(12) PATENT **(11)** Application No. **AU 199882432 B2**
(19) AUSTRALIAN PATENT OFFICE **(10)** Patent No. **754464**

(54) Title
Genes encoding proteins having transglycosylation activity

(51)7 International Patent Classification(s)
C12N 015/54 **C12N 009/10**
A01H 001/00

(21) Application No: **199882432** (22) Application Date: **1998.07.16**

(87) WIPO No: **WO99/05287**

(30) Priority Data

(31) Number	(32) Date	(33) Country
9-200571	**1997.07.25**	**JP**

(43) Publication Date : **1999.02.16**
(43) Publication Journal Date : **1999.04.15**
(44) Accepted Journal Date : **2002.11.14**

(71) Applicant(s)
Suntory Limited

(72) Inventor(s)
Masako Mizutani; Yoshikazu Tanaka; Takaaki Kusumi; Kazuki Saito; Mami Yamazaki; Zhizhong Gong

(74) Agent/Attorney
GRIFFITH HACK,GPO Box 1285K,MELBOURNE VIC 3001

(56) Related Art
WO 96/00291

OPI DATE 16/02/99 APPLN. ID 82432/98
AOJP DATE 15/04/99 PCT NUMBER PCT/JP98/03199

AU9882432

PC

(51) 国際特許分類6
C12N 15/54, 9/10, A01H 1/00

A1

(11) 国際公開番号 WO99/05287

(43) 国際公開日 1999年2月4日(04.02.99)

(21) 国際出願番号 PCT/JP98/03199

(22) 国際出願日 1998年7月16日(16.07.98)

(30) 優先権データ
特願平9/200571 1997年7月25日(25.07.97) JP

(71) 出願人（米国を除くすべての指定国について）
サントリー株式会社(SUNTORY LIMITED)[JP/JP]
〒530-8203 大阪府大阪市北区堂島浜2丁目1番40号 Osaka, (JP)
(72) 発明者；および
(75) 発明者／出願人（米国についてのみ）
水谷正子(MIZUTANI, Masako)[JP/JP]
〒615-8083 京都府京都市西京区桂良町18-21 Kyoto, (JP)
田中良和(TANAKA, Yoshikazu)[JP/JP]
〒520-0246 滋賀県大津市仰木の里2-7-4 Shiga, (JP)
久住高章(KUSUMI, Takaaki)[JP/JP]
〒564-0073 大阪府吹田市山手町2-12-21-402 Osaka, (JP)
斉藤和季(SAITO, Kazuki)[JP/JP]
〒289-1106 千葉県八街市榎戸663-86 Chiba, (JP)
山崎真巳(YAMAZAKI, Mami)[JP/JP]
〒260-0045 千葉県千葉市中央区弁天4-12-6 Chiba, (JP)
鞏　志忠(GONG, Zhizhong)[CN/JP]
〒263-0031 千葉県千葉市稲毛区稲毛東3丁目12番5号
並木コーポ201 Chiba, (JP)
(74) 代理人
弁理士　石田　敬, 外(ISHIDA, Takashi et al.)
〒105-8423 東京都港区虎ノ門三丁目5番1号 虎ノ門37森ビル
青和特許法律事務所 Tokyo, (JP)

(81) 指定国 AU, CA, CN, JP, KR, NZ, US, 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

添付公開書類
国際調査報告書

(54)Title: GENES ENCODING PROTEINS HAVING TRANSGLYCOSYLATION ACTIVITY

(54)発明の名称 糖転移活性を有する蛋白質をコードする遺伝子

(57) Abstract

Genes encoding proteins each having an amino acid sequence represented by any of SEQ ID NOS: 7 to 10 and 12 and showing the activity of transferring a glycosyl group to the 5-position of a flavonoid; genes encoding proteins each having an amino acid sequence derived from any of the above amino acid sequences by modification and showing the activity of transferring a glycosyl group to the 5-position of a flavonoid; and a process for producing the above proteins with the use of these genes. These genes are usable in, for example, artificially improving plant colors.

ABSTRACT

The present invention provides a nucleic acid molecule that codes for a protein having an amino acid sequence described in any of SEQ ID NOs: 7 through 10 or 12 and having activity that transfers a glycoside to the 5 position of a flavonoid, a nucleic acid molecule that codes for a protein having a modified amino acid sequence relative to the above amino acid sequence and having activity that transfers a glycoside to the 5 position of a flavonoid, and a process for producing the above protein using said nucleic acid molecule. This nucleic acid molecule can be used to artificially alter the colour of plants.

SPECIFICATION

GENE CODING FOR A PROTEIN HAVING GLYCOSIDE TRANSFER ACTIVITY

Technical Field

The present invention relates to a gene coding for a protein having activity that transfers a glycoside to the 5 position of a flavonoid, and a process utilizing that gene.

Background Art

The flower industry strives to develop various new varieties. Changing the color of a flower is one way of effectively breeding a new variety. A wide range of colors have been successfully produced for nearly all commercial varieties using classical breeding methods. With these methods, however, since there are restrictions on the gene pool for each species, it is rare for a single species to have a broad range of colored varieties.

Flower colors are based on two types of pigments, namely flavonoids and carotinoids. Flavonoids contribute to color tones ranging from yellow to red and blue, while carotinoids contribute to color tones of orange or yellow. Flavonoid molecules that primarily contribute to flower color are anthocyanins which are glycosides of cyanidin, delphinidin, petunidin, peonidin, malvidin and pelargonidin, and different anthocyans cause remarkable changes in flower color. Moreover, flower color is also affected by auxiliary coloring by colorless flavonoids, metal complex formation, glucosylation, acylation, methylation and vacuolar pH (Forkmann, Plant Breeding, 106, 1, 1991).

The biosynthesis route of anthocyanins, which begins with phenylalanine, has been well understood (e.g., Plant Cell, 7, 1071-1083, 1995), and nearly all genes involved in the biosynthesis have been cloned. For example, among those genes thought to be involved in biosynthesis of malonylshisonin (3-0-(6-0-(p-cumaloyl)-β-D-glucosyl)-5-0-(6-0-malonyl-β-D-glucosyl)-cyanidin), which is an anthocyanin of Perilla, those genes for which homologues have not yet been reported are only the flavonoid-3'-hydroxylase, UDP-glucose:anthocyanin (flavonoid) 5-0-glucosyl transferase (abbreviated as 5GT) and malonyl group transferase genes.

Among these, flavonoid-3'-hydroxylase is known to belong to the cytochrome P450 gene family (Plant Cell, 7, 1071-1083, 1995), and cytochrome P450 genes are surmised to demonstrate structural homology.

The hydroxyl group at the 3 position of flavonoid molecules is typically modified by glucose, and generally glucosylation and other modifications by glycoside are considered to increase the stability and solubility of anthocyanins (The Flavonoids, Chapman & Hall, 1994).

Genes coding for the UDP-glucose:anthocyanidin or flavonoid-3-glucosyl transferase (abbreviated as 3GT) that catalyze this reaction are obtained from numerous plants such as corn, barley, snapdragons and gentians, and their amino acid sequences mutually demonstrate significant homology. For example, the homology between the 3GT amino acid sequences of monocotyledonous corn and dicotyledoneous gentian is 32%, that between the 3GT amino acid sequences of monocotyledonous corn and monocotyledonous barley is 73%, and that between the 3GT amino acid sequences of dicotyledonous gentian and dicotyledonous eggplant is 46%.

In addition, the gene coding for UDP-ramnose:anthocyanidin 3-glucosidoramnosyl transferase (3RT) of petunias has also been cloned.

However, even though the hydroxyl group at the 5 position of the flavonoids of numerous plants is glucosylated, a gene for the enzyme (5GT) that catalyzes this reaction has yet to be obtained.

In addition, although there are examples of measuring the reaction by which glycoside is transferred to the 5 position of petunia and stock anthocyanins (Planta, 160, 341-347, 1984, Planta, 168, 586-591, 1986), these reports only describe the investigation of enzymological properties using crude extracts or partially purified products of flower petals, and there are no examples of this enzyme being purified to its pure form. In addition, since glycosyltransferases are typically biochemically unstable, enzyme purification is difficult.

Although there are hardly any cases in which colour tone is changed by addition of glycoside to a flavonoid molecule, since aromatic acyl groups that have a significant effect on colour tone are linked to a glucose molecule or ramnose molecule within an anthocyanin, regulation of the glycoside transfer reaction is important in terms of controlling anthocyanin biosynthesis, and ultimately in controlling flower colour. Furthermore, as an example of changing flower colour by regulating the expression of glycosyltransferase gene, the reaction by petunia 3RT has been controlled in transformed petunia to modify flower colour.

Plant species, which can be transformed with a foreign gene, include, for example, roses, chrysanthemums, carnations, daisies, petunias, torenia, bellflowers, calanchoes, tulips and gladiolas.

All references, including any patents or patent applications, cited in this specification are hereby incorporated by reference. No admission is made that any reference constitutes prior art. The discussion of the references states what their authors assert, and the applicants reserve the right to challenge the accuracy and pertinency of the cited documents. It will be clearly understood that, although a number of prior art publications are referred to herein, this reference does not constitute an admission that any of these documents forms part of the common general knowledge in the art, in Australia or in any other country.

For the purposes of this specification it will be clearly understood that the word "comprising" means "including but not limited to", and that the word "comprises" has a corresponding meaning.

Disclosure of the invention

The present invention provides an isolated nucleic acid molecule coding for a protein having activity that transfers a glycoside to the 5 position of a flavonoid.

The inventors of the present invention sought to obtain a nucleic acid molecule that codes for a protein having activity that transfers a glycoside to the 5 position of a flavonoid, thereby leading to completion of the present invention.

For example, the 5 position hydroxyl group of the anthocyanins of chrysanthemums and some of the anthocyanins of roses and carnations are not glycosylated. The anthocyanin structure can be changed by introducing the 5GT nucleic acid molecule obtained by the present invention into these plants.

In addition, although it is possible to change flower colour and stabalise flavonoids by acylating flavonoids using the acyl group transferase gene described in International Publication No. WO 96/25500, since the acyl group does not bond directly with the flavonoid, but rather bonds by way of a sugar, simply introducing an acyl group transferase nucleic acid molecule alone is not sufficient for changing flower colour and may even cause the flavonoid not the become stable.

However, by introducing the 5GT nucleic acid molecule in combination with an acyl group transferase nucleic acid molecule, sugar is bounded to the 5 position of the flavonoid thereby further allowing the flavonoid to be acylated. This can be expected to change the anthocyanin structure and cause the flower colour to become bluish.

In addition, if expression of 5GT nucleic acid molecule of a plant in which the 5 position of anthocyanin is glucosylated is suppressed with the antisense method or co-suppression method and so forth, transfer of glucose residue to 5 position can be inhibited. So that, flower colour can be changed. For example, suppressing 5GT activity in gentian or bellflower can be expected to cause flower colour to become reddish.

The inventors of the present invention isolated cDNA of 5GT from Perilla, torenia, verbena and petunia plants using gene recombination technology, and determined the nucleotide sequence of the structural gene. Namely, the inventors of the present invention provide a DNA sequence that codes for 5GT present in the tissue that expresses anthocyanins in these plants. Moreover, since this enzyme transfers glycoside to the 5 position of the anthocyanin pigment, it can be used to change flower colour and increase anthocyanin stability.

Embodiment for Carrying out the Invention

The method of differential displacement, for example, can be used to obtain DNA that codes for the enzyme of the present invention. In Perilla (Perilla frutescens), for example, there are varieties that accumulate anthocyanins (e.g., red forma) and those that do not (e.g., green forma). By cloning DNA present in varieties that accumulate anthocyanins but not present in varieties that do not, it is possible to obtain the DNA that codes for the enzyme of the present invention.

