopeptides and polypeptides having amino acid substitutions, deletions and/or insertions, preferably by a conservative change, relative to the unmodified protein in question and having similar biological and functional activity as the unmodified protein from which they are derived; or in other words, without significant loss of function or activity. Orthologs and paralogs, which are well-known terms by the skilled person, define subcategories of homologs and encompass evolutionary concepts used to describe the ancestral relationships of genes. Paralogs are genes within the same species that have originated through duplication of an ancestral gene; orthologs are genes from different organisms that have originated through speciation, and are also derived from a common ancestral gene. Several different methods are known by those of skill in the art for identifying and defining these functionally homologues sequences. General methods for identifying orthologues and paralogues include phylogenetic methods, sequence similarity and hybridization methods. Percentage similarity and identity can be determined electronically. Examples of useful algorithms are PILEUP (Higgins & Sharp, CABIOS 5:151 (1989), BLAST and BLAST 2.0 (Altschul et al. J. Mol. Biol. 215: 403 (1990). Software for
performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). Preferably, said homologue, orthologue or paralogue has a sequence identity at protein level of at least 50%, preferably 60%, more preferably 70%, even more preferably 80%, most preferably 90% as measured in a BLASTp. For example, non-limiting examples of functionally homologous sequences of TB595 (SEQ ID NO: 2) include At2g20562, At2g31090 (Arabidopsis thaliana); Os05g0209500, Os01g0214500 (Oryza sativa); ABK22794, ABK23667 (Picea sitchensis); XP_001756653, XP_001754417, XP_001756644 (Physcomitrella patens); XP_002522544, XP_002517654 (Ricinus communis); XP_002967828 (Selaginella moellendorffii); XP_002309758, XP_002328037, XP_002317478, XP_002329114, XP_002330749, XP_002326022, XP_002317432 (Populus trichocarpa); ACU14314, ACU14776, ACU17275 (Glycine max); VvXP_002285182, VvXP_002279581, VvXP_002275821 (Vitis vinifera); SbXP_002457141, SbXP_002439444 (Sorghum bicolor). Other examples of functionally homologous sequences of TB595 (SEQ ID NO: 2) can be found via BLAST analysis in specific publicly available databases such as the online EST database of Medicago truncatula (http://compbio.dfcl.harvard.edu/cgi-bin/tgi/gimain.pl?gdb=medicago) and include TC163920, NP7260335, TC168451, TC158280, TC164853.

In one preferred embodiment, said variants have an increased activity.

Further, it will be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the polypeptides of the present invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides. Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed "silent" variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence are a feature of the invention. In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide (i.e. enhanced secondary metabolite production, in the context of the present invention), these conservative variants are, likewise, a feature of the invention.

Conservative substitutions or variations, as used herein, are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 1. Table 1 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.
Substitutions that are less conservative than those in Table 1 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

Substitutions, deletions and insertions introduced into the sequences are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed

<table>
<thead>
<tr>
<th>Residue</th>
<th>Conservative Substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Ser</td>
</tr>
<tr>
<td>Arg</td>
<td>Lys</td>
</tr>
<tr>
<td>Asn</td>
<td>Gln, His</td>
</tr>
<tr>
<td>Asp</td>
<td>Glu</td>
</tr>
<tr>
<td>Gln</td>
<td>Asn</td>
</tr>
<tr>
<td>Cys</td>
<td>Ser</td>
</tr>
<tr>
<td>Glu</td>
<td>Asp</td>
</tr>
<tr>
<td>Gly</td>
<td>Pro</td>
</tr>
<tr>
<td>His</td>
<td>Asn, Gln</td>
</tr>
<tr>
<td>Ile</td>
<td>Leu, Val</td>
</tr>
<tr>
<td>Leu</td>
<td>Ile, Val</td>
</tr>
<tr>
<td>Lys</td>
<td>Arg, Gln</td>
</tr>
<tr>
<td>Met</td>
<td>Leu, Ile</td>
</tr>
<tr>
<td>Phe</td>
<td>Met, Leu, Tyr</td>
</tr>
<tr>
<td>Ser</td>
<td>Thr, Gly</td>
</tr>
<tr>
<td>Thr</td>
<td>Ser, Val</td>
</tr>
<tr>
<td>Trp</td>
<td>Tyr</td>
</tr>
<tr>
<td>Tyr</td>
<td>Trp, Phe</td>
</tr>
<tr>
<td>Val</td>
<td>Ile, Leu</td>
</tr>
</tbody>
</table>
mutagenesis or the other methods known in the art. Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a sequence. The mutations that are made in the polynucleotides of the invention should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function (i.e. enhanced secondary metabolite production, in the context of the present invention).

In a further embodiment, the invention also relates to a chimeric gene, or otherwise expression cassette, comprising the following operably-linked sequences: a) a promoter region capable of directing expression in a host organism; b) a DNA region encoding a polypeptide according to any of claims 1 or 2; c) a 3' polyadenylation and transcript termination region.

The term "host organism" is understood as being any unicellular or multicellular organism into which the chimeric gene according to the invention can be introduced for the purpose of producing secondary metabolites (as described further herein). These organisms are, in particular, bacteria, for example *E. coli*, yeasts, in particular of the genera *Saccharomyces, Kluyveromyces, Pichia, Yarrowia, Hansenula, Kogamataella*, fungi, in particular *Aspergillus*, a baculovirus, algae, animal cells, or, preferably, plant cells and plants. "Plant cell" is understood, according to the invention, as being any cell which is derived from or found in a plant and which is able to form or is part of undifferentiated tissues, such as calli, differentiated tissues such as embryos, parts of plants, plants or seeds.

The term "operably linked" as used herein refers to a linkage in which the regulatory sequence is contiguous with the gene of interest to control the gene of interest, as well as regulatory sequences that act in trans or at a distance to control the gene of interest. For example, a DNA sequence is operably linked to a promoter when it is ligated to the promoter downstream with respect to the transcription initiation site of the promoter and allows transcription elongation to proceed through the DNA sequence. A DNA for a signal sequence is operably linked to DNA coding for a polypeptide if it is expressed as a pre-protein that participates in the transport of the polypeptide. Linkage of DNA sequences to regulatory sequences is typically accomplished by ligation at suitable restriction sites or adapters or linkers inserted in lieu thereof using restriction endonucleases known to one of skill in the art.
The term "regulatory sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operably linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRMA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

In still another embodiment the invention provides for a vector comprising any of the above described polynucleotide sequences or chimeric gene. Further a host cell comprising any of the above described polynucleotides or chimeric gene or vector are also envisaged herein.

The term "vector" as used herein is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. The vector may be of any suitable type including, but not limited to, a phage, virus, plasmid, phagemid, cosmid, bacmid or even an artificial chromosome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., vectors having an origin of replication which functions in the host cell). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and are thereby replicated along with the host genome. Moreover, certain preferred vectors are capable of directing the expression of certain genes of interest. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). Suitable vectors have regulatory sequences, such as promoters, enhancers, terminator sequences, and the like as desired and according to a particular host organism (e.g. plant cell). Typically, a recombinant vector according to the present invention comprises at least one "chimeric gene" or "expression cassette". Expression cassettes are generally DNA constructs preferably including (5' to 3' in the direction of transcription): a promoter region, a polynucleotide sequence, homologue, variant or fragment thereof of the present invention operably linked with the transcription initiation region, and a termination sequence including a stop signal for RNA polymerase and a polyadenylation signal. It is understood that all of these regions should be capable of operating in biological cells, such as plant cells, to be transformed. The promoter region comprising the transcription initiation region, which preferably includes the RNA polymerase binding
site, and the polyadenylation signal may be native to the biological cell to be transformed or may be derived from an alternative source, where the region is functional in the biological cell.

The term "recombinant host cell" ("expression host cell", "expression host system", "expression system" or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A recombinant host cell may be an isolated cell or cell line grown in culture or may be a cell which resides in a living tissue or organism. Host cells can be of bacterial, fungal, plant or mammalian origin.

According to yet another aspect, the invention provides a transgenic plant or a cell derived thereof that is transformed with the above described vector.

