(51) International Patent Classification: A01H 4/00
5/00, C12N 15/05, 15/29

(21) International Application Number:
PCT/NZ2004/000171

(22) International Filing Date: 2 August 2004 (02.08.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
527326 1 August 2003 (01.08.2003) NZ

(71) Applicant (for all designated States except US):
FONTELLA CO-OPERATIVE GROUP LIMITED
[—(NZ)]; 9 Princes Street, Auckland (NZ).

(72) Inventors; and
(75) Inventors/Applicants (for US only): WHITAKER,
David, John [NZ/NZ]; C/- ViaLactia Biosciences, Level
1, 650 Great South Road, Penrose, Auckland (NZ). EL-
BOROUGH, Kieran, Michael [GB/NZ]; C/- ViaLactia
Biosciences, Level 1, 650 Great South Road, Penrose,
Auckland (NZ). SOUTH, Colin, Robert [NZ/NZ]; C/-
ViaLactia Biosciences, Level 1, 650 Great South Road,
Penrose, Auckland (NZ).

(74) Agents: CALHOUN, Douglas, C et al.; A J Park, 6th
Floor Haddock Parker Building, Post Office Square, P O
Box 949, Wellington, 6015 (NZ).

(54) Title: POLYNUCLEOTIDES, POLYPEPTIDES AND THEIR USE IN THE PRODUCTION OF PLANTS WITH ALTERED CONDENSED TANNINS

(57) Abstract: The invention provides a method of altering condensed tannin production in a plant involving transforming a plant into a plant cell with a genetic construct to alter the expression of a polypeptide involved in condensed tannin production in the plant. The invention also provides transformed plants and plant cells and methods for selecting plants with altered condensed tannin content based on determining presence/content of the polypeptide or nucleotide sequences encoding it. The polypeptide has sequence homology with a chitinase or chitinase-like protein.
POLYNUCLEOTIDES, POLYPEPTIDES AND THEIR USE IN THE PRODUCTION OF PLANTS WITH ALTERED CONDENSED TANNINS

TECHNICAL FIELD

The present invention relates to compositions and methods for producing and selecting plants with altered condensed tannin production.

BACKGROUND ART

Proanthocyanidins or condensed tannins (CTs) are flavonoid oligomers which accumulate in vegetative, fruit and seed tissues of many plants (Perter 1998, In Harborne, JB (Ed) The Flavonoids, Advances in Research Science since 1980, Chapman and Hall, New York, p21). CTs possess many hydrophobic aromatic rings and hydroxyl groups that are able to interact with biological molecules, especially proteins, via hydrophobic interactions and hydrogen bonds. Because of their polymeric nature, the interaction of CTs with protein is considerably stronger than that of protein with other monomeric flavonoids. This is presumably because polymeric CTs are able to interact with large proteins at multiple sites, thus increasing the strength of interaction and minimizing dissociation of formed complexes (Fersht 1985, Chemical catalysis, In Enzyme Structure and Mechanism, WH Freeman and Company, New York, p47). The nature of interaction of CT with proteins forms the basis of some of the beneficial properties of CTs.

Cattle and sheep suffer from a potentially lethal condition called “bloat” when proteins are digested too rapidly in the rumen. This leads to the production of a highly stable foam which results in bloat symptoms. When present in forage crops, CTs, bind to forage proteins, prevent their rapid digestion and thereby reduce the potential for bloating and the need for bloat-preventing drenches. In addition, an increase in CT-bound protein by-passing the rumen, results in improved animal nutrition due to increased availability of nitrogen in the latter stages of the digestive process (McMahon et al., 2000, Can J Plant Sci 80,469; Barry and McNabb, 1999, Br. J. Nutr. 81,263). CTs are absent or at sub-optimal levels in the leaves of certain forage crops such as ryegrass and alfalfa. It would thus be advantageous to increase the content of CTs in appropriate tissues of such crops. In some cases however high levels of CT can reduce the palatability of forage crops (Kumar and Singh, 1984, J Agric Food Chem. 32, 447) and reduction of CT content in such forage crops would be beneficial.
CTs have also received attention due to their powerful antioxidant properties and associated beneficial effects on human health and immunity (Ross and Kasum, 2002, *Annu Rev Nutr* 22,19; Bagchi et al., 2000, *Toxicology* 148, 87; Lin et al., 2002, *J Nat Prod* 65, 505). Of particular interest are the CTs from fruit such as cranberry and grape (Foo et al., 2000, *Phytochemistry* 54, 173; Pataki et al., 2002, *Am J Clin Nutr* 75, 894).

Biosynthesis of CTs occurs in the flavonoid biosynthetic pathway (Figure 1). The flavonoids are secondary metabolites derived from products of the phenylpropanoid biosynthetic pathway and the Krebs cycle. The final flavonoid compounds include flavonols (colorless pigments), anthocyanins (pink and red pigments), and condensed tannins (colorless pigments that turn brown on oxidation). The sub-pathways for each of these flavonoid compounds share common initial biosynthesis steps.

Only two biochemical steps specific to CT biosynthesis have been demonstrated *in vitro*. The best characterized step is the reduction of (+)3, 4-cis-leucocyanidin to catechin, catalysed by leucoanthocyanidin reductase (LAR). The LAR enzyme was purified from tissues of the CT-rich legume *Desmodium uncinatum* and the corresponding cDNA was cloned (Tanner et al., 2003 *JBC*, in press). Activity of the recombinant LAR enzyme in *Escherichia coli* and plants confirmed the identity of the protein. The second enzyme which appears specific to condensed tannin biosynthesis is anthocyanidin reductase, encoded by the *BANYULS* (BAN) gene of *Arabidopsis thaliana* and *Medicago truncatula* (Xie et al., 2003, *Science*, 299, 396). *BANYULS* cDNAs from *Arabidopsis* and *Medicago* were expressed in *Escherichia coli*, and the recombinant proteins were shown to convert anthocyanidins to their corresponding 2,3-cis-flavan-3-ols. This activity has yet to be been confirmed in plant extracts.