More specifically, RNA is extracted from the leaves of red forma and green forma, and cDNA is synthesized in accordance with standard methods. This is then separated by electrophoresis to isolate cDNA present in the cDNA library of red forma but not present in the cDNA library of green forma. Next, the red forma cDNA library is screened using the resulting cDNA as a probe to obtain the cDNA that codes for the enzyme of the present invention.

Once cDNA that codes for the enzyme of the present invention is obtained in the manner described above, this cDNA or its fragment is used as a probe to screening the cDNA libraries of other plants. As a result, the DNA that codes for the enzyme of the present invention can be obtained from those plants.

As an example of the screening, in the present invention, the DNA coding for the enzyme of the present invention is cloned from Perilla by the differential display method (Example 1). Next, DNA that codes for the enzyme of the present invention is obtained from verbena by screening of cDNAs from verbena (Verbena hybrida) using the cloned DNA of Example 1 as a probe (Example 2). Moreover, DNA coding for the enzyme of the present invention is obtained from torenia in the same manner (Example 3).

Then, it was confirmed that the proteins encoded in these DNAs have the enzymatic activity of the present invention.

Moreover, the DNA coding for the enzyme of the present invention was obtained from petunia (Example 4).

Examples of the DNAs of the present invention include that which codes for the amino acid sequence described in any one of SEQ ID NOs: 7 through 10 or 12. However, proteins having an amino acid sequence modified by addition and/or deletion of one or more amino acids and/or substitutions by one or more other amino acids are also known to maintain enzymatic activity similar to the original protein. Thus, nucleic acid molecules coding for a protein that has an amino acid sequence modified by addition and/or deletions of one or more amino acids and/or substitutions by one or more other amino acids relative to the amino acid sequence described in any one of SEQ ID NOs:7 through 10 or 12, and still maintains activity of transferring a glycoside to the 5 position of a flavonoid, also belong to the present invention.

The present invention also relates to a nucleic acid molecule coding for a protein which nucleic acid molecule hybridizes to a nucleotide sequence described in any one of SEQ ID NOs: 1 through 4 or 6, or to a nucleotide sequence that codes for an amino acid sequence described therein or to their portions, for example a portion coding for at least six amino acids for a consensus region, under conditions of 2 to 5 x SSC, and for example, 5 x SSC, and 50°, and that has activity of transferring a glycoside to the 5 position of a flavonoid. Furthermore, the optimum hybridisation temperature varies according to the nucleotide sequence and its length, and it is preferable that the hybridisation temperature be lower, the shorter the nucleotide sequence. For example, a temperature of 50°C or lower is preferable in the case of a nucleotide sequence (18 bases) coding for six amino acids.

Although examples of nucleic acid molecules selected by hybridisation in this manner include those which are naturally-occurring such as those derived from plants, examples of which include a nucleic acid molecule derived from verbena and torenia, they may also be derived from other plants, examples of which include petunias, roses, carnations and hyacinths. In addition, nucleic acid molecules selected by hybridisation may also be cDNA or genomic DNA.

Moreover, the present invention also relates to a nucleic acid molecule coding for a protein having an amino acid sequence having homology of 30% or more, preferably 50% or more, for example 60% or 70% or more, and in some cases, 90% or more relative to an amino acid sequence of any of SEQ ID NOs: 7 through 10 or 12, and having activity that transfers a glycoside to the 5 position of a flavonoid. Namely, as indicated in Examples, DNA coding for the enzyme of the present invention demonstrates homology of 20 to 30% in comparison with other glycosyltransferase genes. Thus, the present invention includes nucleic acid molecules coding for a protein that having homology of 30% or more with an amino acid sequence described on any of SEQ ID NOs: 7 through 10 or 12, and has glycosyltransferase activity.

In addition, as is clear from a comparison of the results of Examples 1 through 4, the amino acid sequence of the enzyme of the present invention varies according to the species, with interspecies homology being 50% or more (see Examples 3 and 4), and for example 60 to 70% (see Example 2), while the homology of the amino acid sequences of the enzymes derived from the same species is 90% or more (see Example 1). Thus, nucleic acid molecules coding for a protein that has an amino acid sequence having homology of 50% or more, for example 60 to 70% or more, an in some cases, 90% or more, relative to an amino acid sequence described in any one of SEQ ID NOs: 7 through 10 or 12, and maintains the glycosyltransferase activity of the present invention are included in the present invention.

As described in detail in Examples, DNA having a native nucleotide sequence is obtained by, for example, screening of a cDNA library.

In addition, DNA coding for an enzyme having a modified amino acid sequence can be synthesised using ordinary site-specific mutagenesis and PPCR based on the nucleotide sequence of a native DNA. For example, a DNA fragment containing a site at which a modification is desired to be introduced is obtained by restriction enzyme digestion of cDNA or genomic DNA obtained as described above. By using this as a template, site-specific mutagenesis or PRC is performed using a primer containing the desired mutation to obtain a DNA fragment containing the desired modification. This is then ligated to DNA coding for another portion of the target enzyme.

Alternatively, in order to obtain DNA coding for an enzyme having a shortened amino acid sequence, for example, a DNA coding for an amino acid sequence that is longer than the target amino acid sequence, for example that coding for the entire amino acid sequence, is digested by a desired restriction enzyme, and in the case the resulting DNA fragment does not code for the entire target amino acid sequence, the deficient portion should be supplemented by ligating synthetic DNA.

In addition, by expressing this clone using a nucleic acid molecule expression system in E. coli or yeast and measuring enzyme activity, the resulting nucleic acid molecule can be confirmed to code for glycosyltransferase, and by clarifying the translation region of the gylcosyltransferase gene that transfers glycoside to the 5 position of a flavonoid, a gene is obtained that codes for the glycosyltransferase claimed in the present invention. Moreover, by expressing said nucleic acid molecule, the target transferase protein that transfers a glycoside to the 5 position of a flavonoid can be obtained.

Alternatively, the protein can be obtained by using antibody to an amino acid sequence described in any one of SEQ ID NOs: 7 through 10 or 12.

Thus, the present invention also related to a recombinant vector containing the above-mentioned DNA, and more particularly, to an expression vector and a host transformed with the vector. Both prokaryotes and eukaryotes can be used for the host. Examples of prokaryotes that can be routinely used for the host include bacteria, for example, the genus *Escherichia* such as *Escherichia coli*, and the genus *Bacillus* such as *Bacillus subtilis*.

Examples of eukaryotes that can be used include lower eukaryotes such as eucaryotic microorganisms including fungi such as yeast or mold. Examples of yeast includes the genus Saccharomyces such as Saccharomyces cerevisiae, while examples of molds include the genus Aspergillus such as Aspergillus oryzae and Aspergillus niger, as well as the genus Penicillium. Moreover, animal or plant cells can also be used, examples of animal cells including mouse, hamster, monkey and human cell systems. Moreover, insect cells such as silkworm cells or adult silkworms themselves can be used as hosts.

The expression vectors of the present invention contain an expression control region, such as a promoter, terminator or an origin of replication, depending on the type of host in which they are to be introduced. Examples of promoters of bacterial expression vectors include conventionally used promoters such as trc promoter, tac promoter and lac promoter, while examples of yeast promoters include glyceroaldehyde triphosphate dehydrogenase promoter and PH05 promoter. Examples of mold promoters include amylase and trpC. In addition, examples of promoters for animal cell hosts include viral promoters such as SV40 early promoter and SV40 late promoter.

Preparation of expression vector can be performed in accordance with standard methods using restriction enzyme, ligase and so forth. In addition, transformation of a host by an expression vector can also be performed in accordance with standard methods.

In the process for producing the above-mentioned protein, a host transformed with the expression vector is cultured, cultivated or bred, the target protein can be recovered and purified from the resulting culture in accordance with standard methods, examples of which include filtration, centrifugation, cell homogenation, gel filtration chromatography and ion exchange chromatography.

Furthermore, although the present specification describes transferases derived from Perilla, verbena, torenia and petunia wherein the transferases that transfer glycoside to the 5 position of a flavonoid (which may be simply referred to an "glycosyltransferase" in the present invention), a nucleic acid molecule that codes for said enzyme can be cloned, by entirely or partially altering the purification method of said enzyme so as to purify a glycosyltransferase of another plant, and determining the amino acid sequence of said enzyme. Moreover, by using cDNA of the glycosyltransferase derived from Perilla of the present invention as a probe, cDNA of a different glycosyltransferase was able to be obtained from Perilla, and cDNA of a different glycosyltrsnsferase was able to be obtained from a different plant. Thus, other glycosyltransferase nucleic acid molecules can be obtained by using a portion of the entirety of a glycosyltransferase nucleic acid molecule.

In addition, as indicated in the present specification, by purifying glycosyltransferase from Perilla, verbena, torenia and petunia to obtain antibody to said enzyme in accordance with standard methods, cDNA or chromosomal DNA produces protein which reacts with that antibody that can be cloned. Thus, the present invention is not limited to only nucleic acid molecules of glycosyltransferases derived from Perilla, verbena, torenia and petunia, but also relates to glycosyltransferase in the broad sense.

Moreover, the present invention also relates to a plant, its progeny or their tissue for which colour has been adjusted by introduction of glycosyltransferase nucleic acid molecule, and their form may be that of cut flowers as well.

In addition, UDP-glucose is an example of a glycoside donor in the glycoside transfer reaction of glycoside that include anthocuanin in the present specification.

Examples

The following provides a detailed explanation of the present invention based on the Examples. Unless specified otherwise, the experimental procedure was performed in accordance with the methods described in Molecular Cloning (Cold Spring Harbor, 1989), New Biochemistry Experimental Manual (Kagaku Dojin, 1996) and International Patent Laid-Open Publication No. WO96/25500.

Example 1 Cloning of a Nucleic Acid Molecule Specifically Expressed in Red Forma

(1) Differential Display

Perilla (*Perilla frutescens*) includes varieties that accumulate anthocyanins in their leaves (for example, red forma (Sakata-no-tane)), and varieties that do not accumulate anthocyanins (for example, blue forma (Sakata-no-tane)). The structure of the major anthocyanin is reported to be malonylshisonin (3-0-(6-0-(p-cumaloyl)-β-D-glucosyl)-5-0-(6-0-malonyl-β-D-glucosyl)-cyanidin) (Agri. Biol. Chem., 53:197-198, 1989).

Differential display is a method reported in Science, 257, 967-971 (1992), and is used, for example, to obtain genes that are expressed tissue-specifically.

Total RNA was extracted from the leaves of the above-mentioned two types of Perilla by the hot phenol method (Plant Molecular Biology Manual, Kluwer Academic Publishers, 1994, pp. D5/1-13). Poly A + RNA was purified from the resulting total RNA using an mRNA separator kit (Clonetech). 0.9 µg of poly A + RNA were reverse-transcribed in 33 µl of reaction mixture using oligo-dT primer added an anchor (GenHunter, H-T11G, H-T11A and H-T11C) to obtain single strand cDNA. Using this cDNA as a template, PCR was performed using the same oligo-dT primer added an anchor and synthetic primers (GenHunter, H-AP1 through 8) as primers.

The volume of the PCR reaction mixture was 20 µl, and it contained 2 µl of cDNA solution, 0.2 µM of any one of H-T11G, H-T11A or H-T11C primer, 0.2 µM of any primer from H-AP1 through H-AP8, 0.12 µM dNTP, 5 or 10 µCi of $[^{32}P]dCTP$, 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 0.01% Triton X-100, 1.25 mM $MgCl_2$ and 1 unit of Taq polymerase. The reaction conditions comprised holding the temperature at

72°C for 20 seconds followed by repeating the reaction for 40 cycles with one cycle comprising raising the temperature to 94°C for 30 seconds, lowering to 40°C for 2 minutes and raising to 72°C for 30 seconds, and then holding the temperature at 72°C for 5 minutes.