The term "plant" as used herein refers to vascular plants (e.g. gymnosperms and angiosperms). A "transgenic plant" refers to a plant comprising a recombinant polynucleotide and/or a recombinant polypeptide according to the invention. A transgenic plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, and progeny thereof. A transgenic plant can be obtained by transforming a plant cell with an expression cassette of the present invention and regenerating such plant cell into a transgenic plant. Such plants can be propagated vegetatively or reproductively. The transforming step may be carried out by any suitable means, including by Agrobacterium-mediated transformation and non-Agrobacterium-mediated transformation, as discussed in detail below. Plants can be regenerated from the transformed cell (or cells) by techniques known to those skilled in the art. Where chimeric plants are produced by the process, plants in which all cells are transformed may be regenerated from chimeric plants having transformed germ cells, as is known in the art. Methods that can be used to transform plant cells or tissue with expression vectors of the present invention include both Agrobacterium and non-Agrobacterium vectors. Agrobacterium-mediated gene transfer exploits the natural ability of Agrobacterium tumefaciens to transfer DNA into plant chromosomes and is described in detail in Gheysen, G., Angenon, G. and Van Montagu, M. 1998. Agrobacterium-mediated plant transformation: a scientifically intriguing story with significant applications. In K. Lindsey (Ed.), Transgenic Plant Research. Harwood Academic Publishers, Amsterdam, pp. 1-33 and in Stafford, H.A. (2000) Botanical Review 66: 99-118. A second group of transformation methods is the non-Agrobacterium mediated transformation and these methods are known as direct gene transfer methods. An overview is brought

Methods include particle gun delivery, microinjection, electroporation of intact cells, polyethylene glycol-mediated protoplast transformation, electroporation of protoplasts, liposome-mediated transformation, silicon-whiskers mediated transformation etc.

Hairy root cultures can be obtained by transformation with virulent strains of Agrobacterium rhizogenes, and they can produce high contents of secondary metabolites characteristic to the mother plant. Protocols used for establishing of hairy root cultures vary, as well as the susceptibility of plant species to infection by Agrobacterium (Toivounen et al. 1993; Vanhala et al. 1995). It is known that the Agrobacterium strain used for transformation has a great influence on root morphology and the degree of secondary metabolite accumulation in hairy root cultures. It is possible by systematic clone selection e.g. via protoplasts, to find high yielding, stable, and from single cell derived-hairy root clones. This is possible because the hairy root cultures possess a great somaclonal variation. Another possibility of transformation is the use of viral vectors (Turpen 1999).

Any plant tissue or plant cells capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with an expression vector of the present invention. The term 'organogenesis' means a process by which shoots and roots are developed sequentially from meristematic centers; the term 'embryogenesis' means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include protoplasts, leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g. apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyls meristem).

A "control plant" as used in the present invention refers to a plant cell, seed, plant component, plant tissue, plant organ or whole plant used to compare against transgenic or genetically modified plant for the purpose of identifying a difference in production of secondary metabolite (as described further herein) in the transgenic or genetically modified plant. A control plant may in some cases be a transgenic plant line that comprises an empty vector or marker gene, but does not contain the recombinant polynucleotide of the present invention that is expressed in the transgenic or genetically modified plant being evaluated. In general, a control plant is a plant of the same line or variety as the transgenic or genetically modified plant being tested. A suitable control plant would include a
genetically unaltered or non-transgenic plant of the parental line (wild type) used to generate a
transgenic plant herein.

Plants of the present invention may include, but not limited to, plants or plant cells of agronomically
important crops which are or are not intended for animal or human nutrition, such as maize or corn,
wheat, barley, oat, *Brassica* spp. plants such as *Brassica napus* or *Brassica juncea*, soybean, bean,
alfalfa, pea, rice, sugarcane, beetroot, tobacco, sunflower, cotton, *Arabidopsis*, vegetable plants such
as cucumber, leek, carrot, tomato, lettuce, peppers, melon, watermelon, diverse herbs such as
oregano, basilicum and mint. It may also be applied to plants that produce valuable compounds, e.g.
useful as for instance pharmaceuticals, as ajmalicine, vinblastine, vincristine, ajmaline, reserpine,
rescinnamine, camptothecine, ellipticine, quinine, and quinidine, taxol, morphine, scopolamine,
atropine, cocaine, sanguinarine, codeine, genistein, daidzein, digoxin, calystegins or as food additives
such as anthocyanins, vanillin; including but not limited to the classes of compounds mentioned above.
Examples of such plants include, but not limited to, *Papaver* spp., *Rauwolfia* spp., *Taxus* spp., *Cinchona*
spp., *Eschscholtzia californica*, *Camptotheca acuminata*, *Hyoscyamus* spp., *Berberis* spp., *Coptis* spp.,
*Datura* spp., *Atropa* spp., *Thalictrum* spp., *Peganum* spp. Preferred members of the genus *Taxus*
comprise *Taxus brevifolia*, *Taxus baccata*, *Taxus cuspidate*, *Taxus Canadensis* and *Taxus floridana*.

The polynucleotide sequence, homologue, variant or fragment thereof of the invention may be
expressed in for example a plant cell under the control of a promoter that directs constitutive
expression or regulated expression. Regulated expression comprises temporally or spatially regulated
expression and any other form of inducible or repressible expression. Temporally means that the
expression is induced at a certain time point, for instance, when a certain growth rate of the plant cell
culture is obtained (e.g. the promoter is induced only in the stationary phase or at a certain stage of
development). Spatially means that the promoter is only active in specific organs, tissues, or cells (e.g.
only in roots, leaves, epidermis, guard cells or the like). Other examples of regulated expression
comprise promoters whose activity is induced or repressed by adding chemical or physical stimuli to
the plant cell. In a preferred embodiment the expression is under control of environmental, hormonal,
chemical, and/or developmental signals. Such promoters for plant cells include promoters that are
regulated by (1) heat, (2) light, (3) hormones, such as abscisic acid and methyl jasmonate (4) wounding
or (5) chemicals such as salicylic acid, chitosans or metals. Indeed, it is well known that the expression
of secondary metabolites can be boosted by the addition of for example specific chemicals, jasmonate
and elicitors. In a particular embodiment the co-expression of several (more than one) polynucleotide
sequence or homologue or variant or fragment thereof, in combination with the induction of
secondary metabolite synthesis is beneficial for an optimal and enhanced production of secondary
metabolites. Alternatively, the at least one polynucleotide sequence, homologue, variant or fragment
thereof is placed under the control of a constitutive promoter. A constitutive promoter directs
expression in a wide range of cells under a wide range of conditions. Examples of constitutive plant
promoters useful for expressing heterologous polypeptides in plant cells include, but are not limited to,
the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in
most plant tissues including monocots; the nopaline synthase promoter and the octopine synthase
promoter. The expression cassette is usually provided in a DNA or RNA construct which is typically
called an "expression vector" which is any genetic element, e.g., a plasmid, a chromosome, a virus,
behaving either as an autonomous unit of polynucleotide replication within a cell (i.e. capable of
replication under its own control) or being rendered capable of replication by insertion into a host cell
chromosome, having attached to it another polynucleotide segment, so as to bring about the
replication and/or expression of the attached segment. Suitable vectors include, but are not limited to,
plasmids, bacteriophages, cosmids, plant viruses and artificial chromosomes. The expression cassette
may be provided in a DNA construct which also has at least one replication system. In addition to the
replication system, there will frequently be at least one marker present, which may be useful in one or
more hosts, or different markers for individual hosts. The markers may a) code for protection against a
biocide, such as antibiotics, toxins, heavy metals, certain sugars or the like; b) provide complementation,
by imparting prototrophy to an auxotrophic host; or c) provide a visible phenotype through the
production of a novel compound in the plant. Exemplary genes which may be employed
include neomycin phosphotransferase (NPTII), hygromycin phosphotransferase (HPT), chloramphenicol
acetyltransferase (CAT), nitrilase, and the gentamicin resistance gene. For plant host selection, non-
limiting examples of suitable markers are β-glucuronidase, providing indigo production, luciferase,
providing visible light production, Green Fluorescent Protein and variants thereof, NPTII, providing
kanamycin resistance or G418 resistance, HPT, providing hygromycin resistance, and the mutated aroA
gene, providing glyphosate resistance.