Genetic and molecular studies indicate the regulation of CT biosynthesis in the model plant *Arabidopsis thaliana* is complex. In *Arabidopsis* CTs specifically accumulate in the endothelium during early seed development. Several genes involved in the regulation of CT biosynthesis in *Arabidopsis* seed have been characterized, but unidentified tissue-specific factors also appear necessary. (Nesi et al., 2000, *Plant Cell* 12, 1863; Nesi et al., 2001, *Plant Cell* 13, 2099; Nesi et al., 2002, *Plant Cell* 14, 2463). The maize SN gene has also been implicated in the regulation of CT biosynthesis (Robbins et al., 2003, *J Exp Bot.* 54, 239). SN appears to be orthologous to tt8, one of the regulatory genes characterized in *Arabidopsis*. Like tt8, the SN gene appears to up-regulate CTs only in competent tissues.
Plants with altered CT content may also have altered nutritional value, increased pest and disease resistance or improved nutritional qualities (WO 98/07836). Plants with altered CT content may also lead to improved silage storage through reduced bacterial and fungal activity, reduced methane production in cattle grazed up such plants, reduced parasite load in the intestines of animals grazed on such plants, improved resistance to insects and birds, improved quality of oilseeds and grains and altered flavour, colour and taste characteristics in fruit, fruit juice and wine (WO 02/10412).

As the content of CTs in dietary plants can impart both positive and negative effects, an important objective for agriculture is to be able to manipulate the level of CTs in plants. There is thus a need for compositions and methods for developing both forage and human crop varieties with altered CT content.

WO 02/10412 is directed towards novel regulatory and biosynthetic genes and encoded proteins which can be used in methods to alter tannin production in plants for a variety of agricultural benefits. No examples describing such plants are provided.

Over-expression of BAN in tobacco petals and Arabidopsis leaves leads to loss of anthocyanin and accumulation of CTs (Xie et al, 2003, Science, 299, 396).

WO 98/07836 is directed towards sequences encoding leucoanthocyanidin reductase (LAR) and to use of such sequences to regulate condensed tannin biosynthesis in plants. The sequences are stated to be useful in the production of bloat-sage forage crops and other crops with altered nutritional value, increased disease resistance and pest resistance or improved malting qualities. Prophetic examples of such crops are provided.

It is an object of the invention to provide plants with altered production of CTs or at least to provide the public with a useful choice.

**DISCLOSURE OF THE INVENTION**

In one aspect the invention provides a method of altering condensed tannin production in a plant comprising transformation of a plant cell or plant with a genetic construct comprising a polynucleotide that alters the expression of a polypeptide comprising [SEQ ID NO:1] involved in condensed tannin
production in the plant wherein \( X_1=K, E, I \) or \( R, X_2=Y \) or \( F, X_3=H, Q, D \) or \( G, X_4=S \) or \( T, X_5=Y \) or \( F, X_6=Q, E \) or \( K, X_7=T, V \) or \( A, X_8=Y \) or \( H, X_9=Q, K \) or \( H, X_{10}=D, N, Q \) or \( E, X_{11}=A, S \) or \( P \) and \( X_{12}=K, R \) or \( H \) and wherein the polynucleotide encodes the polypeptide or a fragment thereof or polynucleotide is complementary to said polynucleotide encoding the polypeptide or fragment.

In another aspect the invention provides a method of altering condensed tannin production in a plant comprising transformation of a plant cell or plant with a genetic construct comprising a sequence selected from SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a functionally equivalent variant thereof, wherein said genetic construct encodes a polypeptide involved in condensed tannin production in plants.

In a further aspect the invention provides a transgenic plant having altered condensed tannin content comprising a polynucleotide that alters the expression of a polypeptide comprising \( S X_1 X_2 C C N X_3 T I X_4 D X_5 F X_6 X_7 X_8 X_9 F E X_{10} L F X_{11} X_{12} R N \) (SEQ ID NO:1) involved in condensed tannin production in the plant wherein \( X_1=K, E, I \) or \( R, X_2=Y \) or \( F, X_3=H, Q, D \) or \( G, X_4=S \) or \( T, X_5=Y \) or \( F, X_6=Q, E \) or \( K, X_7=T, V \) or \( A, X_8=Y \) or \( H, X_9=Q, K \) or \( H, X_{10}=D, N, Q \) or \( E, X_{11}=A, S \) or \( P \) and \( X_{12}=K, R \) or \( H \) and wherein the polynucleotide encodes the polypeptide or a fragment thereof or the polynucleotide is complementary to said polynucleotide encoding the polypeptide or fragment.

In a further aspect the invention provides a transgenic plant having altered condensed tannin content comprising a genetic construct comprising a sequence selected from SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a functionally equivalent variant thereof, wherein said genetic construct encodes a polypeptide involved in condensed tannin production in plants.

In preferred embodiments the polynucleotide encodes a polypeptide comprising the amino acid sequence:

\[
S X_1 X_2 C C N X_3 T I X_4 D X_5 F X_6 X_7 X_8 X_9 F E X_{10} L F X_{11} X_{12} R N (SEQ ID NO:1)
\]

wherein \( X_1 \) to \( X_{12} \) are as herein defined, which encodes a polypeptide involved in condensed tannin production in plants.

In preferred embodiments the polynucleotide comprises a sequence selected from a group of related sequences (SEQ ID NOs: 2, 3 and 4) or a variant thereof, wherein the polynucleotide encodes a polypeptide involved in condensed tannin production in plants.
In one embodiment the invention the polynucleotide comprises SEQ ID NO:2 or a variant thereof, and the polynucleotide encodes a polypeptide involved in condensed tannin production in plants.