The DNA fragments amplified in this manner were separated by the same polyacrylamide gel electrophoresis as used for DNA Sequencing. After drying the gel, the gel was exposed to X-ray film. Among the resulting approximately 2,600 bands, there were 36 bands observed only in the red forma as a result of comparing the two varieties. They were cut out of the dried gel and eluted into 100 µl of water. The eluted DNA was precipitated with ethanol and dissolved in 20 µl of water. Using a half amount of each DNA as a template, the PCR reaction was performed as described above, and amplified fragments were obtained for 33 of DNA fragments. Library screening and northern analysis were then performed using these DNA fragments.

(2) Northern Analysis

Northern analysis was performed according to the method described below using the above 33 types of DNA probes. After separating poly A + RNA derived from red forma and green forma with formamide gel containing 1.2% agarose, the poly A + RNA was transferred to a Nylon membrane. This membrane was hybridized with the above-mentioned DNA probes labeled with $[^{32}P]$ for overnight at 65°C in the presence of 5XSSPE, 5X Denhalt's solution, 0.5% SDS and 20 µg/ml of denatured salmon sperm DNA. The hybridized membrane was washed at 65°C in 1XSSPE and 0.1% SDS solution and subjected to autoradiography. As a result, only five probes were specifically expressed in red forma. These clones are predicted to be genes involved in the biosynthesis of anthocyanins.

(3) Screening of cDNA Library

A cDNA library with λgt10 as a vector was prepared using the poly A + RNA obtained from the leaves of red forma and the Complete Rapid Cloning System λgt10 (Amersham). This cDNA library was screened with the five DNA fragments described above to obtain cDNA corresponding to each fragment. Among these, a clone named 3R5 was obtained using a DNA fragment obtained by H-T11A and H-AP3 primers, and this clone demonstrated homology of approximately 26% at the amino acid level with previously reported corn flavonoid-3-0-glucosyl transferase.

In addition, clones designated as 3R4 and 3R6 were obtained by library screening using the same probes, and these demonstrated an extremely high level of homology with 3R5. The complete nucleotide sequences and deduced amino acid sequences of 3R4 and 3R6 are shown in SEQ ID NO: 1 and SEQ ID NO: 2 of the Sequence Listing, respectively. In addition, the deduced amino acid sequences of the proteins encoded by 3R4 and 3R6 demonstrated homology of 92%.

A clone designated as 8R6 was obtained using a DNA fragment obtained by H-T11G and H-AP8 primers, and this clone did not demonstrate significant homology with any sequences reported so far. This sequence is shown in SEQ ID NO: 5 of the Sequence Listing. Although there is a strong possibility that 8R6 is a gene involved in the biosynthesis of anthocyanins, since its structure lacks homology with genes reported so far, it is predicted to be a new gene involved in anthocyanin biosynthesis.

In consideration of the anthocyanin structure in Perilla (the previously mentioned malonylshisonin), it is predicted that this gene is a malonyl transferase. In order to verify this, this gene should be expressed in yeast and <u>E. coli</u> followed by reacting with anthocyanin and malonyl-CoA as substrates. Such an experiment can be carried out using, for example, the method described in International Publication No. WO96/25500. Malonyl transferase gene is useful in terms of artificially altering anthocyanin structure.

(4) Expression of 3R4 cDNA in Yeast

An approximately 1.5 kb DNA fragment obtained by blunting the BstXI cleavaged site of p3R4 using T4 DNA polymerase (Takara Shuzo) and then cutting out at the BamHI cleavage site in the adapter, and an approximately 8 kb DNA fragment obtained by blunting the EcoRI cleaved end of pYE22m and then digesting with BamHI were ligated to obtain a plasmid that was designated as pY3R4.

Furthermore, E. coli strain JM109 having pYE22m was named Escherichia coli SBM335, and deposited at the National Institute of Bioscience and Human-Technology Agency of Industrial Science and Technology as FERM BP-5435. In pY3R4, cDNA coding for glycosyltransferase has been ligated downstream of the promoter for glyceroaldehyde triphosphate dehydrogenase lone of the constitutive yeast promoter, and transcription is controlled by this promoter.

Using pY3R4, yeast Saccharomyces cerevisiae G1315 (Ashikari, et al., Appl. Microbiol. Biotechnol., 30, 515-520, 1989) was transformed according to the method of Ito, et al. (Ito, et al., J. Bacteriol., 153, 163-168, 1983). The transformed yeast was selected according to recovery of tryptophan synthesis ability. The resulting transformed strain was cultured for 24 hours at 30°C with shaking in 10 ml of Burkholder's medium (Burkholder, Amer. J. Bot., 30, 206-210) containing 1% casamino acids.

In order to conduct a control experiment, yeast that spontaneously recovered tryptophan synthesis ability was also cultured in the same manner. After collecting the yeast, the cells were suspended in suspension buffer (100 mM phosphate buffer (pH 8.5), 0.1% (v/v) 2-mercaptoethanol, 10 μM APMSF and 100 μM UDP-glucose) followed by the addition of glass beads (Glass Beads, 425-600 microns Acid-Wash, Sigma) and vigorous shaking to crush the cells. The crushed cells were then centrifuged for 20 minutes at 15,000 rpm and the supernatant was used as a crude enzyme solution for the measurement of enzyme activity described below.

(5) Measurement of Enzymatic Activity

After allowing 50 µl of reaction mixture containing 20 µl of crude enzyme solution (100 mM phosphate buffer (pH 8.5), 670 µM cyanidin-3-glucoside, 1 mM UDP-glucose) for 10 minutes at 30 °C, 50 µl of 50% acetonitrile solution containing 0.1% TFA was added to stop the reaction. Supernatant obtained by centrifuging for 5 minutes at 15,000 rpm was passed through a Samprep LCR4(T)-LC filter (Millipore) so as to remove impurities. This was then analyzed by high-performance liquid chromatography (HPLC). Analysis was performed using a reverse phase column (Asahipak ODP-50, 4.6 mm diameter x 250 mm, Showa Denko), the mobile phase consisted of 0.5% TFA/H_2O for solution A and 0.5% TFA 50% CH_3CN for solution B. The flow rate was 0.6 ml/min. and the fractions were eluted at a gradient of B20% → B100% (20 min) followed by holding at B100% for 5 minutes.

20 µl of reaction mixture was used for analysis. A520 nm, AUFS 0.5 (Shimadzu SPD-10A) and a photodiode array detector (Shimadzu SPD-M6A) at an absorbance of 600-250 nm were used for detection. In the case of reaction of yeast crude enzyme solution that expressed pY3R4, in addition to the substrate cyanidin-3-glucoside (retention time: 17 minutes), a new peak was observed at retention time of 14.5 minutes. Since it was not observed in the case of reaction of yeast crude enzyme solution of the control experiment, this new peak was considered to be generated due to the activity of protein originated from pY3R4. As a result of co-chromatography with cyanidin-3,5-diglucoside, the retention time of this peak coincided with that of cyanidin-3,5-diglucoside, and their absorption spectra were also identical to each other. Based on these observations, 3R4 cDNA of Perilla was found to code for 5GT.

Example 2 Cloning of 5GT Nucleic Acid Molecule of Verbena hybrida

(1) Preparation of cDNA Library

Petals were collected from Verbena variety Hanatemari violet (Suntory) and ground by a mortar and pestle in liquid nitrogen. RNA was extracted from the ground tissues according to a method using guanidine thiocyanate/cesium chloride, and poly A + RNA was obtained by the method recommended by the manufacturer using Oligotex (Takara Shuzo). The method using guanidine thiocyanate/cesium chloride was carried out in accordance with the method described in detail in Methods in Molecular Biology, Vol. 2 (Humana Press Inc., 1984) by R. McGookin and Robert J. Slater, et al.

Using the resulting poly A + RNA as a template, double-stranded cDNA was synthesized using the ZAP-cDNA synthesis kit (Stratagene), then, a cDNA library was prepared using the Uni-ZAP XR Cloning Kit (Stratagene) according to the method recommended by the manufacturer.

(2) Cloning of 5GT cDNA

The λ phage library obtained as described above was screened in the following manner using the p3R4 cDNA of Perilla as a probe. The filters were maintained at 42°C for 1 hour in hybridization buffer (5X SSC, 30% formamide, 50 mM sodium phosphate buffer (pH 7.0), 3% SDS 2% blocking reagent (Boehringer), 0.1% lauroylsarcosine, 80 μg/ml of salmon sperm DNA). DIG-labeled Perilla 5GT cDNA, p3R4 cDNA, fragment was added to the hybridization solution and the filters were incubated for further 16 hours.

After washing the filters with washing solution (5X SSC 50°C, 1% SDS), the positive clones labeled with anti-DIG-alkaline phosphate were immunologically detected using 5-bromo-4-chloro-3-indolylphosphate and nitro blue tetrazolium salt according to the method described by the manufacturer (Boehringer).

As a result, seven positive clones were obtained. These cDNA were excised on plasmid pBluescript SK using the method recommended by Stratagene. When the lengths of the cDNA were investigated by agarose gel electrophoresis, insertion of a maximum length of 2.0 kb was observed.

(3) Determination of Nucleotide Sequence

Plasmids were extracted from the resulting clones, and the nucleotide sequences near the 3' and 5' ends of the cDNA were determined according to the dideoxy sequence method using fluorescent reagent as recommended by Perkin-Elmer with the ABI 373A sequencer (Perkin-Elmer). As a result, five of the seven clones had mutually same nucleotide sequences although the lengths of the cDNA were different. The entire nucleotide sequence of pSHGT8 was determined. Determination of nucleotide sequences was performed as described above by either using the Kilo-Sequence Deletion Kit (Takara Shuzo) to obtain a series of deleted cDNA clones, or by using an oligoprimer specific for the internal sequence of pSHGT8.

(4) Comparison of the Nucleotide Sequence and the Amino Acid Sequence

The cDNA inserted into pSHGT8 had the length of 2062 bp, and included an open reading frame (ORF) consisting of 1386 bp in length (including a stop codon). This sequence is shown in SEQ ID NO: 3. The amino acid sequence of this ORF had homology of 68% with the amino acid sequence of Perilla 5GT encoded by p3R4, and homology of 64% with that encoded by p3R6. In addition, it also had homology of 22 to 25% with the 3GTs of monocotyledonous and dicotyledoneous plants, and homology of 21% with petunia 3RT.

(5) Expression in Yeast and Measurement of Enzymatic Activity

An approximately 2.0 kb DNA fragment obtained by digesting pSHGT8 with BamHI/XhoI, and an approximately 8 kb DNA fragment obtained by digesting pYE22m with BamHI/SalI were ligated, and the resulting plasmid was designated as pYHGT8. pYHGT8 was expressed in yeast cells in the same manner as Example 1, and the enzymatic activity of the protein encoded by pSHGT8 was measured. As a result, in the reaction mixture containing the crude enzyme solution of yeast transformed with pYHGT8, a product was obtained that coincided with cyanidin-3,5-

diglucoside in both retention time and absorption spectrum. Based on this observation, the pSHGT8 cDNA of Verbena was determined to code for 5GT.

Example 3 Cloning of Trenia 5GT Nucleic Acid Molecule

(1). Preparation of cDNA Library

Petals were collected from torenia variety Summer Wave Blue (Suntory) and ground in a mortar and pestle in liquid nitrogen. RNA was extracted from the ground tissues according to a method using guanidine thiocyanate/cesium chloride, and poly A + RNA was obtained by the method recommended by the manufacturer using Oligotex (Takara Shuzo). The method using guanidine thiocyanate/cesium chloride was carried out in accordance with the method described in detail in Methods in Molecular Biology, Vol. 2 (Humana Press Inc., 1984) by R. McGookin and Robert J. Slater, et al.