The term "promoter activity" refers to the extent of transcription of a polynucleotide sequence,
homologue, variant or fragment thereof that is operably linked to the promoter whose promoter
activity is being measured. The promoter activity may be measured directly by measuring the amount
of RNA transcript produced, for example by Northern blot or indirectly by measuring the product
coded for by the RNA transcript, such as when a reporter gene is linked to the promoter.

According to a further aspect of the invention, the above described polynucleotide sequences (and
encoded proteins) can be used for inducing or enhancing the production of secondary metabolites
and/or for the stimulation of biosynthesis of secondary metabolites in any organism wherein said
organisms can be selected from the list comprising bacteria, for example E. coli, yeasts, in particular of
the genera Saccharomyces, Kluyveromyces, Pichia, Yarrowia, Hansenula, Kogamataella, fungi, in
particular *Aspergillus*, a baculovirus, algae, animal cells or, preferably, plant cells and plants. According to a preferred embodiment, the present invention provides methods for inducing or enhancing the production of secondary metabolites chosen from the group comprising alkaloid compounds, phenylpropanoid compounds and terpenoid compounds, for which non-limiting examples are provided further herein. In one particular embodiment, the present invention provides methods for inducing or enhancing the production of at least one secondary metabolite, which is meant to include related structures of secondary metabolites and intermediates or precursors thereof.

Generally, two basic types of metabolites are synthesised in cells, i.e. those referred to as primary metabolites and those referred to as secondary metabolites. A primary metabolite is any intermediate in, or product of the primary metabolism in cells. The primary metabolism in cells is the sum of metabolic activities that are common to most, if not all, living cells and are necessary for basal growth and maintenance of the cells. Primary metabolism thus includes pathways for generally modifying and synthesising certain carbohydrates, amino acids, fats and nucleic acids, with the compounds involved in the pathways being designated primary metabolites. In contrast hereto, secondary metabolites usually do not appear to participate directly in growth and development. They are a group of chemically very diverse molecules produced by a large variety of organisms, including bacteria, fungi, plants and animals, that often have a restricted taxonomic distribution. Secondary metabolites normally exist as members of closely related chemical families, usually of a molecular weight of less than 1500 Dalton, although some bacterial toxins are considerably longer.

Secondary plant metabolites, as used herein, include e.g. alkaloid compounds (e.g. terpenoid indole alkaloid, tropane alkaloid, steroid alkaloids), phenylpropanoid compounds (e.g. quinines, lignans and flavonoids), terpenoid compounds (e.g. monoterpenoids, iridoids, sesquiterpenoids, diterpenoids and triterpenoids). In addition, secondary metabolites include small molecules, such as substituted heterocyclic compounds which may be monocyclic or polycyclic, fused or bridged. Many plant secondary metabolites have value as pharmaceuticals. Non-limiting examples of plant pharmaceuticals include e.g. taxol, digoxin, atropine, scopolamine, colchicine, diosgenin, codeine, cocaine, morphine, quinine, shikonin, ajmaline, vinblastine, vincristine, and others.

As used herein, the definition of "alkaloids", of which more than 15,000 structures have been described already, refers to all nitrogen-containing natural products which are not otherwise classified as amino acid peptides and proteins, amines, cyanogenic glycosides, glucosinolates, antibiotics, phytohormones or primary metabolites (such as purine and pyrimidine bases). Alkaloids can be divided into the following major groups: (i) "true alkaloids", which contain nitrogen in the heterocycle and originate from amino acids, for which atropine, nicotine and morphine are characteristic examples; (ii)
"protoalkaloids", which contain nitrogen in the side chain and also originate from amino acids, for which mescaline, adrenaline and ephedrine are characteristic examples; (iii) "polyamine alkaloids" which are derivatives of putrescine, spermidine and spermine; (iv) "peptide and cyclopeptide alkaloids"; (v) "pseudoalkaloids", which are alkaloid-like compounds which do not originate from amino acids, including terpene-like and steroid-like alkaloids as well as purine-like alkaloids such as caffeine, theobromine and theophylline.

Plants synthesize alkaloids for various defence-related reactions, e.g. actions against pathogens or herbivores. Enzymes and genes have been partly characterised only in groups of nicotine and tropane alkaloids, indole alkaloids and isoquinoline alkaloids (Suzuki et al. 1999). Nicotine and tropane alkaloid share partly the same biosynthetic pathway. Many plants belonging to, for example, the Solanaceae family, have been used for centuries because of their active substances hyoscyamine and scopolamine. Also other Solanaceae plants belonging to the genera Atropa, Datura, Duboisia and Scopolia produce these valuable alkaloids. In medicine they find important applications in ophthalmology, anaesthesia, and in the treatment of cardiac and gastrointestinal diseases. Although a lot of information is available on the pharmacological effects of tropane alkaloid, surprisingly little is known about how plants synthesize these substances and almost nothing is known about how this synthesis is regulated. Nicotine is found in the genus Nicotiana and also other genera of Solanaceae and is also present in many other plants including lycopods and horsetails (Flores et al. 1991). Although much is known of the alkaloid metabolite content in different organs of tobacco, surprisingly little is known about the biosynthesis, metabolism and regulation of various nicotine alkaloids in tobacco callus and cell cultures.

The "calystegins" constitute a unique subgroup of the tropane alkaloid class (Goldmann et al. 1990). They are characterized by the absence of an N-methyl substituent and a high degree of hydroxylation. Trihydroxylated calystegins are summarized as the calystegin A-group, tetrahydroxylated calystegins as the B-group, and pentahydroxylated derivatives form the C-group. Calystegins represent a novel structural class of tropane alkaloids possessing potent glycosidase inhibitory properties next to longer known classes of the monocyclic pyrrolidines (e.g. dihydroxymethyl dihydroxy pyrrolidine), pyrrolines and piperidines (e.g. deoxyxojirimycin), and the bicyclic pyrrolizidines (e.g. australine) and indolizidines (e.g. swainsonine and castanospermine). Glycosidase inhibitors are potentially useful as antidiabetic, antiviral, antimitastatic, and immunomodulatory agents.

As used herein, "phenylpropanoids" or "phenylpropanes", refer to aromatic compounds with a propyl side-chain attached to the aromatic ring, which can be derived directly from phenylalanine. The ring often carries oxygenated substituents (hydroxyl, methoxy and methylenedioxy groups) in the para-
position. Natural products in which the side-chain has been shortened or removed can also be derived from typical phenylpropanes. Phenylpropanoids are found throughout the plant kingdom. Most plant phenolics are derived from the phenylpropanoid and phenylpropanoid-acetate pathways and fulfil a very broad range of physiological roles in plants. For example, polymeric lignins reinforce specialized cell wall. Closely related are the lignans which vary from dimers to higher oligomers. Lignans can either help defend against various pathogens or act as antioxidants in flowers, leaves and roots. The flavonoids comprise an astonishingly diverse group of more than 4500 known compounds. Among their subclasses are the anthocyanins (pigments), proanthocyanidins or condensed tannins (feeding deterrents and wood protectants), and isoflavonoids (defensive products and signaling molecules). The coumarins, furanocoumarins, and stilbenoids protect against bacterial and fungal pathogens, discourage herbivory, and inhibit seed germination.

As used herein, "terpenoids" or otherwise "isoprenoids" refer to the large and diverse class of naturally-occurring organic chemicals similar to terpenes and can be found in all classes of living organisms. Terpenoids are molecules derived from a five-carbon isoprene unit that are assembled and modified in different ways and have diverse activities. Plant terpenoids are used extensively for their aromatic qualities and contribute to e.g. the scent of eucalyptus, the flavors of cinnamon, clover and ginger, and the color of yellow flowers. They play a role in traditional herbal remedies and may have antibacterial, antineoplastic, and other pharmaceutical functions. Well-known terpenoids include citral, menthol, camphor, salvinorin A and cannabinoids and are also used to flavour and/or scent a variety of commercial products. The steroids and sterols in animals are biologically produced from terpenoid precursors. They also include pharmaceuticals e.g. taxol, artemisinin.