In a further embodiment the invention the polynucleotide comprises SEQ ID NO:3 or a variant thereof, and the polynucleotide encodes a polypeptide involved in condensed tannin production in plants. A preferred variant is the polypeptide encoded by F286 (see US Published Patent Application 2004/0049808, paragraph 67).

In a further embodiment the invention the polynucleotide comprises SEQ ID NO:4 or a variant thereof, and the polynucleotide encodes a polypeptide involved in condensed tannin production in plants.

In further embodiments, the polynucleotide is selected from naturally occurring polynucleotide orthologue of SEQ ID NOs: 2, 3 and 4, and variants of such orthologues, wherein the polynucleotide orthologues and variants encode polypeptides involved in condensed tannin production in plants. Exemplary orthologues are disclosed herein and identified as SEQ ID NOs: 5-9 of the sequence listing.

Fragments of SEQ ID NOs: 2, 3 and 4 are useful as probes, primers, gene silencing molecules, and in polynucleotide-based selection methods of the invention. Polynucleotides comprising fragments of orthologues and/or variants are useful in the invention.

Isolated polynucleotides are useful in methods for altering condensed tannin production in plants and in traditional breeding programs for selection of individuals with altered CT content.

In a further aspect the invention provides a plant cell which comprises one or more of the genetic constructs comprising a polynucleotide as described above.

In a further aspect the invention provides methods for selecting plants with altered condensed tannin content comprising the steps of:

(a) contacting a sample from a plant with a polynucleotide selected from SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 and orthologues thereof, or a fragment of one of these, as a probe or primer...
(b) determining presence/content of complementary sequences to assess altered expression of a polypeptide involved in condensed tannin production in a plant; and
(c) selection of a plant showing such altered expression.

In a further aspect the invention provides methods for selecting plants with altered condensed tannin content comprising the steps of:

(a) contacting a plant sample or plant antibody raised against a polypeptide coded by a nucleotide sequence comprising SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a naturally occurring orthologue thereof to assess altered expression of a polypeptide involved in condensed tannin production in a plant; and
(b) selection of a plant shown such altered expression.

In a further aspect the invention provides plants selected by a method of the invention.

The term “plant selected by a method of the invention” is intended to include any whole plant, part, portion propagules or progeny of a plant selected by a method of the invention.

Definitions

The term “transgenic plant” is intended to include a whole plant or any part of a plant, propagules and progeny of a plant.

The term ‘propagule’ means any part of a plant that may be used in reproduction or propagation, either sexual or asexual, including seeds and cuttings.

The term “polynucleotide(s),” as used herein, means a single or double-stranded deoxyribonucleotide or ribonucleotide polymer of any length, and include as non-limiting examples, coding and non-coding sequences of a gene, sense and antisense sequences, exons, introns, genomic DNA, cDNA, pre-mRNA, mRNA, rRNA, siRNA, miRNA, tRNA, ribozymes, recombinant polynucleotides, isolated and purified naturally occurring DNA or RNA sequences, synthetic RNA and DNA sequences, nucleic acid probes, primers, fragments, genetic constructs, vectors and modified polynucleotides.
A "fragment" of a polynucleotide sequence provided herein is a subsequence of contiguous nucleotides that is capable of specific hybridization to a target of interest, e.g., a sequence that is at least 15 nucleotides in length. A fragment of a polynucleotide sequence can be used in antisense, gene silencing, triple helix or ribozyme technology, or as a primer, a probe, included in a microarray, or used in polynucleotide-based selection methods of the invention.

The term "primer" refers to a short polynucleotide, usually having a free 3’OH group, that is hybridized to a template and used for priming polymerization of a polynucleotide complementary to the target.

The term "probe" refers to a short polynucleotide that is used to detect a polynucleotide sequence that is complementary to the probe in a hybridization-based assay.

The term "polypeptide", as used herein, encompasses amino acid chains of any length, including full-length proteins, in which amino acid residues are linked by covalent peptide bonds. Polypeptides may be purified natural products, or may be produced partially or wholly using recombinant or synthetic techniques.

A "fragment" of a polypeptide is a subsequence of the polypeptide that performs a function that is required for the biological activity and/or provides three dimensional structure of the polypeptide.

The term "isolated" as applied to the polynucleotide or polypeptide sequences disclosed herein is used to refer to sequences that are removed from their natural cellular environment. An isolated molecule may be obtained by any method or combination of methods including biochemical, recombinant, and synthetic techniques.

The term "recombinant" refers to a polynucleotide sequence that is removed from sequences that surround it in its natural context and/or is recombined with sequences that are not present in its natural context.

A "recombinant" polypeptide sequence is produced by translation from a "recombinant" polynucleotide sequence.
As used herein, the term "variant" refers to polynucleotide or polypeptide sequences different from the specifically identified sequences, wherein one or more nucleotides or amino acid residues is deleted, substituted, or added. Variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variants may be from the same or from other species and encompass homologous, paralogues and orthologues. The term "variant" encompasses both cDNA and genomic sequences. Variant sequences (polynucleotide or polypeptide) preferably exhibit at least 50%, more preferably at least 70%, more preferably at least 80%, more preferably at least 90%, more preferably at least 95%, more preferably at least 98%, and most preferably at least 99% identity to a sequences of the present invention. The term variant, with reference to polynucleotides, also encompasses sequences complementary to any of the polynucleotides used in the invention.