Using the resulting poly A + RNA as a template, double-strand cDNA was synthesized using the ZAP-cDNA synthesis kit of Strategene, then, a cDNA library was prepared using the Uni-ZAP XR Cloning Kit (Stratagene) according to the method recommended by the manufacturer.

(2) Cloning of 5GT cDNA

The λ phage library obtained as described above was screened in the same manner as Example 2 using the p3R4 cDNA of Perilla as a probe. As a result, eight positive clones were obtained. After excision of the cDNA on plasmid pBluescript SK, the lengths of the cDNA were investigated by agarose gel electrophoresis, which revealed that a maximum length of insertion was 1.6 kb.

(3) Determination of Nucleotide Sequence

Plasmids were extracted from the resulting clones, and the nucleotide sequences near both 5' and 3' ends were determined in the same manner as Example 2. As a result, six of the eight clones were considered to have mutually same nucleotide sequences although the lengths of the cDNA were different. The entire nucleotide sequence of pSTGT5 cDNA was determined.

(4) Comparison of the Nucleotide Sequence and the Amino Acid Sequence

The cDNA encoded in pSTGT5 was of 1671 bp in length, and included an open reading frame (ORF) consisting of 1437 bp in length (including a stop codon). This sequence is shown in SEQ ID NO: 4. The amino acid sequence of this ORF had homology of 58% with the amino acid sequence of Perilla 5GT encoded by p3R4, and homology of 57% with that encoded by p3R6, and, homology of 57% with that encoded by Verbena pSHGT8. In addition, it also had homology of 19 to 23% with the 3GT of monocotyledonous and dicotyledoneous plants, and homology of 20% with petunia 3RT.

(5) Expression of 5GT nucleic acid molecule

An approximately 1.6 kb DNA fragment obtained by digesting pSTGT5 with SmaI/KpnI, and an approximately 8 kb DNA fragment obtained by blunting the EcoRI-digested site of pYE22m and then digesting with KpnI were ligated, and the resulting plasmid was designated as pYTGT5. pYTGT5 was expressed in yeast cells in the same manner as Example 1, and the enzymatic activity of the protein encoded by pSTGT5 was measured. As a result, in the reaction mixture containing the crude enzyme solution of yeast transformed with pYTGT5, a product was obtained that coincided with cyanidin-3,5-diglucoside in both retention time and absorption spectrum. Based on this observation, the pSTGT5 cDNA of Torenia was determined to code for 5GT.

Example 4 Cloning of Petunia 5GT Nucleic Acid Molecule

(1) Preparation of cDNA Library

A cDNA library was prepared by RNA extracted from petals of the Petunia variety Old Glory Blue in the manner described in detail by T. Holton, et al. (Plant Journal, 1993 4: 1003-1010)

(2) Cloning of 5GT cDNA

The cDNA library was screened in the same manner as Example 2 using the mixture of 5GT cDNAs of Perilla, torenia and verbena obtained in the manner described above as probes. As a result, four positive cDNA clones were obtained and excised on plasmid pBluescript SK. The lengths of the cDNA were investigated by agarose gel electrophoresis, cDNA of a maximum length of 2.0 kb was observed.

(3) Determination of the Nucleotide Sequence

Plasmids were extracted from the resulting clones, and the nucleotide sequence near the 5' end was determined in the same manner as Example 2. As a result, two of the four clones, pSPGT1, were appeared to code an amino acid sequence with a high degree of homology with those of 5GT from Perilla, torenia and verbena obtained thus far. Therefore, the entire nucleotide sequence of pSPGT1 was determined.

(4) Comparison of the Nucleotide Sequence and the Amino Acid Sequence

The pSPGT1 cDNA was 2015 bp in length, and included an open reading frame (ORF) consisting of 1407 bp (including a stop codon). This sequence is shown in SEQ ID NO: 6. The amino acid sequence of this ORF had homology of 57% with that of 5GT encoded by p3R4 of Perilla, homology of 54% with that encoded by p3R6, 55% with that encoded by pSHGT8 of verbena, and 51% of that encoded by pTGT5 of torenia. In addition, it also had homology of 20 to 29% with the 3GT of monocotyledonous and dicotyledoneous plants, and homology of 20% with petunia 3RT. Based on this observation, pSPGT1 cDNA obtained from petunia is considered to code for 5GT.

Industrial Applicability

As has been described above, cDNA coding for enzymes that transfer a glycoside to the 5 position of a flavonoid originating in Perilla, verbena, torenia and petunia were cloned and their nucleotide sequences were determined. In addition, the isolated cDNAs were clearly shown to code for 5GT by the enzymatic activity of their protein expressed in yeast. Introducing of these cDNAs into a suitable plant expression vector and transferring the resulting expression constructs into a plant makes it possible to provide, increase or decrease 5GT activity in the transformed plant, which leads to regulation of flower color. In addition, by using this enzyme, the structure of anthocyans can be altered or more stable anthocyans can be synthesized either in plants or in vitro.

EDITORIAL NOTE-NO.82432/98

This specification contains a sequence listing pages 24-41 following the description and claim pages follows starting from page no 22 to 23

Sequence

Sequence ID No.: 1

Sequence length: 1507

Sequence type: Nucleic acid

Number of strands: Double-strand

Topology: Straight chain

Source:

Biological name: Perilla (Perilla frutescens)

Tissue type: Leaf

Direct source:

Library name: cDNA library

Clone name: p3R4

Sequence:

GAAAATTTCC ACAAAA ATG GTC CGC CGC CGC GTG CTG CTA GCA ACG TTT 49
Met Val Arg Arg Arg Val Leu Leu Ala Thr Phe
1 5 10

CCT GCG CAA GGC CAC ATA AAT CCC GCC CTC CAA TTC GCC AAG AGA CTC 97
Pro Ala Gln Gly His Ile Asn Pro Ala Leu Gln Phe Ala Lys Arg Leu
15 20 25

CTA AAA GCC GGC ACT GAC GTC ACA TTT TTC ACG AGC GTT TAT GCA TGG 145
Leu Lys Ala Gly Thr Asp Val Thr Phe Phe Thr Ser Val Tyr Ala Trp
30 35 40

CGC CGC ATG GCC AAC ACA GCC TCC GCC GCT GCC GGA AAC CCA CCG GGC 193
Arg Arg Met Ala Asn Thr Ala Ser Ala Ala Ala Gly Asn Pro Pro Gly
45 50 55

CTC GAC TTC GTG GCG TTC TCC GAC GGC TAC GAC GAC GGG CTG AAG CCC 241
Leu Asp Phe Val Ala Phe Ser Asp Gly Tyr Asp Asp Gly Leu Lys Pro
60 65 70 75

TGC GGC GAC GGG AAG CGC TAC ATG TCC GAG ATG AAA GCC CGC GGC TCC 289
Cys Gly Asp Gly Lys Arg Tyr Met Ser Glu Met Lys Ala Arg Gly Ser
80 85 90

GAG GCC TTA AGA AAC CTC CTT CTC AAC AAC CAC GAC GTC ACG TTC GTC 337
Glu Ala Leu Arg Asn Leu Leu Leu Asn Asn His Asp Val Thr Phe Val
95 100 105

```
GTC TAC TCC CAC CTC TTT GCA TGG GCG GCG GAG GTG GCG CGT GAG TCC     385
Val Tyr Ser His Leu Phe Ala Trp Ala Ala Glu Val Ala Arg Glu Ser
        110                 115                 120
CAG GTC CCG AGC GCC CTT CTC TGG GTC GAG CCC GCC ACC GTG CTG TGC     433
Gln Val Pro Ser Ala Leu Leu Trp Val Glu Pro Ala Thr Val Leu Cys
    125                 130                 135
ATA TAT TAC TTC TAC TTC AAC GGC TAC GCA GAC GAG ATC GAC GCC GGT     481
Ile Tyr Tyr Phe Tyr Phe Asn Gly Tyr Ala Asp Glu Ile Asp Ala Gly
140                 145                 150                 155
TCC GAC GAA ATT CAG CTC CCT CGG CTT CCA CCC CTG GAG CAG CGC AGT     529
Ser Asp Glu Ile Gln Leu Pro Arg Leu Pro Pro Leu Glu Gln Arg Ser
                160                 165                 170
CTT CCG ACC TTT CTG CTG CCG GAG ACA CCG GAG AGA TTC CGG TTG ATG     577
Leu Pro Thr Phe Leu Leu Pro Glu Thr Pro Glu Arg Phe Arg Leu Met
            175                 180                 185
ATG AAG GAG AAG CTG GAA ACT TTA GAC GGT GAA GAG AAG GCG AAA GTG     625
Met Lys Glu Lys Leu Glu Thr Leu Asp Gly Glu Glu Lys Ala Lys Val
        190                 195                 200
TTG GTG AAC ACG TTT GAT GCG TTG GAG CCC GAT GCA CTC ACG GCT ATT     673
Leu Val Asn Thr Phe Asp Ala Leu Glu Pro Asp Ala Leu Thr Ala Ile
    205                 210                 215
GAT AGG TAT GAG TTG ATC GGG ATC GGG CCG TTG ATT CCC TCC GCC TTC     721
Asp Arg Tyr Glu Leu Ile Gly Ile Gly Pro Leu Ile Pro Ser Ala Phe
220                 225                 230                 235
TTG GAC GGC GGA GAT CCC TCC GAA ACG TCT TAC GGC GGC GAT CTT TTC     769
Leu Asp Gly Gly Asp Pro Ser Glu Thr Ser Tyr Gly Gly Asp Leu Phe
                240                 245                 250
GAA AAA TCG GAG GAG AAT AAC TGC GTG GAG TGG TTG GAC ACG AAG CCG     817
Glu Lys Ser Glu Glu Asn Asn Cys Val Glu Trp Leu Asp Thr Lys Pro
            255                 260                 265
AAA TCT TCG GTG GTG TAT GTG TCG TTT GGG AGC GTT TTG AGG TTT CCA     865
Lys Ser Ser Val Val Tyr Val Ser Phe Gly Ser Val Leu Arg Phe Pro
        270                 275                 280
AAG GCA CAA ATG GAA GAG ATT GGG AAA GGG CTA TTA GCC TGC GGA AGG     913
Lys Ala Gln Met Glu Glu Ile Gly Lys Gly Leu Leu Ala Cys Gly Arg
    285                 290                 295
```

```
CCG TTT TTA TGG ATG ATA CGA GAA CAG AAG AAT GAC GAC GGC GAA GAA    961
Pro Phe Leu Trp Met Ile Arg Glu Gln Lys Asn Asp Asp Gly Glu Glu
300                 305                 310                 315
GAA GAA GAA GAG TTG AGT TGC ATT GGG GAA TTG AAA AAA ATG GGG AAA   1009
Glu Glu Glu Glu Leu Ser Cys Ile Gly Glu Leu Lys Lys Met Gly Lys
                320                 325                 330
ATA GTT TCG TGG TGC TCG CAG TTG GAG GTT CTG GCG CAC CCT GCG TTG   1057
Ile Val Ser Trp Cys Ser Gln Leu Glu Val Leu Ala His Pro Ala Leu
            335                 340                 345
GGA TGT TTC GTG ACG CAT TGT GGG TGG AAC TCG GCT GTG GAG AGC TTG   1105
Gly Cys Phe Val Thr His Cys Gly Trp Asn Ser Ala Val Glu Ser Leu
        350                 355                 360
AGT TGC GGG GTT CCG GTG GTG GCG GTG CCG CAG TGG TTT GAT CAG ACG   1153
Ser Cys Gly Val Pro Val Val Ala Val Pro Gln Trp Phe Asp Gln Thr
    365                 370                 375
ACG AAT GCG AAG CTG ATT GAG GAT GCG TGG GGG ACA GGG GTG AGA GTG   1201
Thr Asn Ala Lys Leu Ile Glu Asp Ala Trp Gly Thr Gly Val Arg Val
380                 385                 390                 395
AGA ATG AAT GAA GGG GGT GGG GTT GAT GGA TCT GAG ATA GAG AGG TGT   1249
Arg Met Asn Glu Gly Gly Gly Val Asp Gly Ser Glu Ile Glu Arg Cys
                400                 405                 410
GTG GAG ATG GTG ATG GAT GGG GGT GAG AAG AGC AAA CTA GTG AGA GAA   1297
Val Glu Met Val Met Asp Gly Gly Glu Lys Ser Lys Leu Val Arg Glu
            415                 420                 425
AAT GCC ATA AAA TGG AAG ACT TTG GCC AGA GAA GCC ATG GGA GAG GAT   1345
Asn Ala Ile Lys Trp Lys Thr Leu Ala Arg Glu Ala Met Gly Glu Asp
        430                 435                 440
GGA TCT TCA CTC AAG AAT CTC AAC GCC TTT CTT CAT CAA GTT GCA CGT   1393
Gly Ser Ser Leu Lys Asn Leu Asn Ala Phe Leu His Gln Val Ala Arg
    445                 450                 455
GCT TAATACACAA AATGGCTTTC CACTTTTAAT CTACTCAAAC ACCGGTTCAA        1446
Ala
460                                                               1507
ATAAATATCC CCTTCCACTT CTTTCTATTT CACTATCACA TTTATAATTT TAGTAACAAA 1506
A
```