Terpenoids are classified with reference to the number of isoprene units that comprise the particular terpenoid. For example a monoterpenoid comprises two isoprene units; a sesquiterpenoid comprises three isoprene units and a diterpenoid four isoprene units. Polyterpenoids comprise multiple isoprene units. There are many thousands of examples of terpenoids. Artemisinin is a sesquiterpene lactone endoperoxide and is a natural product produced by the plant Artemisia annua. Artemisinin is typically used in combination with anti-malarial therapeutics, for example lumefantrine, mefloquine, amodiaquine, sulfadoxine, chloroquine, in artemisinin combination therapies (ACT). Endoperoxides like artemisinin, for example dihydroartemisinin, artemether, and sodium artesunate have been used in the treatment of malaria. Examples of monoterpenoids include linalool, citronellol, menthol, geraniol and terpineol. Linalool and citronellol are used as a scent in soap, detergents, shampoo and lotions. Linalool is also an intermediate in the synthesis of vitamin E. Menthol is isolated from peppermint or other mint oils and is known for its anaesthetic properties; it is often included sore throat medications and oral medications e.g. for the treatment of bad breath in toothpaste and mouth wash. Geraniol is
known for its insect repellent properties and and is also used as a scent in perfumes. Terpineol is also 
used as an ingredient in perfumes and cosmetics and as flavouring. It is apparent that in addition to the 
pharmaceutical applications of monoterpenoids such as perillyl alcohol there are additional uses as 
scents, flavourings and as insect deterrents.

The synthesis of terpenoids involves a large number of enzymes with different activities. For example 
isoprene units are synthesized from monosaturated isoprene units by prenyltransferases into multiples 
of 2, 3 or 4 isoprene units. These molecules serve as substrates for terpene synthase enzymes, also 
called terpene cyclase. Plant terpene synthases are known in the art.

Taxanes are diterpenoids which were originally described in the genes Taxus (yews). The term "taxoids"
is equivalent with the word "taxanes". As an example, taxol is a highly substituted, polyoxygenated 
cyclic diterpenoid characterised by the taxane ring system, which presents an excellent antitumoral 
activity against a range of cancers. Paclitaxel is the generic name for taxol, which is a registered 
trademark of Bristol-Myers Squibb. The taxol biosynthetic pathway is considered to require 19 
enzymatic steps from the universal diterpenoid precursor geranylgeranyl diphosphate which is 
cyclised, in the committed step to taxa-4(5), ll(12)-diene. The latter is then functionalised by a series of 
8 cytochrome P450-mediated oxygenations and two CoA-dependent acylations, and undergoes 
oxidation at C9 and ring expansion to the oxetane function en route to 10-deacetyl baccatin III, to 
which the ClO acetate group is added and the side chain at C13 is then appended - in five additional 
steps to afford Taxol.

With "production" of secondary metabolites is meant both intracellular production as well as secretion 
into the medium. The term "modulates" or "modulation" as used herein refers to an increase or a 
decrease in production or biosynthesis of secondary metabolites.

Often an increase of a secondary metabolite is desired but sometimes a decrease of a secondary 
metabolite is wanted. Said decrease can for example refer to the decrease of an undesired 
intermediate product. With an increase in the production of one or more metabolites it is understood 
that said production may be enhanced by at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or at least 
100% relative to the untransformed organism (e.g. a plant or plant cell) which was used to transform 
with an expression vector comprising an expression cassette further comprising at least one 
polynucleotide or homologue or variant or fragment thereof of the invention. Conversely, a decrease in 
the production of the level of a secondary metabolite may be decreased by at least 20%, 30%, 40%, 
50%, 60%, 70%, 80%, 90% or at least 100% relative to the untransformed organism (e.g. plant or plant
cell) which was used to transform with an expression vector comprising an expression cassette further comprising at least one polynucleotide or homologue or variant or fragment thereof of the invention.

An "induced production" of a secondary metabolite means that there is no detectable production of secondary metabolite(s) by the untransformed plant (cell) but that detection becomes possible upon carrying out the transformation according to the invention.

An "enhanced production" of a secondary metabolite means that there exists already a detectable amount of secondary metabolite(s) by the untransformed plant cell but that detection becomes possible upon carrying out the transformation according to the invention and that an increase of secondary metabolite(s) can be measured by at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more then 90% compared to basal secretion by the untransformed plant (cell).

Preferably, the organism (such as a plant or plant cell line) transformed with a polynucleotide of the present invention is induced before it produces secondary metabolites. The wording "inducing the production" means that for example the cell culture, such as a plant cell culture, is stimulated by the addition of an external factor. External factors include the application of heat, the application of cold, the addition of acids, bases, metal ions, fungal membrane proteins, sugars and the like. One approach that has been given interesting results for better production of plant secondary metabolites is elicitation. Elicitors are compounds capable of inducing defence responses in plants. These are usually not found in intact plants but their biosynthesis is induced after wounding or stress conditions. Commonly used elicitors are jasmonates, mainly jasmonic acid and its methyl ester, methyl jasmonate.

Jasmonates are linoleic acid derivatives of the plasma membrane and display a wide distribution in the plant kingdom. They were originally classified as growth inhibitors or promoters of senescence but now it has become apparent that they have pleiotropic effects on plant growth and development. Jasmonates appear to regulate cell division, cell elongation and cell expansion and thereby stimulate organ or tissue formation. They are also involved in the signal transduction cascades that are activated by stress situations such as wounding, osmotic stress, desiccation and pathogen attack. Methyl jasmonate (MeJA) is known to induce the accumulation of numerous defence-related secondary metabolites (e.g. phenolics, alkaloids and sesquiterpenes) through the induction of genes coding for the enzymes involved in the biosynthesis of these compounds in plants. Jasmonates can modulate gene expression from the (post)transcriptional to the (post)translational level, both in a positive as in a negative way. Genes that are upregulated are e.g. defence and stress related genes (PR proteins and enzymes involved with the synthesis of phytoalexins and other secondary metabolites) whereas the activity of housekeeping proteins and genes involved with photosynthetic carbon assimilation are down-regulated. For example: the biosynthesis of phytoalexins and other secondary products in plants can also be
boosted up by signal molecules derived from micro-organisms or plants (such as peptides, oligosaccharides, glycopeptides, salicylic acid and lipophilic substances) as well as by various abiotic elicitors like UV-light, heavy metals (Cu, VOS04, Cd) and ethylene. The effect of any elicitor is dependent on a number of factors, such as the specificity of an elicitor, elicitor concentration, the duration of the treatment and growth stage of the culture.

Generally, secondary metabolites can be measured intracellularly or in the extracellular space by methods known in the art. Such methods comprise analysis by thin-layer chromatography, high pressure liquid chromatography, capillary electrophoresis, gas chromatography combined with mass spectrometric detection, radioimmuno-assay (RIA) and enzyme immuno-assay (ELISA). For example, taxol and taxanes can be measured as described in Ketchum et al. (2007); tobacco alkaloid content can be analysed by GC-MS (Millet et al. 2009); Medicago flavonoid and triterpene saponin content can be analysed by Reversed phase UPLC/ICR/FT-MS (see for example as described in Example section).

When the polynucleotides of the present invention are used to modulate the production of secondary metabolites in plants or plant cells, the invention can be practiced with any plant variety for which cells of the plant can be transformed with an expression cassette of the current invention and for which transformed cells can be cultured in vitro. Suspension culture, callus culture, hairy root culture, shoot culture or other conventional plant cell culture methods may be used (as described in: Drugs of Natural Origin, G. Samuelsson, 1999, ISBN 9186274813). By "plant cells" it is understood any cell which is derived from a plant and can be subsequently propagated as callus, plant cells in suspension, organized tissue and organs (e.g. hairy roots). Tissue cultures derived from the plant tissue of interest can be established. Methods for establishing and maintaining plant tissue cultures are well known in the art (see, e.g. Trigiano R.N. and Gray D.J. (1999), "Plant Tissue Culture Concepts and Laboratory Exercises", ISBN: 0-8493-2029-1; Herman E.B. (2000), "Regeneration and Micropropagation: Techniques, Systems and Media 1997-1999", Agricell Report). Typically, the plant material is surface-sterilized prior to introducing it to the culture medium. Any conventional sterilization technique, such as chlorinated bleach treatment can be used. In addition, antimicrobial agents may be included in the growth medium. Under appropriate conditions plant tissue cells form callus tissue, which may be grown either as solid tissue on solidified medium or as a cell suspension in a liquid medium.