The term "percent identity" or "percent identical" when used in connection with the polynucleotide molecules and polypeptides, is defined as the percentage of nucleic acid residues in a candidate polynucleotide sequence, or the percentage of amino acid residues in a candidate polypeptide sequence, that are identical with a subject polynucleotide sequence or polypeptide molecule sequence (such as the polynucleotide sequence of SEQ ID NO: 2), after aligning the candidate and subject sequences to achieve the maximum percent identity, and not considering any nucleic acid residue substitutions as part of the polynucleotide sequence identity. When making the comparison, the candidate polynucleotide sequence or polypeptide sequence (which may be a portion of a larger polynucleotide sequence or polypeptide sequence) is the same length as the subject polynucleotide sequence or polypeptide sequence, and no gaps are introduced into the candidate polynucleotide sequence or polypeptide sequence in order to achieve the best alignment.

Polynucleotide sequence identity can be determined in the following manner. The subject polynucleotide sequence is used to search a polynucleotide sequence database, such as the GenBank database (accessibe at web site http://www.ncbi.nlm.nih.gov/blast/), using the program BLASTN version 2.2.5 (based on Altschul et al., 1997, Nucleic Acids Res, 25, 3389). The programs are used in the ungapped mode. Default filtering is used to remove sequence homologies due to regions of low complexity. The default parameters of BLASTN, BLASTX and tBLASTX are utilized.

Polypeptide sequence identity can be determined in the following manner. The subject polypeptide sequence is used to search a polypeptide sequence database, such as the GenBank
database (accessible at web site http://www.ncbi.nlm.nih.gov/blast/), using the BLASTP program. The program is used in the ungapped mode. Default filtering is used to remove sequence homologies due to regions of low complexity. The default parameters of BLASTP are utilized.

According to one embodiment, "variant" polynucleotides and polypeptides, with reference to each of the polynucleotides and polypeptides of the present invention, produce an E value of 0.01 or less when compared to the polynucleotide or polypeptide of the present invention using BLASTN, FASTA or BLASTP algorithms set to default parameters. That is, a variant polynucleotide or polypeptide is any sequence that has at least a 99% probability of being the same as the polynucleotide or polypeptide of the present invention. According to a preferred embodiment, a variant polynucleotide is a sequence having the same number or fewer nucleic acids than a polynucleotide of the present invention that has at least a 99% probability of being the same as the polynucleotide of the present invention, measured as having an E value of 0.01 or less using the BLASTN or FASTA algorithms set at parameters described above. Similarly, according to a preferred embodiment, a variant polypeptide is a sequence having the same number or fewer amino acids than a polypeptide of the present invention that has at least a 99% probability of being the same as a polypeptide of the present invention, measured as having an E value of 0.01 or less using the BLASTP algorithm set at the parameters described above.

Alternatively, variant polynucleotides of the present invention hybridize to the polynucleotide sequences recited in SEQ ID NO: 2-9, or complements thereof, antisense sequences and complements thereof, under stringent conditions. As used herein, "stringent conditions" refers to hybridization conditions such as prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65°C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65°C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65°C.

Variant polynucleotides of the present invention also encompasses polynucleotides that differ from the sequences of the invention but that, as a consequence of the degeneracy of the genetic code, encode a polypeptide having similar activity to a polypeptide encoded by a polynucleotide of the present invention. A sequence alteration that does not change the amino acid sequence of the polypeptide is a "silent variation". Except for ATG (methionine) and TGG (tryptophan), other codons for the same amino acid may be changed by art recognized techniques, e.g., to optimize codon expression in a particular host organism. Conservative substitutions of one or several amino acids of a described polypeptide sequence without significantly altering its
biological activity are also included in the invention. A skilled artisan will be aware of methods for making phenotypically silent amino acid substitutions (see, e.g., Bowie et al., 1990, Science 247, 1306).

Additionally, polynucleotides comprising sequences that differ from the polynucleotide sequences identified above as a result of deletions and/or insertions totaling less than 10% of the total sequence length are also useful in methods and products of the present invention. Similarly, polypeptides comprising sequences that differ from the polypeptide sequences identified as a result of amino acid substitutions, insertions, and/or deletions totaling less than 10% of the total sequence length are useful in methods and products of the present invention. In certain embodiments, variants of the inventive polypeptides and polynucleotides possess biological activities that are the same or similar to those of the inventive polypeptides or polynucleotides.

The term "hybridize under stringent conditions", and grammatical equivalents thereof, refers to the ability of a polynucleotide molecule to hybridize to a target polynucleotide molecule (such as a target polynucleotide molecule immobilized on a DNA or RNA blot, such as a Southern blot or Northern blot) under defined conditions of temperature and salt concentration. The ability to hybridize under stringent hybridization conditions can be determined by initially hybridizing under less stringent conditions then increasing the stringency to the desired stringency.

With respect to polynucleotide molecules greater than about 100 bases in length, typical stringent hybridization conditions are no more than 25 to 30°C (for example, 10°C) below the melting temperature (Tm) of the native duplex (see generally, Sambrook et al., Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Ausubel et al., 1987, Current Protocols in Molecular Biology, Greene Publishing.). Tm for polynucleotide molecules greater than about 100 bases can be calculated by the formula Tm = 81.5 + 0.41% (G + C-log (Na+)).( Sambrook et al., Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Bolton and McCarthy, 1962, PNAS 84:1390)

With respect to polynucleotide molecules having a length less than 100 bases, exemplary stringent hybridization conditions are 5 to 10°C below Tm. On average, the Tm of a polynucleotide molecule of length less than 100 bp is reduced by approximately (500/oligonucleotide length)^o C.
With respect to the DNA mimics known as peptide nucleic acids (PNA’s) (Nielsen et al., Science. 1991 Dec 6;254(5037):1497-500) Tm values are higher than those for DNA-DNA or DNA-RNA hybrids, and can be calculated using the formula described in Giesen et al., Nucleic Acids Res. 1998 Nov 1;26(21):5004-6. Exemplary stringent hybridization conditions for a DNA-PNA hybrid having a length less than 100 bases are 5 to 10° C below the Tm.