Sequence ID No.: 2

Sequence length: 1470

Sequence type: Nucleic acid

Number of strands: Double-strand

Topology: Straight chain

Source:

Biological name: Perilla (Perilla frutescens)

Tissue type: Leaf

Direct source:

Library name: cDNA library

Clone name: p3R6

Sequence:

```
ACCAAACCAA AACAAAATTT CCACAAAA ATG GTC CGC CGC CGC GTG CTG CTA      48
                              Met Val Arg Arg Arg Val Leu Leu
                              1               5
GCA ACG TTT CCG GCG CAA GGC CAC ATA AAT CCC GCC CTC CAA TTC GCC     96
Ala Thr Phe Pro Ala Gln Gly His Ile Asn Pro Ala Leu Gln Phe Ala
    10                  15                  20
AAG AGA CTC CTA AAA GCC GGC ACT GAC GTC ACG TTT TTC ACG AGC GTT    144
Lys Arg Leu Leu Lys Ala Gly Thr Asp Val Thr Phe Phe Thr Ser Val
25                  30                  35                  40
TAT GCA TGG CGC CGC ATG GCC AAC ACA GCC TCC GCC GCT GCC GGA AAC    192
Tyr Ala Trp Arg Arg Met Ala Asn Thr Ala Ser Ala Ala Ala Gly Asn
                45                  50                  55
CCA CCG GGC CTC GAC TTC GTG GCG TTC TCC GAC GGC TAC GAC GAC GGG    240
Pro Pro Gly Leu Asp Phe Val Ala Phe Ser Asp Gly Tyr Asp Asp Gly
            60                  65                  70
CTG AAG CCC GGC GGC GAC GGG AAG CGC TAC ATG TCC GAG ATG AAA GCC    288
Leu Lys Pro Gly Gly Asp Gly Lys Arg Tyr Met Ser Glu Met Lys Ala
        75                  80                  85
CGC GGC TCC GAG GCC TTA AGA AAC CTC CTT CTC AAC AAC GAC GAC GTC    336
Arg Gly Ser Glu Ala Leu Arg Asn Leu Leu Leu Asn Asn Asp Asp Val
    90                  95                  100
ACT TTC GTC GTC TAC TCC CAC CTC TTT GCA TGG GCG GCG GAG GTG GCG    384
Thr Phe Val Val Tyr Ser His Leu Phe Ala Trp Ala Ala Glu Val Ala
105                 110                 115                 120
```

```
CGT TTG TCC CAC GTC CCG ACC GCC CTT CTC TGG GTC GAG CCC GCC ACC      432
Arg Leu Ser His Val Pro Thr Ala Leu Leu Trp Val Glu Pro Ala Thr
                125                 130                 135
GTG CTG TGC ATA TAC CAC TTC TAC TTC AAC GGC TAC GCA GAC GAG ATC      480
Val Leu Cys Ile Tyr His Phe Tyr Phe Asn Gly Tyr Ala Asp Glu Ile
            140                 145                 150
GAC GCC GGT TCC AAT GAA ATT CAG CTC CCT CGG CTT CCA TCC CTG GAG      528
Asp Ala Gly Ser Asn Glu Ile Gln Leu Pro Arg Leu Pro Ser Leu Glu
        155                 160                 165
CAG CGC AGT CTT CCG ACG TTT CTG CTG CCT GCG ACG CCG GAG AGA TTC      576
Gln Arg Ser Leu Pro Thr Phe Leu Leu Pro Ala Thr Pro Glu Arg Phe
    170                 175                 180
CGG TTG ATG ATG AAG GAG AAG CTG GAA ACT TTA GAC GGT GAA GAG AAG      624
Arg Leu Met Met Lys Glu Lys Leu Glu Thr Leu Asp Gly Glu Glu Lys
185                 190                 195                 200
GCG AAA GTA TTG GTG AAC ACG TTT GAT GCG TTG GAG CCC GAT GCA CTC      672
Ala Lys Val Leu Val Asn Thr Phe Asp Ala Leu Glu Pro Asp Ala Leu
                205                 210                 215
ACG GCT ATT GAT AGG TAT GAG TTG ATC GGG ATC GGG CCG TTG ATT CCC      720
Thr Ala Ile Asp Arg Tyr Glu Leu Ile Gly Ile Gly Pro Leu Ile Pro
            220                 225                 230
TCC GCC TTC TTG GAC GGC GAA GAT CCC TCC GAA ACG TCT TAC GGC GGC      768
Ser Ala Phe Leu Asp Gly Glu Asp Pro Ser Glu Thr Ser Tyr Gly Gly
        235                 240                 245
GAT CTT TTC GAA AAA TCG GAG GAG AAT AAC TGC GTG GAG TGG TTG AAC      816
Asp Leu Phe Glu Lys Ser Glu Glu Asn Asn Cys Val Glu Trp Leu Asn
    250                 255                 260
TCG AAG CCG AAA TCT TCG GTG GTG TAT GTG TCG TTT GGG AGC GTT TTG      864
Ser Lys Pro Lys Ser Ser Val Val Tyr Val Ser Phe Gly Ser Val Leu
265                 270                 275                 280
AGG TTT CCA AAG GCA CAA ATG GAA GAG ATT GGG AAA GGG CTA TTA GCC      912
Arg Phe Pro Lys Ala Gln Met Glu Glu Ile Gly Lys Gly Leu Leu Ala
                285                 290                 295
TGC GGA AGG CCC TTT TTA TGG ATG ATA CGA GAA CAG AAG AAT GAC GAC      960
Cys Gly Arg Pro Phe Leu Trp Met Ile Arg Glu Gln Lys Asn Asp Asp
            300                 305                 310
```

```
GGC GAA GAA GAA GAA GAA GAA GAA GAG TTG AGT TGC ATT GGG GAA TTG   1008
Gly Glu Glu Glu Glu Glu Glu Glu Glu Leu Ser Cys Ile Gly Glu Leu
        315                 320                 325
AAA AAA ATG GGG AAA ATA GTG TCG TGG TGC TCG CAG TTG GAG GTT CTG   1056
Lys Lys Met Gly Lys Ile Val Ser Trp Cys Ser Gln Leu Glu Val Leu
    330                 335                 340
GCG CAC CCT GCG TTG GGA TGT TTC GTG ACG CAT TGT GGG TGG AAC TCG   1104
Ala His Pro Ala Leu Gly Cys Phe Val Thr His Cys Gly Trp Asn Ser
345                 350                 355                 360
GCT GTG GAG AGC TTG AGT TGC GGG ATT CCG GTG GTG GCG GTG CCG CAG   1152
Ala Val Glu Ser Leu Ser Cys Gly Ile Pro Val Val Ala Val Pro Gln
                365                 370                 375
TGG TTT GAT CAG ACG ACG AAT GCG AAG CTG ATT GAG GAT GCG TGG GGG   1200
Trp Phe Asp Gln Thr Thr Asn Ala Lys Leu Ile Glu Asp Ala Trp Gly
            380                 385                 390
ACA GGG GTG AGA GTG AGA ATG AAT GAA GGG GGT GGG GTT GAT GGA TGT   1248
Thr Gly Val Arg Val Arg Met Asn Glu Gly Gly Gly Val Asp Gly Cys
        395                 400                 405
GAG ATA GAA AGG TGT GTG GAG ATG GTG ATG GAT GGG GGT GAC AAG ACC   1296
Glu Ile Glu Arg Cys Val Glu Met Val Met Asp Gly Gly Asp Lys Thr
    410                 415                 420
AAA CTA GTG AGA GAA AAT GCC ATC AAA TGG AAG ACT TTG GCC AGA CAA   1344
Lys Leu Val Arg Glu Asn Ala Ile Lys Trp Lys Thr Leu Ala Arg Gln
425                 430                 435                 440
GCC ATG GGA TAGGATGGAT CTTCACTCAA CAATCTCAAC GCCTTTCTTC           1393
Ala Met Gly
        443
GTCAAGTTGC ACACTTTTAA TCTGCTCAAA CAGCGGTTCA AATAAATATC CCCTTCCACT 1453
TAAAAAAAAA AAAAAAA                                                1470
```

Sequence ID No.: 3

Sequence length: 2062

Sequence type: Nucleic acid

Number of strands: Double-strand

Topology: Straight chain

Source:

Biological name: Verbena (Verbena hybrida)

Tissue type: Petal

Direct source:

Library name: cDNA library

Clone name: pSHGT8

Sequence:

```
ATTTTACCAA AAAAATAAAA AAAAA ATG AGC AGA GCT CAC GTC CTC TTG GCC     52
                            Met Ser Arg Ala His Val Leu Leu Ala
                             1               5

ACA TTC CCA GCA CAG GGA CAC ATA AAT CCC GCC CTT CAA TTC GCC AAG    100
Thr Phe Pro Ala Gln Gly His Ile Asn Pro Ala Leu Gln Phe Ala Lys
 10                  15                  20                  25

CGT CTC GCA AAT GCC GAC ATT CAA GTC ACA TTC TTC ACC AGC GTC TAC    148
Arg Leu Ala Asn Ala Asp Ile Gln Val Thr Phe Phe Thr Ser Val Tyr
                 30                  35                  40

GCA TGG CGC CGC ATG TCC AGA ACC GCC GCT GGC TCA AAC GGG CTC ATC    196
Ala Trp Arg Arg Met Ser Arg Thr Ala Ala Gly Ser Asn Gly Leu Ile
             45                  50                  55

AAT TTT GTG TCG TTT TCC GAC GGG TAT GAC GAC GGG TTA CAG CCC GGA    244
Asn Phe Val Ser Phe Ser Asp Gly Tyr Asp Asp Gly Leu Gln Pro Gly
         60                  65                  70

GAC GAT GGG AAG AAC TAC ATG TCG GAG ATG AAA AGC AGA GGT ATA AAA    292
Asp Asp Gly Lys Asn Tyr Met Ser Glu Met Lys Ser Arg Gly Ile Lys
     75                  80                  85

GCC TTG AGC GAT ACT CTT GCA GCC AAT AAT GTC GAT CAA AAA AGC AGC    340
Ala Leu Ser Asp Thr Leu Ala Ala Asn Asn Val Asp Gln Lys Ser Ser
90                   95                  100                 105

AAA ATC ACG TTC GTG GTG TAC TCC CAC CTC TTT GCA TGG GCG GCC AAG    388
Lys Ile Thr Phe Val Val Tyr Ser His Leu Phe Ala Trp Ala Ala Lys
                110                 115                 120

GTG GCG CGT GAG TTC CAT CTC CGG AGC GCG CTA CTC TGG ATT GAG CCA    436
Val Ala Arg Glu Phe His Leu Arg Ser Ala Leu Leu Trp Ile Glu Pro
            125                 130                 135

GCT ACG GTG TTG GAT ATA TTT TAC TTT TAT TTC AAC GGC TAT AGC GAC    484
Ala Thr Val Leu Asp Ile Phe Tyr Phe Tyr Phe Asn Gly Tyr Ser Asp
        140                 145                 150
```

```
GAA ATC GAT GCG GGT TCG GAT GCT ATT CAC TTG CCC GGA GGA CTC CCA    532
Glu Ile Asp Ala Gly Ser Asp Ala Ile His Leu Pro Gly Gly Leu Pro
    155                 160                 165
GTG CTG GCC CAG CGT GAT TTA CCG TCT TTC CTT CTT CCT TCC ACG CAT    580
Val Leu Ala Gln Arg Asp Leu Pro Ser Phe Leu Leu Pro Ser Thr His
170                 175                 180                 185
GAG AGA TTC CGT TCA CTG ATG AAG GAG AAA TTG GAA ACT TTA GAA GGT    628
Glu Arg Phe Arg Ser Leu Met Lys Glu Lys Leu Glu Thr Leu Glu Gly
                190                 195                 200
GAA GAA AAA CCT AAG GTC TTG GTG AAC AGC TTT GAT GCG TTG GAG CCT    676
Glu Glu Lys Pro Lys Val Leu Val Asn Ser Phe Asp Ala Leu Glu Pro
            205                 210                 215
GAT GCG CTC AAG GCC ATT GAT AAG TAC GAG ATG ATT GCA ATC GGG CCG    724
Asp Ala Leu Lys Ala Ile Asp Lys Tyr Glu Met Ile Ala Ile Gly Pro
        220                 225                 230
TTG ATT CCT TCC GCA TTC TTG GAC GGT AAA GAT CCT TCG GAC AGG TCT    772
Leu Ile Pro Ser Ala Phe Leu Asp Gly Lys Asp Pro Ser Asp Arg Ser
    235                 240                 245
TTC GGC GGA GAT TTG TTC GAG AAA GGG TCG AAT GAC GAC GAT TGC CTC    820
Phe Gly Gly Asp Leu Phe Glu Lys Gly Ser Asn Asp Asp Asp Cys Leu
250                 255                 260                 265
GAA TGG TTG AGC ACG AAT CCT CGA TCT TCG GTG GTT TAC GTT TCG TTC    868
Glu Trp Leu Ser Thr Asn Pro Arg Ser Ser Val Val Tyr Val Ser Phe
                270                 275                 280
GGA AGC TTC GTT AAT ACG ACG AAG TCG CAA ATG GAA GAG ATA GCA AGA    916
Gly Ser Phe Val Asn Thr Thr Lys Ser Gln Met Glu Glu Ile Ala Arg
            285                 290                 295
GGG CTG TTA GAT TGT GGG AGG CCG TTT TTG TGG GTG GTA AGA GTA AAC    964
Gly Leu Leu Asp Cys Gly Arg Pro Phe Leu Trp Val Val Arg Val Asn
        300                 305                 310
GAA GGA GAA GAG GTA TTG ATA AGT TGC ATG GAG GAG TTG AAA CGA GTG   1012
Glu Gly Glu Glu Val Leu Ile Ser Cys Met Glu Glu Leu Lys Arg Val
    315                 320                 325
GGG AAA ATT GTA TCT TGG TGT TCT CAA TTG GAA GTC CTG ACG CAT CCC   1060
Gly Lys Ile Val Ser Trp Cys Ser Gln Leu Glu Val Leu Thr His Pro
330                 335                 340                 345
```

```
TCG TTG GGA TGT TTC GTG ACA CAC TGC GGG TGG AAT TCG ACT CTA GAG    1108
Ser Leu Gly Cys Phe Val Thr His Cys Gly Trp Asn Ser Thr Leu Glu
                350                 355                 360
AGT ATA TCT TTC GGG GTT CCG ATG GTG GCT TTT CCG CAG TGG TTC GAT    1156
Ser Ile Ser Phe Gly Val Pro Met Val Ala Phe Pro Gln Trp Phe Asp
            365                 370                 375
CAA GGG ACG AAT GCG AAG CTG ATG GAG GAT GTG TGG AGG ACG GGT GTG    1204
Gln Gly Thr Asn Ala Lys Leu Met Glu Asp Val Trp Arg Thr Gly Val
        380                 385                 390
AGA GTG AGA GCT AAT GAG GAG GGT AGC GTC GTT GAT GGT GAT GAA ATT    1252
Arg Val Arg Ala Asn Glu Glu Gly Ser Val Val Asp Gly Asp Glu Ile
    395                 400                 405
AGG AGA TGT ATT GAG GAG GTT ATG GAT GGG GGA GAA AAG AGT AGG AAA    1300
Arg Arg Cys Ile Glu Glu Val Met Asp Gly Gly Glu Lys Ser Arg Lys
410                 415                 420             425
CTT AGA GAG AGT GCT GGC AAG TGG AAG GAT TTG GCA AGA AAA GCT ATG    1348
Leu Arg Glu Ser Ala Gly Lys Trp Lys Asp Leu Ala Arg Lys Ala Met
                430                 435                 440
GAG GAA GAT GGA TCT TCA GTT AAC AAC CTC AAG GTC TTT CTT GAT GAG    1396
Glu Glu Asp Gly Ser Ser Val Asn Asn Leu Lys Val Phe Leu Asp Glu
            445                 450                 455
GTT GTA GGT ATC TAAAGACGTA AATGAGGTCC CCATAGGCAA AATTGCAAAT       1448
Val Val Gly Ile
        460 461
TTCATCTCGT AAGTTGAATA CTTTTTGGCT TTAATTTTGT TCGAGTTTGT TTTTCAAAAT  1508
TTATCTTGTA ATTTTACATT GAGTGTAAAT TTAGTCTGAT TTTAACTGGA AAAATATAAA  1568
ATTCATTGTT GAGACTCTTC ATCAAAATCA TCTGATTTCC TTTATTGTCT TGGTCAAAAT  1628
TCTCATATCA ATTGGAAAAA ATAAATTTCA AAATCGTCCA ATTTTGAACC AAGAAAGAAG  1688
TATAATTTGA CCAAAATAAT AAAAGGATTC AAGTGATCTT GATGAAGTGT CTGAGCGACG  1748
AGTTCTATAT TTTTCCACCG AATTTCTAAC GAGTTTTTGA ATTTTTTTTA GCCAAAATCG  1808
GACTAACTTT GTACAAAATG AAAAGTTATA TGATGAAATT TTAAAAAACA AACTCAGACA  1868
ATAATAAAGC CCGAAAGTAG TAAAATTACC TGACGAAATT TGCAATTTCG CCTCCTATTT  1928
TAATTTTTTT GGTGTGTTTA ATAAATCGGT TATTTTACTT TTAATTAAAA TAAAAGTGAG  1988
ATGCATGATA GCTTGGTGAG TATATATGAG TTGATGGTAA TGTACGATAT TTTCTAAAAA  2048
AAAAAAAAAA AAAA                                                    2062
```

Sequence ID No.: 4

Sequence length: 1671

Sequence type: Nucleic acid

Number of strands: Double-strand

Topology: Straight chain

Source:

Biological name: Trenia

Tissue type: Petal

Direct source:

Library name: cDNA library

Clone name: pSTGT5

Sequence:

```
AACACATAAA AAAAAAATAA AAGAAGAAAT AATTAAAAAA AAAA ATG GTT AAC          53
                                                 Met Val Asn
                                                  1
AAA CGC CAT ATT CTA CTA GCA ACA TTC CCA GCA CAA GGC CAC ATA AAC      101
Lys Arg His Ile Leu Leu Ala Thr Phe Pro Ala Gln Gly His Ile Asn
     5                  10                  15
CCT TCT CTC GAG TTC GCC AAA AGG CTC CTC AAC ACC GGA TAC GTC GAC      149
Pro Ser Leu Glu Phe Ala Lys Arg Leu Leu Asn Thr Gly Tyr Val Asp
20                  25                  30                  35
CAA GTC ACA TTC TTC ACG AGT GTA TAC GCA TTG AGA CGC ATG CGC TTC      197
Gln Val Thr Phe Phe Thr Ser Val Tyr Ala Leu Arg Arg Met Arg Phe
                40                  45                  50
GAA ACC GAT CCG AGC AGC AGA ATC GAT TTC GTG GCA TKT YCA GAT TCT      245
Glu Thr Asp Pro Ser Ser Arg Ile Asp Phe Val Ala Xaa Xaa Asp Ser
            55                  60                  65
TAC GAT GAT GGC TTA AAG AAA GGC GAC GAT GGC AAA AAC TAC ATG TCG      293
Tyr Asp Asp Gly Leu Lys Lys Gly Asp Asp Gly Lys Asn Tyr Met Ser
        70                  75                  80
GAG ATG AGA AAG CGC GGA ACG AAG GCC TTA AAG GAC ACT CTT ATT AAG      341
Glu Met Arg Lys Arg Gly Thr Lys Ala Leu Lys Asp Thr Leu Ile Lys
    85                  90                  95
CTC AAC GAT GCT GCG ATG GGA AGT GAA TGT TAC AAT CGC GTG AGC TTT      389
Leu Asn Asp Ala Ala Met Gly Ser Glu Cys Tyr Asn Arg Val Ser Phe
100                 105                 110                 115
```

```
GTG GTG TAC TCT CAT CTA TTT TCG TGG GCA GCT GAA GTG GCG CGT GAA    437
Val Val Tyr Ser His Leu Phe Ser Trp Ala Ala Glu Val Ala Arg Glu
                120                 125                 130
GTC GAC GTG CCG AGT GCC CTT CTT TGG ATT GAA CCG GCT ACG GTT TTC    485
Val Asp Val Pro Ser Ala Leu Leu Trp Ile Glu Pro Ala Thr Val Phe
            135                 140                 145
GAT GTG TAC TAT TTT TAC TTC AAT GGG TAT GCC GAT GAT ATC GAT GCG    533
Asp Val Tyr Tyr Phe Tyr Phe Asn Gly Tyr Ala Asp Asp Ile Asp Ala
        150                 155                 160
GGC TCA GAT CAA ATC CAA CTG CCC AAT CTT CCG CAG CTC TCC AAG CAA    581
Gly Ser Asp Gln Ile Gln Leu Pro Asn Leu Pro Gln Leu Ser Lys Gln
    165                 170                 175
GAT CTC CCC TCT TTC CTA CTC CCT TCG AGC CCC GCG AGA TTC CGA ACC    629
Asp Leu Pro Ser Phe Leu Leu Pro Ser Ser Pro Ala Arg Phe Arg Thr
180                 185                 190                 195
CTA ATG AAA GAA AAG TTC GAC ACG CTC GAC AAA GAA CCG AAA GCG AAG    677
Leu Met Lys Glu Lys Phe Asp Thr Leu Asp Lys Glu Pro Lys Ala Lys
                200                 205                 210
GTC TTG ATA AAC ACG TTC GAC GCA TTA GAA ACC GAA CAA CTC AAA GCC    725
Val Leu Ile Asn Thr Phe Asp Ala Leu Glu Thr Glu Gln Leu Lys Ala
            215                 220                 225
ATC GAC AGG TAT GAA CTA ATA TCC ATC GGC CCA TTA ATC CCA TCA TCG    773
Ile Asp Arg Tyr Glu Leu Ile Ser Ile Gly Pro Leu Ile Pro Ser Ser
        230                 235                 240
ATA TTC TCA GAT GGC AAC GAC CCC TCA TCA AGC AAC AAA TCC TAC GGT    821
Ile Phe Ser Asp Gly Asn Asp Pro Ser Ser Ser Asn Lys Ser Tyr Gly
    245                 250                 255
GGA GAC CTC TTC AGA AAA GCC GAT GAA ACT TAC ATG GAC TGG CTA AAC    869
Gly Asp Leu Phe Arg Lys Ala Asp Glu Thr Tyr Met Asp Trp Leu Asn
260                 265                 270                 275
TCA AAA CCC GAA TCA TCG GTC GTT TAC GTT TCG TTC GGG AGC CTC CTG    917
Ser Lys Pro Glu Ser Ser Val Val Tyr Val Ser Phe Gly Ser Leu Leu
                280                 285                 290
AGG CTC CCG AAA CCC CAA ATG GAA GAA ATA GCA ATA GGG CTT TCA GAC    965
Arg Leu Pro Lys Pro Gln Met Glu Glu Ile Ala Ile Gly Leu Ser Asp
            295                 300                 305
```

```
ACC AAA TCG CCA GTT CTC TGG GTG ATA AGA AGA AAC GAA GAG GGC GAC    1013
Thr Lys Ser Pro Val Leu Trp Val Ile Arg Arg Asn Glu Glu Gly Asp
        310                 315                 320
GAA CAA GAG CAA GCA GAA GAA GAA GAG AAG CTG CTG AGC TTC TTT GAT    1061
Glu Gln Glu Gln Ala Glu Glu Glu Glu Lys Leu Leu Ser Phe Phe Asp
    325                 330                 335
CGT CAC GGA ACT GAA CGA CTC GGG AAA ATC GTG ACA TGG TGC TCA CAA    1109
Arg His Gly Thr Glu Arg Leu Gly Lys Ile Val Thr Trp Cys Ser Gln
340                 345                 350                 355
TTG GAT GTT CTG ACG CAT AAG TCG GTG GGA TGC TTC GTG ACG CAT TGC    1157
Leu Asp Val Leu Thr His Lys Ser Val Gly Cys Phe Val Thr His Cys
                360                 365                 370
GGT TGG AAT TCT GCT ATC GAG AGC CTG GCT TGT GGT GTG CCC GTG GTG    1205
Gly Trp Asn Ser Ala Ile Glu Ser Leu Ala Cys Gly Val Pro Val Val
            375                 380                 385
TGC TTT CCT CAA TGG TTC GAT CAA GGG ACT AAT GCG AAG ATG ATC GAA    1253
Cys Phe Pro Gln Trp Phe Asp Gln Gly Thr Asn Ala Lys Met Ile Glu
        390                 395                 400
GAT GTG TGG AGG AGT GGT GTG AGA GTC AGA GTG AAT GAG GAA GGC GGC    1301
Asp Val Trp Arg Ser Gly Val Arg Val Arg Val Asn Glu Glu Gly Gly
    405                 410                 415
GTT GTT GAT AGG CGT GAG ATT AAG AGG TGC GTC TCG GAG GTT ATA AAG    1349
Val Val Asp Arg Arg Glu Ile Lys Arg Cys Val Ser Glu Val Ile Lys
420                 425                 430                 435
AGT CGA GAG TTG AGA GAA AGC GCA ATG ATG TGG AAG GGT TTG GCT AAA    1397
Ser Arg Glu Leu Arg Glu Ser Ala Met Met Trp Lys Gly Leu Ala Lys
                440                 445                 450
GAA GCT ATG GAT GAA GAA CGT GGA TCA TCA ATG AAC AAT CTG AAG AAT    1445
Glu Ala Met Asp Glu Glu Arg Gly Ser Ser Met Asn Asn Leu Lys Asn
            455                 460                 465
TTT ATT ACT AGG ATT ATT AAT GAA AAT GCC TCA TAAGTTGTAC             1488
Phe Ile Thr Arg Ile Ile Asn Glu Asn Ala Ser
        470                 475         478
TATATATGTT ATTATTGTTG TTATGGACGT CGAATTAAGT ATTAGTTAAA TGATATGTAT  1548
TTAGAGGAAG GCCAAAACGG GCTACACCCG GCAGGCCACG GGTTGGAAAA GCCCGCCATG  1608
ATTTAAAATA TATATTTTAA AATAAATATT TTCTACTATT AAACTAAAAA AAAAAAAAAA  1668
AAA                                                                1671
```

Sequence ID No.: 5

Sequence length: 1437

Sequence type: Nucleic acid

Number of strands: Double-strand

Topology: Straight chain

Source:

Biological name: Perilla (Perilla frutescens)

Tissue type: Leaf

Direct source:

Library name: cDNA library

Clone name: p8R6

Sequence:

```
TTCAAAACTC ATAACGTGAT TGAGCTAATG TGCACATCTT CCTCTTCAAA GTCTACAGTG    60
TCATCCTACC AGCATCATCA TGATCAATCT CTTTATAATG AGGAGAATGG AGTAACAAGG   120
AGTGGGTTTT GTTACTCAGC TTCAACCTAC GTACGTACTA CTACTGACTC AACTCTCAAG   180
AGAATGAATA TAATATATAA TGGGCGATAG ATCTTTGTAG ATATGTAGGT GTAGCCTGCA   240
GGTGGTTAAT TAATTTCCGG TGTGGGAAAA TAAATAAATA AATAAATATA GCG ATG AGC   299
                                                           Met Ser
                                                            1
AGC AGC AGC AGC AGA AGG TGG AGA GAG AAT GAG GGG ATG CGA AGG ACA      347
Ser Ser Ser Ser Arg Arg Trp Arg Glu Asn Glu Gly Met Arg Arg Thr
        5                   10                  15
TTG CTG GGG TTG GGT TTG GGG CAG TTG GTT TCT TTC GAT TTG GCT ATC      395
Leu Leu Gly Leu Gly Leu Gly Gln Leu Val Ser Phe Asp Leu Ala Ile
    20                  25                  30
ATG ACC TTT TCT GCT TCT TTG GTT TCA ACC ACA GTG GAT GCA CCA CTT      443
Met Thr Phe Ser Ala Ser Leu Val Ser Thr Thr Val Asp Ala Pro Leu
35                  40                  45                  50
ACT ATG TCG TTC ACT ACA TAC ACT GTT GTG GCC CTG CTC TAT GGA ACC      491
Thr Met Ser Phe Thr Thr Tyr Thr Val Val Ala Leu Leu Tyr Gly Thr
                55                  60                  65
ATC TTG CTT TAC CGC CGC CAC AAA TTC TTG GTT CCA TGG TAC TGG TAT      539
Ile Leu Leu Tyr Arg Arg His Lys Phe Leu Val Pro Trp Tyr Trp Tyr
            70                  75                  80
```

```
GCT CTC CTG GGG TTC GTG GAC GTC CAC GGC AAT TAT CTT GTT AAT AAA    587
Ala Leu Leu Gly Phe Val Asp Val His Gly Asn Tyr Leu Val Asn Lys
        85                  90                  95
GCA TTC GAG TTG ACA TCG ATT ACG AGT GTG AGC ATA CTG GAT TGT TGG    635
Ala Phe Glu Leu Thr Ser Ile Thr Ser Val Ser Ile Leu Asp Cys Trp
    100                 105                 110
ACA ATC GTG TGG TCC ATC ATC TTT ACA TGG ATG TTC CTA GGC ACA AAA    683
Thr Ile Val Trp Ser Ile Ile Phe Thr Trp Met Phe Leu Gly Thr Lys
115                 120                 125                 130
TAC TCT GTA TAC CAG TTT GTC GGT GCT GCT ATT TGT GTA GGA GGC CTC    731
Tyr Ser Val Tyr Gln Phe Val Gly Ala Ala Ile Cys Val Gly Gly Leu
                135                 140                 145
CTC CTC GTG CTT CTT TCC GAC TCA GGG GTC ACT GCT GCT GGT TCG AAT    779
Leu Leu Val Leu Leu Ser Asp Ser Gly Val Thr Ala Ala Gly Ser Asn
            150                 155                 160
CCT CTT TTG GGT GAT TTT CTT GTC ATA ACA GGC TCT ATT TTG TTC ACA    827
Pro Leu Leu Gly Asp Phe Leu Val Ile Thr Gly Ser Ile Leu Phe Thr
        165                 170                 175
CTC AGC ACT GTT GGT CAG GAA TAC TGC GTG AAG AGG AAA GAT CGT ATT    875
Leu Ser Thr Val Gly Gln Glu Tyr Cys Val Lys Arg Lys Asp Arg Ile
    180                 185                 190
GAA GTA GTA GCA ATG ATC GGT GTA TTT GGT ATG CTC ATC AGT GCA ACC    923
Glu Val Val Ala Met Ile Gly Val Phe Gly Met Leu Ile Ser Ala Thr
195                 200                 205                 210
GAG ATT ACT GTG CTG GAG AGG AAT GCC CTC TCA TCA ATG CAG TGG TCT    971
Glu Ile Thr Val Leu Glu Arg Asn Ala Leu Ser Ser Met Gln Trp Ser
                215                 220                 225
ACT GGA CTT TTG GCA GCC TAT GTT GTT TAT GCA CTG TCC AGC TTC CTC   1019
Thr Gly Leu Leu Ala Ala Tyr Val Val Tyr Ala Leu Ser Ser Phe Leu
            230                 235                 240
TTC TGC ACA CTC ACC CCT TTT CTT CTC AAG ATG AGT GGC GCT GCA TTT   1067
Phe Cys Thr Leu Thr Pro Phe Leu Leu Lys Met Ser Gly Ala Ala Phe
        245                 250                 255
TTC AAT CTT TCC ATG CTT ACA TCT GAT ATG TGG GCT GTT GCA ATT AGG   1115
Phe Asn Leu Ser Met Leu Thr Ser Asp Met Trp Ala Val Ala Ile Arg
    260                 265                 270
```

```
ACA TTC ATA TAC AAC CAG GAG GTT GAT TGG TTA TAC TAT TTG GCC TTT    1163
Thr Phe Ile Tyr Asn Gln Glu Val Asp Trp Leu Tyr Tyr Leu Ala Phe
275                 280                 285                 290
TGT CTC GTT GTT GTT GGA ATA TTC ATA TAT ACA AAA ACA GAG AAG GAT    1211
Cys Leu Val Val Val Gly Ile Phe Ile Tyr Thr Lys Thr Glu Lys Asp
                295                 300                 305
CCT AAC AAT ACG AGA GCC CTT GAG AAT GGA AAC TTG GAT CAT GAA TAT    1259
Pro Asn Asn Thr Arg Ala Leu Glu Asn Gly Asn Leu Asp His Glu Tyr
            310                 315                 320
AGT CTC CTT GAG GAT CAA GAT GAC ACA CCA AGA AAA CCA TAGCTAGCTT     1308
Ser Leu Leu Glu Asp Gln Asp Asp Thr Pro Arg Lys Pro
        325                 330                 335
TGCCCACAAT CTTTTCATCA ACAGTTTTAA ATAATTCGTG AGGGGGAGAG AGATCGAGAT  1368
ACTAATTAAT GGACGTCTAT TATATAGTTG GAGGTTTTTG TTTTATTTAT TTATTTGAGT  1428
AAAAAAAAA                                                          1437
```

Sequence ID No.: 6

Sequence length: 2105

Sequence type: Nucleic acid

Number of strands: Double-strand

Topology: Straight chain

Source:

Biological name: Petunia

Tissue type: Leaf

Direct source:

Library name: cDNA library

Clone name: pSPGT1

Sequence:

```
AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT AGGCACCCCA GGCTTTACAC     60
TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAATT TCACACAGGA    120
AACAGCTATG ACCATGATTA CGCCAAGCTC GAAATTAACC CTCACTAAAG GGAACAAAAG    180
CTGGAGCTCC ACGCGGTGGC GGCCGCTCTA GAACTAGTGG ATCCCCCGGG CTGCAGGAAT    240
TCCGTTGCTG TCGCCACAAT TTACAAACCA AGAAATTAAG CATCCCTTTC CCCCCCTTAA    300
AAAACATACA AGTTTTTAAT TTTTCACTAA GCAAGAAAAT ATG GTG CAG CCT CAT GTC  358
                                            Met Val Gln Pro His Val
                                            1               5
```

```
ATC TTA ACA ACA TTT CCA GCA CAA GGC CAT ATT AAT CCA GCA CTT CAA      406
Ile Leu Thr Thr Phe Pro Ala Gln Gly His Ile Asn Pro Ala Leu Gln
            10                  15                  20
TTT GCC AAG AAT CTT GTC AAG ATG GGC ATA GAA GTG ACA TTT TCT ACA      454
Phe Ala Lys Asn Leu Val Lys Met Gly Ile Glu Val Thr Phe Ser Thr
        25                  30                  35
AGC ATT TAT GCC CAA AGC CGT ATG GAT GAA AAA TCC ATT CTT AAT GCA      502
Ser Ile Tyr Ala Gln Ser Arg Met Asp Glu Lys Ser Ile Leu Asn Ala
    40                  45                  50
CCA AAA GGA TTG AAT TTC ATT CCA TTT TCC GAT GGC TTT GAT GAA GGT      550
Pro Lys Gly Leu Asn Phe Ile Pro Phe Ser Asp Gly Phe Asp Glu Gly
55                  60                  65                  70
TTT GAT CAT TCA AAA GAC CCT GTA TTT TAC ATG TCA CAA CTT CGT AAA      598
Phe Asp His Ser Lys Asp Pro Val Phe Tyr Met Ser Gln Leu Arg Lys
                75                  80                  85
TGT GGA AGT GAA ACT GTC AAA AAA ATA ATT CTC ACT TGC TCT GAA AAT      646
Cys Gly Ser Glu Thr Val Lys Lys Ile Ile Leu Thr Cys Ser Glu Asn
            90                  95                  100
GGA CAG CCT ATA ACT TGC CTA CTT TAC TCC ATT TTC CTT CCT TGG GCA      694
Gly Gln Pro Ile Thr Cys Leu Leu Tyr Ser Ile Phe Leu Pro Trp Ala
        105                 110                 115
GCA GAG GTA GCA CGT GAA GTT CAC ATC CCT TCT GCT CTT CTT TGG AGT      742
Ala Glu Val Ala Arg Glu Val His Ile Pro Ser Ala Leu Leu Trp Ser
    120                 125                 130
CAA CCA GCA ACA ATA TTG GAC ATA TAT TAC TTC AAC TTT CAT GGA TAT      790
Gln Pro Ala Thr Ile Leu Asp Ile Tyr Tyr Phe Asn Phe His Gly Tyr
135                 140                 145                 150
GAA AAA GCT ATG GCT AAT GAA TCC AAT GAT CCA AAT TGG TCC ATT CAA      838
Glu Lys Ala Met Ala Asn Glu Ser Asn Asp Pro Asn Trp Ser Ile Gln
                155                 160                 165
CTT CCC GGG CTT CCA CTA CTG GAA ACT CGA GAT CTT CCT TCA TTT TTA      886
Leu Pro Gly Leu Pro Leu Leu Glu Thr Arg Asp Leu Pro Ser Phe Leu
            170                 175                 180
CTT CCT TAT GGT GCA AAA GGG AGT CTT CGA GTT GCA CTT CCA CCA TTC      934
Leu Pro Tyr Gly Ala Lys Gly Ser Leu Arg Val Ala Leu Pro Pro Phe
        185                 190                 195
```

```
AAA GAA TTG ATA GAC ACA TTA GAT GCT GAA ACC ACT CCT AAG ATT CTT      982
Lys Glu Leu Ile Asp Thr Leu Asp Ala Glu Thr Thr Pro Lys Ile Leu
    200                 205                 210
GTG AAT ACA TTT GAT GAA TTA GAG CCT GAG GCA CTC AAT GCA ATT GAA     1030
Val Asn Thr Phe Asp Glu Leu Glu Pro Glu Ala Leu Asn Ala Ile Glu
215                 220                 225                 230
GGT TAT AAG TTT TAT GGA ATT GGA CCG TTG ATT CCT TCT GCT TTC TTG     1078
Gly Tyr Lys Phe Tyr Gly Ile Gly Pro Leu Ile Pro Ser Ala Phe Leu
                235                 240                 245
GGT GGA AAT GAC CCT TTA GAT GCT TCA TTT GGT GGT GAT CTT TTT CAA     1126
Gly Gly Asn Asp Pro Leu Asp Ala Ser Phe Gly Gly Asp Leu Phe Gln
            250                 255                 260
AAT TCA AAT GAC TAT ATG GAA TGG TTA AAC TCA AAG CCA AAT TCA TCA     1174
Asn Ser Asn Asp Tyr Met Glu Trp Leu Asn Ser Lys Pro Asn Ser Ser
        265                 270                 275
GTT GTT TAT ATA TCT TTT GGG AGT CTA ATG AAT CCA TCT ATT AGC CAA     1222
Val Val Tyr Ile Ser Phe Gly Ser Leu Met Asn Pro Ser Ile Ser Gln
    280                 285                 290
ATG GAG GAG ATA TCA AAA GGG TTG ATA GAC ATA GGA AGG CCG TTT TTA     1270
Met Glu Glu Ile Ser Lys Gly Leu Ile Asp Ile Gly Arg Pro Phe Leu
295                 300                 305                 310
TGG GTG ATA AAA GAA AAT GAA AAA GGC AAA GAA GAA GAG AAT AAA AAG     1318
Trp Val Ile Lys Glu Asn Glu Lys Gly Lys Glu Glu Glu Asn Lys Lys
                315                 320                 325
CTT GGT TGT ATT GAA GAA TTG GAA AAA ATA GGA AAA ATA GTT CCA TGG     1366
Leu Gly Cys Ile Glu Glu Leu Glu Lys Ile Gly Lys Ile Val Pro Trp
            330                 335                 340
TGT TCA CAA CTT GAA GTT CTA AAA CAT CCA TCT TTA GGA TGT TTT GTT     1414
Cys Ser Gln Leu Glu Val Leu Lys His Pro Ser Leu Gly Cys Phe Val
        345                 350                 355
TCT CAT TGT GGA TGG AAT TCA GCC TTA GAG AGT TTA GCT TGT GGA GTG     1462
Ser His Cys Gly Trp Asn Ser Ala Leu Glu Ser Leu Ala Cys Gly Val
    360                 365                 370
CCA GTT GTG GCA TTT CCT CAA TGG ACA GAT CAA ATG ACA AAT GCC AAA     1510
Pro Val Val Ala Phe Pro Gln Trp Thr Asp Gln Met Thr Asn Ala Lys
375                 380                 385                 390
```

```
CAA GTT GAA GAT GTG TGG AAA AGT GGA GTA AGA GTG AGA ATA AAT GAA      1558
Gln Val Glu Asp Val Trp Lys Ser Gly Val Arg Val Arg Ile Asn Glu
                395                 400                 405

GAT GGT GTT GTT GAA AGT GAG GAA ATC AAA AGG TGT ATT GAA TTG GTA      1606
Asp Gly Val Val Glu Ser Glu Glu Ile Lys Arg Cys Ile Glu Leu Val
            410                 415                 420

ATG GAT GGA GGA GAG AAA GGG GAA GAA TTG AGA AAG AAT GCT AAG AAA      1654
Met Asp Gly Gly Glu Lys Gly Glu Glu Leu Arg Lys Asn Ala Lys Lys
        425                 430                 435

TGG AAA GAA TTG GCT AGA GAA GCT GTG AAG GAA GGT GGA TCT TCA CAC      1702
Trp Lys Glu Leu Ala Arg Glu Ala Val Lys Glu Gly Gly Ser Ser His
    440                 445                 450

AAG AAT TTA AAG GCT TTT ATT GAT GAT GTT GCC AAA GGG TTT TAATATTTAC   1754
Lys Asn Leu Lys Ala Phe Ile Asp Asp Val Ala Lys Gly Phe
455                 460                 465         468

AGGCTTTTGC CGTGATATTA CTTCCCCTAG TTGGCGATTC ACTCTTTGTG GACTTGCTTG    1814
ACAAAAAACT GAGGGAATGT GCTAAGACAC GCTAATGCTT TAAGAAGTCA TTTCCAAGGC    1874
TTGAAGCCTG CTTTTAAAAC TTATTAGCCA GTAATCTATA GGGTTCTCTT CTATTTTTCT    1934
CTGTCTCTCT TTTTAGCCTT TTTCTTTCCA AGGTTTAAGA ATAGCGTGAA CATAGCTTAG    1994
TACGTAGTCT TGGTATCTCT ATCTTACCAA GTGCAAGATT ATGCTTATGC TGTCCTCCTA    2054
AATTTCTTAA TAAAATGCAA GATGAAAAAG TACAAAAAAA AAAAAAAAAA A             2105
```

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. An isolated nucleic acid molecule coding for a protein having activity that transfers a glycoside to the 5 position of a flavonoid.

2. A nucleic acid molecule according to claim 1 that codes for a protein having an amino acid sequence described in any one of SEQ ID NOs: 7 to 10 or 12, and having activity that transfers a glycoside to the 5 position of a flavonoid, or a protein having an amino acid sequence modified by addition and/or deletion of one or more amino acids and/or substitutions by one or more other amino acids relative to said amino acids and maintains activity that transfers a glycoside to the 5 position of a flavonoid.

3. A nucleic acid molecule according to claim 1 that codes for a protein having an amino acid sequence that has homology of 30% or more with an amino acid sequence described in any one of SEQ ID NOs: 7 to 10 or 12, and has activity that transfers a glycoside to the 5 position of a flavonoid.

4. A nucleic acid molecule according to claim 1 that codes for a protein having an amino acid sequence that has homology of 50% or more with an amino acid sequence described in any one of SEQ ID NOs: 7 to 10 or 12, and has activity that transfers a glycoside to the 5 position of a flavonoid.

5. A nucleic acid molecule according to any one of claims 1 to 4 that codes for a protein, wherein said nucleic acid molecule can be hybridized under conditions of 5 x SCC and 50°C with all or a portion of a nucleotide sequence that codes for an amino acid sequence described in any one of SEQ ID NOs: 7 to 10 or 12, and has activity

that transfers a glycoside to the 5 position of a flavonoid.

6. A vector containing a nucleic acid molecule according to any one of claims 1 to 5.

7. A host transformed with a vector according to claim 6.

8. A protein encoded by a nucleic acid molecule according to any one of claims 1 to 5.

9. A process for producing a protein comprising culturing or breeding a host according to claim 7, and recovering a protein having activity that transfers a glycoside to the 5 position of a flavonoid from said host.

10. A plant into which is introduced a nucleic acid molecule according to any one of claims 1 to 5, or its progeny or tissue having identical properties.

11. A cut flower of a plant according to claim 10 or its progeny having identical properties.

12. A nucleic acid molecule according to claim 1, substantially as herein described with reference to any of the examples.

Dated this 12th day of September 2002
SUNTORY LIMITED
By their Patent Attorneys
GRIFFITH HACK
Fellows Institute of Patent and
Trade Mark Attorneys of Australia