A number of suitable culture media for callus induction and subsequent growth on aqueous or solidified media are known. Exemplary media include standard growth media, many of which are commercially available (e.g., Sigma Chemical Co., St. Louis, Mo.). Examples include Schenk-Hildebrandt (SH) medium, Linsmaier-Skoog (LS) medium, Murashige and Skoog (MS) medium, Gamborg's B5 medium, Nitsch & Nitsch medium, White's medium, and other variations and supplements well known
to those of skill in the art (see, e.g., Plant Cell Culture, Dixon, ed. IRL Press, Ltd. Oxford (1985) and George et al., Plant Culture Media, Vol 1, Formulations and Uses Exegetics Ltd. Wilts, UK. (1987)). For the growth of conifer cells, particularly suitable media include 1/2 MS, 1/2 L.P., DCR, Woody Plant Medium (WPM), Gamborg's B5 and its modifications, DV (Durzan and Ventimiglia, In Vitro Cell Dev. Biol. 30:219-227 (1994)), SH, and White's medium.

In yet another aspect, the invention envisages a method for the production or for stimulating the biosynthesis of secondary metabolites in plant or plant cell comprising the step of contacting said plant or plant cell derived thereof with a polypeptide of the present invention.

Further, the present invention also relates to a method for the production or for stimulating the biosynthesis of secondary metabolites in a plant or a plant cell comprising the steps of transforming said plant or plant cell with a vector of the present invention or with a chimeric gene of the present invention and allowing the plant or the plant cell to grow.

In a particular embodiment the current invention can be combined with other known methods to enhance the production and/or the secretion of secondary metabolites in plant cell cultures such as (1) by improvement of the plant cell culture conditions, (2) by the transformation of the plant cells with a transcription factor capable of upregulating genes involved in the pathway of secondary metabolite formation, (3) by the addition of specific elicitors to the plant cell culture, and (4) by the induction of organogenesis.

The polynucleotides and polypeptides of the present invention can be used in all types of plants to boost the plant's own secondary metabolite production or in plants that are transformed with a combination of genetic material that can lead to the generation of novel metabolic pathways (for example through the interaction with metabolic pathways resident in the host organism or alternatively silent metabolic pathways can be unmasked) and eventually lead to the production of novel classes of compounds. This novel or reconstituted metabolic pathways can have utility in the commercial production of novel, valuable compounds.

The following examples are intended to promote a further understanding of the present invention. While the present invention is described herein with reference to illustrated embodiments, it should be understood that the invention is not limited hereto. Those having ordinary skill in the art and access to the teachings herein will recognize additional modifications and embodiments within the scope thereof. Therefore, the present invention is limited only by the claims attached herein.
EXAMPLE 1. Unraveling the transcriptional network that dictates Taxol biosynthesis

A full genome-wide cDNA-AFLP based transcript profiling on elicited Taxus baccata cells was carried out as described (Rischer et al., 2006; Vuylsteke et al., 2007). Samples were taken from the growth medium at days 5 (T-72) and 7 (T-24) and production medium to which at day 1 (=T0) either methyl jasmonate (MeJA, 100 μM), vanadyl sulfate (50 μM) or an equivalent amount of ethanol was added (samples at 0-1-2-4-8-16-24-48-96H). About 8,000 Taxus genes could be visualised by cDNA-AFLP and the expression of about 800 gene tags were modulated by elicitation. From this collection 680 tag sequences have been obtained, almost all representing novel Taxus sequences. For 400 of these tag sequences a hit in the public database, and thus an indication on the possible function or activity was encountered. Among this list, 8 of the known Taxol biosynthesis genes (Croteau et al., 2006) are present, confirming the value and relevance of the sequence platform (Fig. 1). Additionally, full-length open reading frame sequences have been isolated for genes encoding the Jasmonate ZIM-domain (JAZ) transcription factors, known JA-inducible repressor proteins involved in JA signaling (Chico et al., 2008), indicating that the sequence platform also contains regulatory genes, of JA signaling in general and potentially of taxol biosynthesis in particular (Fig. 1). One such gene (SEQ ID NO: 1; TB595 in Fig. 1) encodes for a hypothetical protein that potentially encodes a peptide hormone that may regulate taxol biosynthesis in planta and in vitro cultures.


To obtain the full-length (FL) cDNA clone corresponding to the short sequence TB595 tag isolated in the cDNA-AFLP analysis, the use of gene-specific primers, RT-PCR, 5′- and 3′-RACE (InVitrogen Life Technologies) techniques were combined to yield a FL TB595 cDNA clone (SEQ ID NO: 1). For transient over-expression of the TB595 gene (SEQ ID NO: 1) in plant protoplasts, the corresponding TB595 FL-ORF sequence was introduced into the p2GW7 high copy vector (Karimi et al. 2002) in which the gene is put under the control of the p35S promoter, using the Gateway™ technology (InVitrogen Life Technologies).

To obtain promoter sequences of taxol synthesis genes from Taxus species, synthetic DNA was made based by GenScript USA Inc for the promoters of taxadiene synthase (p7XS, 1436 nt, EMBL accession EF153471) and taxane 13-a-hydroxylase (pT130H, 1060 nt, EMBL accession EF153469). Promoter sequences were directly synthesized between attB sites (Gateway™ technology; InVitrogen Life Technologies) and recombined into the Gateway™ entry vector pDONR221 and subsequently to the
pGWL7 vector (Karimi et al. 2005), to create the reporter constructs in which the promoters are fused in front of the firefly luciferase open reading frame (pTXS-fLUC and pT130H-fLUC).

For protoplast transfections the automated and standardized transient expression assay in tobacco cells (De Sutter et al. 2005) was applied. In this assay all transcriptional readouts are measured using a dual firefly/Renilla luciferase assay, in which the former is controlled by the reporter promoter and the latter by the 35S CaMV promoter, which serves as internal normalization.

To analyse the regulatory potential of the TB595 gene (SEQ ID NO: 1), it was co-transfected with both of the promoters, in the absence or presence of methyl jasmonate (MeJA) elicitation. It could be demonstrated that transfection with TB595 (SEQ ID NO: 1) and MeJA has a synergistic effect on the transactivation of the pTXS-fLUC reporter construct (Fig. 2).


Promoter sequences for the Catharanthus roseus strictosidine synthase (pSTR; EMBL accession Y10182), and the Nicotiana tabacum putrescine N-methyltransferase (pPMTI; EMBL accession AF126810), and quinolinate phosphoribosyltransferase (pQPRT2; EMBL accession AJ748263) genes were PCR amplified between the Gateway™ attB sites, and Gateway™ recombined into the pDON R221 vector and subsequently to the pGWL7 vector (Karimi et al. 2005), to create the reporter constructs in which the promoters are fused in front of the firefly luciferase open reading frame. For protoplast transfections the automated and standardized transient expression assay in tobacco cells (De Sutter et al. 2005) was applied. In this assay all transcriptional readouts are measured using a dual firefly/Renilla luciferase assay, in which the former is controlled by the reporter promoter and the latter by the 35S CaMV promoter, which served as internal normalization. To analyse the regulatory potential of the TB595 gene (SEQ ID NO: 1) it was co-transfected with all of the promoters, in the absence or presence of methyl jasmonate (MeJA) elicitation. It could be demonstrated that transfection with TB595 and MeJA has a synergistic effect on all of the alkaloid synthesis reporter constructs (Fig. 3). Furthermore, TB595 was also co-transfected with the pQPRT2 promoter and the known regulatory transcription factors (TFs) of nicotine biosynthesis in Nicotiana species, NIORCI (De Sutter et al. 2005) and NbbHLH1 (Todd et al. 2010). It could be demonstrated that transfection with TB595 and MeJA has a synergistic effect on the combined activity of these two TFs in the presence of MeJA (Fig. 3).

For stable over-expression of the TB595 gene (SEQ ID NO: 1) in plants or plant cell cultures, the corresponding TB595 FL-ORF sequence was introduced in the pK7WG2D binary vector (Karimi et al. 2002) in which the gene is put under the control of the p35S promoter, using the Gateway™ technology (Invitrogen Life Technologies).