The term "genetic construct" refers to a polynucleotide molecule, usually double-stranded DNA, which may have inserted into it another polynucleotide molecule (the insert polynucleotide molecule) such as, but not limited to, a cDNA molecule. A genetic construct may contain the necessary elements that permit transcribing the insert polynucleotide molecule, and, optionally, translating the transcript into a polypeptide. The insert polynucleotide molecule may be derived from the host cell, or may be derived from a different cell or organism and/or may be a recombinant polynucleotide. Once inside the host cell the genetic construct may become integrated in the host chromosomal DNA. The genetic construct may be linked to a vector.

The term “vector” refers to a polynucleotide molecule, usually double stranded DNA, which is used to transport the genetic construct into a host cell. The vector may be capable of replication in at least one host system, such as E. coli.

The term "expression construct" refers to a genetic construct that includes the necessary elements that permit transcribing the insert polynucleotide molecule, and, optionally, translating the transcript into a polypeptide. An expression construct typically comprises in a 5’ to 3’ direction:

a) a promoter functional in the host cell into which the construct will be transformed,

b) the polynucleotide to be expressed, and

c) a terminator functional in the host cell into which the construct will be transformed.

The term “coding region” or “open reading frame” (ORF) refers to the sense strand of a genomic DNA sequence or a cDNA sequence that is capable of producing a transcription product and/or a polypeptide under the control of appropriate regulatory sequences. The coding sequence is identified by the presence of a 5’ translation start codon and a 3’ translation stop codon. When inserted into a genetic construct, a “coding sequence” is capable of being expressed when it is operably linked to promoter and terminator sequences.
"Operably-linked" means that the coding sequence is in the correct reading frame for translation and its expression is placed under the control of regulatory elements that include promoters, tissue-specific regulatory elements, temporal regulatory elements, enhancers, silencers and terminators.

The term “noncoding region” refers to untranslated sequences that are upstream of the translational start site and downstream of the translational stop site. These sequences are also referred to respectively as the 5' UTR and the 3' UTR. These regions include elements required for transcription initiation and termination and for regulation of translation efficiency.

Terminators are sequences, which terminate transcription, and are found in the 3' untranslated ends of genes downstream of the translated sequence. Terminators are important determinants of mRNA stability and in some cases have been found to have spatial regulatory functions.

The term “promoter” refers to nontranscribed cis-regulatory elements upstream of the coding region that regulate gene transcription. Promoters comprise cis-initiator elements which specify the transcription initiation site and conserved boxes such as the TATA box, and motifs that are bound by transcription factors.

A “transgene” is a polynucleotide that is taken from one organism and introduced into a different organism by transformation. The transgene may be derived from the same species or from a different species as the species of the organism into which the transgene is introduced.

A “transgenic plant” refers to a plant which contains new genetic material as a result of genetic manipulation or transformation. The new genetic material may be derived from a plant of the same species as the resulting transgenic plant or from a different species.

An “inverted repeat” is a sequence that is repeated, where the second half of the repeat is in the complementary strand, e.g.,

(5')GATCTA......TAGATC(3')
(3')CTAGAT......ATCTAG(5')

Read-through transcription will produce a transcript that undergoes complementary base-pairing to form a hairpin structure provided that there is a 3-5 bp spacer between the repeated regions.
The term “involved in condensed tannin production” means that such polynucleotides and polypeptides may function to activate, inhibit or otherwise modulate CT production in a plant. Such modulation is intended to include alteration of the type of CT monomer/s present and alteration of the level of polymerization of CTs. Such polynucleotides and polypeptide may act within the CT biosynthetic pathway or may function in regulation of the CT biosynthetic pathway.

The term “polypeptide involved in condensed tannin production” refers to a polypeptide, an aggregate of a polypeptide such as a dimer or other multimer, a fusion polypeptide, a polypeptide fragment, a polypeptide variant, or derivative thereof capable of affecting the production of CT in a plant.

The applicants have identified polynucleotides involved in condensed tannin biosynthesis in plants. A polynucleotide sequence, up-regulated in a naturally occurring *Gossypium hirsutum* (cotton) mutant, which produces unusually high levels of CTs, relative to wild-type plants, was identified by subtractive methods. This cotton polynucleotide sequence was also used to identify and isolate further orthologous polynucleotides from *Arabidopsis thaliana*, *Lolium perenne* (ryegrass) and other species. Where available, ‘in-silico’ data indicated that the orthologous polynucleotides are expressed in tissues where condensed tannins are expected to be produced. The applicants further identified a consensus polypeptide sequence present in all of the polypeptides encoded by the orthologous polynucleotide sequences, as shown in Figure 3.

The invention provides transgenic plants with altered CT content, including plants with both increased and decreased levels of CTs and methods for the production of such plants. The invention further provides methods for selection of plants with altered condensed tannin content, as compared to the majority of similar wild-type plants. Such methods based on use of polynucleotides or polypeptides described herein may be used in traditional breeding programs in the development of varieties with altered condensed tannin content.

In a first aspect the invention uses an isolated polynucleotide encoding a polypeptide comprising the amino acid sequence:

\[ S \ X_1 \ X_2 \ C \ C \ N \ X_3 \ T \ I \ X_4 \ D \ X_5 \ F \ X_6 \ X_7 \ X_8 \ X_9 \ F \ E \ X_{10} \ L \ F \ X_{11} \ X_{12} \ R \ N \ (SEQ \ ID \ NO:1) \]

wherein \( X_1 = K \), \( E \), \( I \) or \( R \), \( X_2 = Y \) or \( F \), \( X_3 = H \), \( Q \), \( D \) or \( G \), \( X_4 = S \) or \( T \), \( X_5 = Y \) or \( F \), \( X_6 = Q \), \( E \) or \( K \), \( X_7 = T \), \( V \) or \( A \), \( X_8 = Y \) or \( H \), \( X_9 = Q \), \( K \) or \( H \), \( X_{10} = D \), \( N \), \( Q \) or \( E \), \( X_{11} = A \), \( S \) or \( P \) and \( X_{12} = K \), \( R \) or \( H \).
The isolated polynucleotide of this aspect encodes a polypeptide involved in condensed tannin production in plants.