For plant cell transformations the ternary vector system (van der Fits et al. 2000) was applied. The plasmid pBBRIMCS-5.virGN54D was used as a ternary vector. The binary plasmid was introduced into Agrobacterium tumefaciens strain LBA4404 already bearing the ternary plasmid by electro-transformation. Alternatively the A. tumefaciens PMP90-C58Cirf™ strain was used. For hairy root transformation the binary plasmid was introduced in the Agrobacterium rhizogenes strain LBA94.2/12.

The resulting Agrobacterium strains are used to create transgenic Taxus baccata hairy root or cell lines (Exposito et al. 2010). Taxoid content is analysed by HPLC as previously described (Bonfill et al. 2007; Ketchum et al. 2007).

EXAM PLE 5. Functional analysis of the TB595 gene by stable transformation of other medicinal plant cells.

For plant cell transformations the ternary vector system (van der Fits et al. 2000) was applied. The plasmid pBBRIMCS-5.virGN54D was used as a ternary vector. The binary plasmid was introduced into Agrobacterium tumefaciens strain LBA4404 already bearing the ternary plasmid by electro-transformation. For whole plant transformation the binary plasmid was introduced in the Agrobacterium tumefaciens strain PMP90-C58Cirf. For hairy root transformation the binary plasmid was introduced in the Agrobacterium rhizogenes strain LBA94.2/12.

The resulting Agrobacterium strains were used to create transgenic plant (Nicotiana tabacum), hairy root (Nicotiana tabacum and Medicago truncatula) or cell lines (Nicotiana tabacum).

Leaves of Nicotiana tabacum cv. Xanthi plantlets grown in in vitro conditions were used for the establishment of hairy roots, after their infection with Agrobacterium rhizogenes LBA 9402 carrying the pK7WG2D expression vector with TB595 or GUS under the control of CaMV 35S promoter, and the nptI gene as a marker. The Agrobacterium was cultured on solid YEB medium supplemented with 100 mg/L spectinomycin, 300 mg/L Streptomycin and 100μg/mL rifampicin in the dark at 28°C during 48h. Leaves were infected by wounding the mid-ribs with a sterile needle inoculated with the desired Agrobacterium strain. Hairy roots appeared 2-4 weeks after the infection and they were excised and
cultured individually on MS medium supplemented with 30 g/l sucrose and 500 mg/L cefotaxime to eliminate the bacteria and 100 mg/L kanamycin for hairy root selection. Hairy root lines were kept in the dark at 25 °C and routinely subcultured every 3-4 weeks. After several subcultures, 1 ± 0.1 g of the root lines were transferred to 10 mL MS liquid medium and kept in a rotary shaker at 100 rpm, 25 °C and in the dark, and 1 ± 0.5 g of the root lines were transferred to half-strength Gamborg’s B5 salts, full strength Gamborg’s vitamins (B5/2). Subculturing was performed every 3-4 weeks in fresh medium.

For jasmonate elicitor treatment, 2.5 ± 0.5 g of N. tabacum hairy roots carrying theTB595 gene and control roots were inoculated in plates with B5/2 solid culture medium and were supplemented at the beginning of the culture with 100 µM methyl jasmonate. The cultures were kept in the dark at 25 °C. Samples were taken after 8 days.

Nicotine alkaloids were extracted modifying the protocol described in Sheng et al. (2005). 100 ± 20 mg powdered lyophilized roots were resuspended in 200 µL NH3-H2O 6% and 4 mL methanohdichloromethane in 1:3 (v:v), kept in an ultrasound bath for 15 min and filtered. After evaporation, the supernatant was washed 3 times with methyl tert-butyl ethendichlormethane (MTBE:DCM) in 1:0.3 (v:v), evaporated again and kept at -20°C. Finally, just before its analysis, the extract was resuspended in 400 µL MTBE:DCM 1:0.3 (v:v). The samples were determined using a Flame Ionization Detector (FID) Gas Chromatography system (Hewlett-Packard HP5890 Series II Plus). The analyses were performed on an RTX-35 amine column (30 m, 0.25 mm i.d. x 0.5µm d.f.) by using split mode, with 1 mL/min of Helium flux and a programmed oven temperature from 60 °C to 240 °C, increasing 10 °C/min. Injector and detector temperatures were kept at 270 °C. Aliquots of 1 µL were injected by a split-splitless 6890 AutoSampler (Agilent Technologies). The nicotine alkaloids were identified by comparing retention times with the standards: Nicotine, Nornicotine (LGC Promochem, Barcelona, Spain), Anabasine (Chromadex, Irvine, CA, USA) and Anatabine (TransM IT, Marburg, Germany).

Expression of the TB595 gene in transgenic tobacco roots was verified and confirmed with RT-PCR (Hakkinen et al., 2007). RNA was isolated using the "REAL ARNzol SPIN KIT" (REAL, Valencia, Spain), cDNA was prepared from 1 µg of RNA with SuperscriptII reverse transcriptase (Invitrogen, California, USA) and qRT-PCR was performed using SYBR Green PCR Mastermix (Roche, USA) in a 384-well platform system (LightCycler ® 480 Instrument, Roche, USA). Gene specific primers were designed with Primer3 software version 0.4.0 (Table 2) and the amplification efficiency of each primer pair was determined empirically by 10-fold serial dilutions of cDNA and calculated as described by Qiagen. Only those primer pairs with an efficiency of over 0.8 were used. Expression levels were normalized to the levels of the β-ATPase from Nicotiana tabacum.
TABLE 2: Sequences of the primers used to amplify the genes by quantitative real-time PCR.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer Sequence (SEQ ID)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-ATPase</td>
<td>Sense 5'-CCATCAACACCCACGAAAGAATCC-3' (SEQ ID NO: 4)</td>
<td>Häkkinen et al., 2007</td>
</tr>
<tr>
<td></td>
<td>Reverse 5'-GATGACCTGGCAAGGACCC-3' (SEQ ID NO: 5)</td>
<td></td>
</tr>
<tr>
<td>PMT</td>
<td>Sense 5'-TGGATGAGCAATTCAACA-3' (SEQ ID NO: 6)</td>
<td>Häkkinen et al., 2007</td>
</tr>
<tr>
<td></td>
<td>Reverse 5'-AACCAATTCCCTCCGCGGATG-3' (SEQ ID NO: 7)</td>
<td></td>
</tr>
<tr>
<td>QPRT</td>
<td>Sense 5'-ATACGGAGGGCTCAGGAATG-3' (SEQ ID NO: 8)</td>
<td>Häkkinen et al., 2007</td>
</tr>
<tr>
<td></td>
<td>Reverse 5'-GTCAGTGCTTTTACGGAATGC-3' (SEQ ID NO: 9)</td>
<td></td>
</tr>
<tr>
<td>TB595</td>
<td>Sense 5'-TGAGCTGCATAGTCCGTGGTTC-3' (SEQ ID NO: 10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reverse 5'-TCGCATGATTTGAGGAAGCTT-3' (SEQ ID NO: 11)</td>
<td></td>
</tr>
</tbody>
</table>

The main alkaloid found both in control and TB595 root lines was nicotine. This alkaloid was produced in quantities 30-50 times higher than the other alkaloids studied. The improvement of nicotine production in response to MeJA elicitation, was significantly higher in the TB595 root lines (annotated with 1, 6, 7, 9, 10, 11, and 12 in Fig. 4) than in the control lines (annotated with C1, C2 and C11 in Fig. 4). Nicotine content in the TB595 root lines increased, on average, 3.5 times after MeJA elicitation, whereas the average increase observed in the control root lines was only 2.2 (Fig. 4). This analysis demonstrated that TB595 can stimulate nicotine biosynthesis in tobacco roots.

*A. rhizogenes* mediated transformation of *M. truncatula* (ecotype Jemalong J5) hairy roots was done according to a protocol adapted from Boisson-Dernier et al. (2001). Expression of the TB595 gene in transgenic *Medicago* roots was confirmed with RT-PCR. *Medicago* flavonoid and triterpene saponin content is analysed by Reversed phase UPLC/ICR/FT-MS (Pollier et al., 2011).