The invention further uses an isolated polynucleotide comprising a sequence selected from a group of related sequences derived from *Gossypium hirsutum* (cotton), *Arabidopsis thaliana* and *Lolium perenne* (ryegrass) (SEQ ID NOs: 2, 3 and 4 respectively) or a variant thereof, wherein the polynucleotide encodes a polypeptide involved in condensed tannin production in plants.

The present invention also encompasses use of polynucleotide variants of SEQ ID NOs: 2, 3 and 4. Such polynucleotide variants preferably exhibit at least 50%, more preferably at least 70%, more preferably at least 80%, more preferably at least 90%, more preferably at least 95%, more preferably at least 98%, and most preferably at least 99% identity to any one of the sequences set forth in SEQ ID NOs: 2, 3, or 4.

The invention further provides naturally occurring polynucleotide orthologues of SEQ ID NOs: 2, 3 or 4, and variants of such orthologues. Exemplary orthologues are disclosed herein and identified as SEQ ID NO: 5-9 of the sequence listing.

The polynucleotide molecules can be isolated by using a variety of techniques known to those of ordinary skill in the art. By way of example, such polynucleotides can be isolated through use of the polymerase chain reaction (PCR) described in Mullis *et al.*, Eds. 1994 The Polymerase Chain Reaction, Birkhauser, incorporated herein by reference.

The isolated polynucleotides can be amplified using primers, as defined herein, derived from the polynucleotide sequences. Variant polynucleotides may also be identified using such primers. Typically, the polynucleotide sequence of a primer, useful to amplify variants of polynucleotide molecules by PCR, may be based on a conserved region of amino acid sequence of the polypeptides.

Further, by way of example, inverse PCR permits acquisition of unknown sequences, flanking the polynucleotide sequences disclosed herein, starting with primers based on a known region (Triglia *et al.*, 1998, *Nucleic Acids Res* 16, 8186, incorporated herein by reference). The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template. Divergent primers are designed from the known region.
Further methods for identifying variant polynucleotides by methods of the invention include use of all, or portions of, the polynucleotides having the sequence set forth in SEQ ID NO: 2-4 as hybridization probes to screen a plant genomic or cDNA libraries. The technique of hybridizing labelled polynucleotide probes to polynucleotides immobilized on nitrocellulose filters or nylon membranes can be used to screen the genomic or cDNA libraries. Exemplary hybridization and wash conditions are: hybridization for 20 hours at 65°C in 5.0 X SSC, 0.5% sodium dodecyl sulfate, 1X Denhardt's solution; washing (three washes of twenty minutes each at 55°C) in 1.0 X SSC, 1% (w/v) sodium dodecyl sulfate, and optionally one wash (for twenty minutes) in 0.5 X SSC, 1% (w/v) sodium dodecyl sulfate, at 60°C. An optional further wash (for twenty minutes) can be conducted under conditions of 0.1 X SSC, 1% (w/v) sodium dodecyl sulfate, at 60°C.

The variant sequences of the invention can also be identified by computer-based methods, using public domain sequence alignment algorithms and sequence similarity search tools to search sequence databases (public domain databases include Genbank, EMBL, Swiss-Prot, PIR and others). See, e.g., Nucleic Acids Res. 29: 1-10 and 11-16, 2001 for examples of online resources. Similarity searches retrieve and align sequences for comparison with a target sequence to be analyzed (i.e., a query sequence). Sequence comparison algorithms use scoring matrices to assign an overall score to each of the alignments.

Polynucleotide and polypeptide sequences may be aligned, and percentage of identical residues in a specified region may be determined against other polynucleotide and polypeptide sequences, using computer algorithms that are publicly available. Two exemplary algorithms for aligning and identifying the similarity of polynucleotide sequences are the BLASTN and FASTA algorithms. Polynucleotides may also be analyzed using the BLASTX algorithm, which compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database. The similarity of polypeptide sequences may be examined using the BLASTP algorithm. The BLASTN algorithm Version 2.0.11 [Jan-20-2000] set to the default parameters described in the documentation and distributed with the algorithm, are preferred for use in the determination of polynucleotide variants used in the present invention. The following parameters are preferred for determination of alignments and similarities using the BLASTX program version 2.0.11 [Jan-20-2000] Unix running command: blastall -p blastn -d swissprot TrembLeDB -e 10 -G 0 -E 0 -v 30 -b 30 -i queryseq -o results. The BLASTP algorithm, is preferred for use in the determination of polypeptide variants used in the
present invention. The following running parameters are preferred for determination of alignments and similarities using BLASTP that contribute to the E values and percentage identity of polypeptide sequences: blastall -p blastp -d swissprotbrembledb -e 10 -G 0 -E 0 -v 30 -b 30 -i queryseq -o results; the parameters are: -p Program Name [String]; -d Database [String]; -e Expectation value (E) [Real]; -G Cost to open a gap (zero invokes default behavior) [Integer]; -E Cost to extend a gap (zero invokes default behavior) [Integer]; -v Number of one-line descriptions (v) [Integer]; -b Number of alignments to show (b) [Integer]; -I Query File [File In]; -o BLAST report Output File [File Out] Optional.

The use of the BLAST family of algorithms, including BLASTN, BLASTP, and BLASTX, is described in the publication of Altschul et al., Nucleic Acids Res. 25: 3389-3402, 1997. The BLASTN software is available on the NCBI server (http://ncbi.nlm.nih.gov) under /blast/executables/ and is available from the National Center for Biotechnology Information (NCBI), National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894 USA.

The FASTA software package is available from the University of Virginia (University of Virginia, PO Box 9025, Charlottesville, VA 22906-9025). Version 2.0u4, February 1996, set to the default parameters described in the documentation and distributed with the algorithm, may be used in the determination of variants used in the present invention. The use of the FASTA algorithm is described in Pearson and Lipman, 1988, Proc Natl Acad Sci USA, 85, 2444; and Pearson, 1990, Methods in Enzymol. 183, 63.

The "hits" to one or more database sequences by a queried sequence produced by BLASTN, FASTA, BLASTX, BLASTP or a similar algorithm, align and identify similar portions of sequences. The hits are arranged in order of the degree of similarity and the length of sequence overlap. Hits to a database sequence generally represent an overlap over only a fraction of the sequence length of the queried sequence.

The BLASTN, FASTA and BLASTP algorithms also produce "Expect" values for alignments. The Expect value (E) indicates the number of hits one can "expect" to see over a certain number of contiguous sequences by chance when searching a database of a certain size. The Expect value is used as a significance threshold for determining whether the hit to a database, such as the preferred EMBL database, indicates true similarity. For example, an E value of 0.1 assigned to a polynucleotide hit is interpreted as meaning that in a database of the size of the EMBL
database, one might expect to see 0.1 matches over the aligned portion of the sequence with a similar score simply by chance. By this criterion, the aligned and matched portions of the polynucleotide sequences then have a probability of 90% of being the same. For sequences having an E value of 0.01 or less over aligned and matched portions, the probability of finding a match by chance in the EMBL database is 1% or less using the BLASTN, FASTA, BLASTX or BLASTP algorithm.

Multiple sequence alignments of a group of related sequences can be carried out with PILEUP, which uses progressive, pairwise alignments. (Feng and Doolittle, 1987, J. Mol. Evol. 25, 351).

Pattern recognition software applications are available for finding motifs or signature sequences. For example, MEME (Multiple Em for Motif Elicitation) finds motifs and signature sequences in a set of sequences, and MAST (Motif Alignment and Search Tool) uses these motifs to identify similar or the same motifs in query sequences. The MAST results are provided as a series of alignments with appropriate statistical data and a visual overview of the motifs found. MEME and MAST were developed at the University of California, San Diego.

PROSITE (Bairoch and Bucher, 1994, Nucleic Acids Res. 22, 3583; Hofmann et al., 1999, Nucleic Acids Res. 27, 215) is a method of identifying the functions of uncharacterized proteins translated from genomic or cDNA sequences. The PROSITE database (www.expasy.org/prosite) contains biologically significant patterns and profiles and is designed so that it can be used with appropriate computational tools to assign a new sequence to a known family of proteins or to determine which known domain(s) are present in the sequence (Falquet et al., 2002, Nucleic Acids Res. 30, 235). Prosearch is a tool that can search SWISS-PROT and EMBL databases with a given sequence pattern or signature.

Proteins can be classified according to their sequence relatedness to other proteins in the same genome (paralogues) or a different genome (orthologues). Orthologous genes are genes that evolved by speciation from a common ancestral gene. These genes normally retain the same function as they evolve. Paralogous genes are genes that are duplicated within a genome. These genes may acquire new specificities or modified functions which may be related to the original one. Phylogenetic analysis methods are reviewed in Tatusov et al., 1997, Science 278, 631-637.)
A partial cDNA sequence may be used, in methods well-known in the art to identify the corresponding full length polynucleotide sequence containing an open reading frame (ORF) necessary for over-expression of the encoded polypeptide in transgenic organisms or cell/s of such organisms. Such organisms would include bacteria, yeast, fungi, insects, mammals and plants. Such methods would include PCR-based methods, hybridization-based method, computer/database-based methods. In order to physically assemble full-length clones, standard molecular biology approaches can be utilized (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987).

The involvement of variant polynucleotides in condensed tannin production may be assessed by the transgenic methodologies described herein together with analysis of condensed tannin content of the resulting transgenic plants as described in Example 1.

Described herein is an isolated polypeptide comprising the amino acid sequence:

\[
S X_1 X_2 C C N X_3 T I X_4 D X_5 F X_6 X_7 X_8 X_9 F E X_{10} L F X_{11} X_{12} R N (SEQ~ID~NO:1)
\]

wherein \(X_1 = K, E, I\) or R, \(X_2 = Y\) or F, \(X_3 = H, Q, D\) or G, \(X_4 = S\) or T, \(X_5 = Y\) or F, \(X_6 = Q, E\) or K, \(X_7 = T, V\) or A, \(X_8 = Y\) or H, \(X_9 = Q, K\) or H, \(X_{10} = D, N, Q\) or E, \(X_{11} = A, S\) or P and \(X_{12} = K, R\) or H. The polypeptide is involved in condensed tannin production in plants.

The isolated polypeptide may be selected from a group of related polypeptides comprising sequences derived from *Gossypium hirsutum* (cotton), *Arabidopsis thaliana* and *Lolium perenne* (SEQ ID NO’s: 10, 11 and 12 respectively) or a variant thereof, wherein the polypeptide is involved in condensed tannin production in plant.

Also described herein are polypeptide variants of SEQ ID NOs: 10, 11 and 12. Such variants preferably exhibit at least 50%, more preferably at least 70%, more preferably at least 80%, more preferably at least 90%, more preferably at least 95%, more preferably at least 98%, and most preferably at least 99% identity to any one of the sequences set forth in SEQ ID NOs: 10, 11 and 12.

Also described herein are naturally occurring polypeptide orthologues of SEQ ID NO’s 10, 11 and 12, and variants of such orthologues. Exemplary orthologues are disclosed herein and identified as SEQ ID NOs: 13-17 of the sequence listing.
The polypeptides may be prepared using peptide synthesis methods well known in the art such as direct peptide synthesis using solid phase techniques (e.g. Stewart et al., 1969, in Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco California, or automated synthesis, for example using an Applied Biosystems 431A Peptide Synthesizer (Foster City, California).