Metabolite extracts for saponin profiling are obtained from *M. truncatula* hairy roots grown for 21 days in liquid medium. The hairy roots are harvested and rinsed with purified water under vacuum filtration and snap-frozen in liquid nitrogen. Subsequently, the roots are ground in liquid nitrogen, and 400 mg of ground material is extracted with 1 ml methanol at room temperature for 10 minutes, followed by centrifugation for 10 minutes at 14000 rpm. 500 µl of the supernatant is evaporated to dryness under vacuum and the residue is treated with 400 µl water and 200 µl cyclohexane. Samples are centrifugated (10 minutes at 14000 rpm) and 200 µl of the aqueous phase is retained for analysis.
For reversed phase LC, an Acquity UPLC™ BEH C18 column (150 x 2.1 mm, 1.7 µm; Waters, Milford, MA) is serially coupled to an Acquity UPLC™ BEH C18 column (100 x 2.1 mm, 1.7 µm) and mounted on an ultrahigh performance LC system consisting of an Accela pump (Thermo Electron Corporation, Bremen, Germany) and Accela autosampler (Thermo Electron Corporation). The Accela LC system is hyphenated to a LTQ FT Ultra (Thermo Electron Corporation) via an electrospray ionization source. The following gradient is run using waternaconitriile (99:1, v:v) acidified with 0.1% acetic acid (solvent A) and acetonitrile:water (99:1, v:v) acidified with 0.1% acetic acid (solvent B): time 0 min, 5% B; 30 min, 55% B; 35 min, 100% B. The loop size, flow and column temperature are 25 µl, 300 µl/min and 80° C. respectively. Full loop injection is applied. Negative ionization is obtained using the following parameter values: capillary temperature 150°C, sheath gas 25 (arb), aux gas 3 (arb), and spray voltage 4.5 kV. Full FT-MS spectra between m/z 120-1400 are recorded at a resolution of 100,000. For identification, full MS spectra are interchanged with a dependent MS² scan event in which the most abundant ion in the previous full MS scan are fragmented, and two dependent MS² scan events of the two most abundant daughter ions. The collision energy is set at 35%.

**Example 6. Functional analysis of TB595 orthologs from other plant species.**

To identify sequences orthologous to TB595 from other plant species, BLAST analysis (Altschul et al., 1997) was performed. This analysis indicated that the TB595 sequence was highly conserved in the plant kingdom. Candidate orthologs could be identified in gymnosperms, angiosperms (monocots and dicots), mosses and ferns. Next, the protein sequences of the candidate orthologs from A. thaliana and M. truncatula were aligned with that of TB595 with ClustalW and the resulting alignments were manually adjusted (Fig. 5). The phylogenetic tree was generated in MEGA 4.0.1 software (Tamura et al., 2007), by the Neighbor-Joining method, and bootstrapping was done with 10,000 replicates. The evolutionary distances were computed with the Poisson correction method, and all positions containing gaps and missing data were eliminated from the data set (complete deletion option). This analysis indicated that TB595 is most related to A. thaliana At2g20562 and M. truncatula MtTC163920 (Fig. 6). Besides these genes, A. thaliana and M. truncatula possess 1 and 7 additional isoforms of the TB595 gene, respectively (Fig. 6).

The full-length sequences from the 2 A. thaliana and the 8 M. truncatula TB595-like genes were PCR-amplified. For hpRNAi-mediated gene silencing the PCR products were cloned by Gateway™ recombination into the binary vector pK7GW1WG2D(l l) (Karimi et al., 2002). For overexpression the PCR products were cloned by Gateway™ recombination into the binary vector pK7WG2D (Karimi et al., 2002). For whole plant transformation the binary plasmid was introduced in the Agrobacterium tumefaciens strain PM P90-CS8Clrif. For hairy root transformation the binary plasmid was introduced
in the Agrobacterium rhizogenes strain LBA94.2/12. The resulting Agrobacterium strains were used to create transgenic plant (Arabidopsis thaliana), or hairy root (Medicago truncatula) lines.

A. rhizogenes mediated transformation of M. truncatula (ecotype Jemalong J5) hairy roots with the TB595-like genes was done according to a protocol adapted from Boisson-Dernier et al. (2001).

Expression of the TB595 orthologs in transgenic Medicago roots was verified and confirmed with RT-PCR. Medicago flavonoid and triterpene saponin content is analysed by Reversed phase UPLC/ICR/FT-MS (Pollier et al., 2011), as further detailed in Example 5.

Homozygous transgenic Arabidopsis T3 plants overexpressing or silencing the Arabidopsis TB595-like genes were generated by A. tumefaciens mediated transformation with the floral dip method (Clough & Bent, 1998). Expression of the TB595 orthologs in transgenic tobacco roots was verified and confirmed with RT-PCR. Arabidopsis phenolics and glucosinolate content is analysed by Reversed phase UPLC/ICR/FT-MS. Arabidopsis leaf tissue is harvested and snap-frozen in liquid nitrogen. Following ball-milling with a Retsch mill (25 Hz) for 20s, the homogenized plant material is extracted with 0.5 mL methanol. Of the supernatans, 0.4 mL is lyophilized and the pellet re-dissolved in 0.8 mL milliQ water/cyclohexane (1/1, v/v). From the water phase, 10 µL is used for phenolic profiling. Extracts are analyzed with an Accela ultrahigh performance liquid chromatography (UHPLC) system (Thermo Electron Corporation, Bremen, Germany) consisting of an Accela autosampler coupled to an Accela pump and further hyphenated to a LTQ FT Ultra (Thermo Electron Corporation) mass spectrometer (MS) consisting of a linear ion trap (IT) MS connected with a Fourier transform ion cyclotron resonance (FT-ICR) MS. The separation is performed on a reversed phase Acquity UPLC HSS C18 column (150 mm x 2.1 mm, 1.8 µm; Waters, Milford, MA) with aqueous 0.1% acetic acid and acetonitrile/water (99/1, v/v, acidified with 0.1% acetic acid) as solvents A and B. At a flow of 300 µL/min and a column temperature of 60°C, the following gradient is applied: 0 min 1% B, 30 min 60% B, 35 min 100% B. The autosampler temperature is 5°C. Analytes are negatively ionized with an electrospray ionization (ESI) source using the following parameter values: spray voltage 3.5 kV, capillary temperature 300°C, sheath gas 40 (arb), aux gas 20 (arb). Full FT-ICR-MS spectra between 120 and 1400 m/z are recorded (1.2-1.7 sec/scan) at a resolution of 100,000. In parallel, four data dependent MSn spectra are recorded on the ITMS using the preliminary low resolution data obtained during the first 0.1 sec of the previous full FT-ICR-MS scan: a MS2 scan of the most abundant m/z ion of the full FT-ICR-MS scan, followed by two MS3 scans of the most abundant first product ions and a final MS4 scan of the most abundant second product ion obtained from the base peak in the MS2 spectrum. MSn scans are obtained with 35% collision energy.
**Example 7. Determination of the mature, processed, TB595 peptide sequence.**

The TB595 ORF encodes for a peptide of 73 amino acids, of which the first 27 aa putatively correspond to signal peptide, as predicted by the SignalP software (Fig. 7 & 8) (Emanuelsson et al. 2007). Hence the predicted mature peptide consists of 46 aa, and its sequence was highly conserved in the plant kingdom, as determined with the ClustalW software (Fig. 5 & 8), see also Example 6 (Larkin et al. 2007).

To verify the nature of TB595 as a potential signal peptide, the ORF encoding the marker protein Venus-GFP was fused to the C-terminus of TB595, and the fusion gene put under the control of the p35S promoter in a binary vector, using the Gateway™ technology (InVitrogen Life Technologies). The binary plasmid was introduced into the *Agrobacterium tumefaciens* strain and the resulting strain used to perform either leaf infiltrations of *Nicotiana benthamiana* plants as described in Boruc et al. (2010) or stable transformation of *Nicotiana tabacum* BY-2 cells and Arabidopsis thaliana cells as described (Goossens et al. 2003; Pauwels et al., 2010). Confocal microscopy to determine the localization of the TB595 peptide was determined as described (Boruc et al. 2010). To estimate the size of the active TB595 peptide, the ORF encoding the marker protein Venus-GFP was fused to the C-terminus of TB595, and the fusion gene put under the control of the p35S promoter in a binary vector, using the Gateway™ technology (InVitrogen Life Technologies). The binary plasmid was introduced into the *Agrobacterium tumefaciens* strain and the resulting strain used to stably transform *Nicotiana tabacum* BY-2 cells as described (Goossens et al. 2003). Accumulation of the HIS-tagged TB595-peptide is verified by western blot analysis.

**Example 8. Elicitation of medicinal plant cells by synthetic and recombinant TB595 peptide.**

Both synthetic and recombinant versions (from TB595-expressing *Saccharomyces cerevisiae* or tobacco BY-2 cells or *Escherichia coli*) of the mature peptide are generated.

An N-terminal His or FLAG tagged TB595 peptide (46aa) for expression and purification from *E. coli* is generated. His or FLAG tag cleavage recognition sites incorporated between the tag and the peptide itself can be used to obtain the native peptide post-purification (http://www.qiagen.com/products/protein/purification/tagzymesystem/default.aspx).

Due to the highly hydrophobic nature of the peptide the peptide is also expressed with an excretory signal in yeast. This is achieved by generating a fusion of the M-Falpha-1 prepro-leader sequence (89aa; Rothblatt et al. 1987) N-terminal to the HIS-tagged TB595 for excretion of the tagged protein into the medium. The peptide is subsequently purified using the HIS tag before it is cleaved to obtain the native peptide.
The pure synthetic or recombinant peptides are applied in various concentrations to suspension cells, hairy roots and whole plants from various medicinal plant species, in particular *Taxus baccata*, *Nicotiana tabacum*, and *Medicago truncatula*, in which we subsequently analyse accumulation of taxoids, pyridine alkaloids and triterpene saponins/flavonoids, respectively.
REFERENCES


1. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide encoded by a polynucleotide comprising SEQ ID NO: 1;
(b) a polypeptide comprising a polypeptide sequence having a least 75% identity to the polypeptide encoded by a polynucleotide sequence having SEQ ID NO: 1;
(c) a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 2;
(d) a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 3;
(e) a polypeptide comprising an amino acid sequence with at least 75% identity to SEQ ID NO: 2;
(f) a polypeptide comprising an amino acid sequence with at least 75% identity to SEQ ID NO: 3;
(g) fragments and/or variants of the polypeptides according to (a), (b), (c), (d), (e) or (f).

2. An isolated polypeptide according to claim 1 wherein said polypeptide sequence is consisting of an amino acid sequence as set forth in SEQ ID NO: 2 or 3 and polypeptide sequences having at least 75% identity to SEQ ID NO: 2 or 3.

3. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising a polynucleotide sequence having the sequence SEQ ID NO: 1;
(b) a polynucleotide comprising a polynucleotide sequence having at least 70% identity to the sequence having SEQ ID NO: 1;
(c) a polynucleotide which encodes the polypeptide sequence as set forth in SEQ ID NO: 2;
(d) a polynucleotide which encodes the polypeptide sequence as set forth in SEQ ID NO: 3;
(e) fragments and variants of the polynucleotides according to (a), (b), (c) or (d).

4. A chimeric gene comprising the following operably-linked sequences: a) a promoter region capable of directing expression in a plant or plant cell; b) a DNA region encoding a polypeptide as defined in claim 3; c) a 3' polyadenylation and transcript termination region.

5. A vector comprising a polynucleotide sequence as defined in claim 3 or a chimeric gene as defined in claim 4.

6. A host cell comprising a polynucleotide sequence as defined in claim 3 or a chimeric gene as defined in claim 4 or a vector as defined in claim 5.

7. A transgenic plant or a cell derived thereof that is transformed with a vector as defined in claim 5 or that comprises a chimeric gene as defined in claim 4.

8. A transgenic plant or a cell derived thereof according to claim 7 which is from the group comprising Taxus spp., Medicago spp., Catharanthus spp., Nicotiana spp., Artemisia spp.
9. Use of a polynucleotide as defined in claim 3 or a polypeptide as defined in claims 1 or 2 or a chimeric gene as defined in claim 4 to produce secondary metabolites in a plant or plant cell and/or to stimulate the biosynthesis of secondary metabolites in a plant or plant cell.

10. Use according to claim 9 wherein said plant or plant cell belongs to the group comprising Taxus spp., Medicago spp., Catharanthus spp., Nicotiana spp., Artemisia spp.

11. Use according to any of claims 9 or 10 wherein said secondary metabolites comprise alkaloid compounds, phenylpropanoid compounds or terpenoid compounds.

12. A method for the production or for stimulating the biosynthesis of secondary metabolites in a plant or plant cell comprising the step of contacting said plant or plant cell derived thereof with a polypeptide as defined in claims 1 or 2.

13. A method for the production or for stimulating the biosynthesis of secondary metabolites in a plant or a plant cell comprising the steps of transforming said plant or plant cell with a vector as defined in claim 5 or with a chimeric gene as defined in claim 4 and allowing the plant or the plant cell to grow.
Figure 3 continued

**pPMT1**

- CON -
- TB595 -
- CON + MeJA
- TB595 + MeJA

**pQPRT2**

- CON -
- OB -
- TB595/OB -
- CON +MeJA
- OB + MeJA
- TB595/OB + MeJA
**SignalP-NN result:**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Position</th>
<th>Value</th>
<th>Cutoff</th>
<th>signal peptide?</th>
</tr>
</thead>
<tbody>
<tr>
<td>max. C</td>
<td>24</td>
<td>0.568</td>
<td>0.32</td>
<td>YES</td>
</tr>
<tr>
<td>max. Y</td>
<td>28</td>
<td>0.293</td>
<td>0.33</td>
<td>NO</td>
</tr>
<tr>
<td>max. S</td>
<td>18</td>
<td>0.994</td>
<td>0.87</td>
<td>YES</td>
</tr>
<tr>
<td>mean S</td>
<td>1-27</td>
<td>0.944</td>
<td>0.48</td>
<td>YES</td>
</tr>
<tr>
<td>D</td>
<td>1-27</td>
<td>0.619</td>
<td>0.43</td>
<td>YES</td>
</tr>
</tbody>
</table>

# Most likely cleavage site between pos. 27 and 28: VGA-VV
FIGURE 8

MGECRPLGFLGLPFAVSVAVVWILGTLSCICGCICFSGLANMAVGLMKLPKIMRWSIHQIPC
<table>
<thead>
<tr>
<th>Box No.</th>
<th>Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:</td>
</tr>
<tr>
<td>a.</td>
<td>(means)</td>
</tr>
<tr>
<td></td>
<td>□ on paper</td>
</tr>
<tr>
<td></td>
<td>☑ in electronic form</td>
</tr>
<tr>
<td>b.</td>
<td>(time)</td>
</tr>
<tr>
<td></td>
<td>☑ in the international application as filed</td>
</tr>
<tr>
<td></td>
<td>□ together with the international application in electronic form</td>
</tr>
<tr>
<td></td>
<td>□ subsequently to this Authority for the purpose of search</td>
</tr>
<tr>
<td>2.</td>
<td>In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.</td>
</tr>
<tr>
<td>3.</td>
<td>Additional comments:</td>
</tr>
</tbody>
</table>
### INTERNATIONAL SEARCH REPORT

**PCT/EP2012/050969**

**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)

EPO-Internal, BIOSIS, EMBL, FSTA, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>

[X] Further documents are listed in the continuation of Box C. [X] 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

"A" document member of the same patent family

**Date of the actual completion of the international search**

10 July 2012

**Date of mailing of the international search report**

30/07/2012

**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

Obel, Ni col ai
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>
| A        | DATABASE Tai r [O]nline]  
29 September 2008 (2008-09-29), XP002679284, retrieved from arabi dopsi s  
Database accession no. AT2G20562  
the whole document | 1-13 |
<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td>US 2004031072 A1</td>
<td>12-02-2004</td>
<td>NONE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 2450446 A2</td>
<td>09-05-2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2010505449 A</td>
<td>25-02-2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW 200817509 A</td>
<td>16-04-2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2007072224 A2</td>
<td>28-06-2007</td>
</tr>
<tr>
<td>US 2008292735 A1</td>
<td>27-11-2008</td>
<td>NONE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 102171344 A</td>
<td>31-08-2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2010192244 A1</td>
<td>29-07-2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2009063312 A2</td>
<td>22-05-2009</td>
</tr>
</tbody>
</table>