Mutated forms of the polypeptides may also be produced during such syntheses.

The polypeptides may also be purified from natural sources using a variety of techniques that are well known in the art (e.g. Deutscher, 1990, Ed, Methods in Enzymology, Vol. 182, Guide to Protein Purification,).

Alternatively the polypeptides may be expressed recombinantly in suitable host cells and separated from the cells as discussed below.

The polypeptide variants may be identified by screening expression libraries using antibodies raised against previously identified polypeptides. Such antibodies may also be used in the isolation of polypeptide variants. The DNA sequence of polynucleotides identified through expression screening, as well as with the DNA sequences of the polynucleotides already identified may be used in computer based approaches as described above to identify polynucleotide sequences encoding polypeptide variants. Such polynucleotide sequences may be utilized for the recombinant production of polypeptide variants of the invention. Such procedures are standard practice for art skilled workers (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987).

The genetic constructs used in the present invention comprise one or more polynucleotide sequences of the invention and/or polynucleotides encoding polypeptides of the invention, and may be useful for transforming, for example, bacterial, fungal, insect, mammalian or plant organisms. The genetic constructs are intended to include expression constructs as herein defined.

A host cell which comprises a genetic construct or vector comprising a polynucleotide as described above may be used in the invention. Host cells may be derived from, for example, bacterial, fungal, insect, mammalian or plant organisms.

Host cells comprising genetic constructs, such as expression constructs, of the invention are useful in methods well known in the art (e.g. Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987; Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing, 1987) for recombinant production of polypeptides described herein. Such methods may involve the culture of host cells in an appropriate medium in conditions suitable for or conducive to expression of a polypeptide of the invention. The expressed recombinant polypeptide, which may optionally be secreted into the culture, may then be separated from the medium, host cells or culture medium by methods well known in the art (e.g. Deutscher, Ed, 1990, Methods in Enzymology, Vol 182, Guide to Protein Purification).

The host cells may also be useful in methods for production of an enzymic product generated by an expressed polypeptide of the invention. Such methods may involve culturing the host cells of the invention in a medium suitable for expression of a recombinant polypeptide of the invention, optionally in the presence of additional enzymic substrate for the expressed polypeptide of the invention. The enzymic product produced may then be separated from the host cells or medium by a variety of art standard methods.

Plant cells may be transformed to comprise a genetic construct of the invention, and plant cells modified to alter expression of a polynucleotide or polypeptide of the invention.

Condensed tannin production in a plant may also be altered through methods of the invention. Such methods may involve the transformation of plant cells and plants, with a construct of the invention designed to alter expression of a polynucleotide or polypeptide involved in condensed tannin production in such plant cells and plants. Such methods also include the transformation of plant cells and plants with a combination of the construct of the invention and one or more other constructs designed to alter expression of one or more polynucleotides or polypeptides known to be involved in condensed tannin production in such plant cells and plants.

A number of plant transformation strategies are available (e.g. Birch, 1997, Ann Rev Plant Phys Plant Mol Biol, 48, 297). For example, strategies may be designed to increase expression of a polynucleotide/polypeptide in a plant cell tissue or organ in which it is normally expressed or to ectopically express a polynucleotide/polypeptide is a cell tissue or organ which it is not normally expressed. The expressed polynucleotide/polypeptide may be dervied from the plant species to be transformed or may be derived from a different plant species.

Alternatively, strategies may be designed to reduce expression of a polynucleotide/polypeptide in a plant cell, tissue or organ in which it is normally expressed. Such strategies are known as gene silencing strategies.

Temporal regulation of expression of a polynucleotide/polypeptide may also be achieved in the above strategies, through use of a suitable promoter.

Genetic constructs for expression of genes in transgenic plants typically include promoters for driving the expression of one or more cloned polynucleotide, terminators and selectable marker sequences.

The promoters suitable for use in the constructs of this invention are functional in a cell, tissue or organ of a monocot or dicot plant and include cell-, tissue- and organ-specific promoters, cell cycle specific promoters, temporal promoters, inducible promoters, constitutive promoters that are active in most plant tissues, and recombinant promoters. Choice of promoter will depend upon the temporal and spatial expression of the cloned polynucleotide, so desired. The promoters may be those normally associated with a transgene of interest, or heterologous promoters which are derived from genes of other plants, viruses, and plant pathogenic bacteria and fungi. Those skilled in the art will, without undue experimentation, be able to select promoters that are suitable for use in modifying and modulating plant traits using genetic constructs comprising the polynucleotide sequences of the invention. Examples of constitutive plant promoters include the CaMV 35S promoter, the nopaline synthase promoter and the octopine synthase promoter, and the Ubi 1 promoter from maize. Specific plant promoters
which are active in specific tissues, respond to internal developmental signals or external abiotic or biotic stresses are described in the scientific literature. Exemplary promoters are described, e.g., in WO 02/00894, which is herein incorporated by reference.

Exemplary terminators that are commonly used in plant transformation genetic construct include, e.g., the cauliflower mosaic virus (CaMV) 35S terminator, the *Agrobacterium tumefaciens* nopaline synthase or octopine synthase terminators, the *Zea mays* zin gene terminator, the *Oryza sativa* ADP-glucose pyrophosphorylase terminator and the *Solanum tuberosum* PI-II terminator.

Selectable markers commonly used in plant transformation include the neomycin phosphotransferase II gene (NPT II) which confers kanamycin resistance, the aadA gene, which confers spectinomycin and streptomycin resistance, the phosphinothricin acetyl transferase (*bar* gene) for Ignite (AgrEvo) and Basta (Hoechst) resistance, and the hygromycin phosphotransferase gene (*hpt*) for hygromycin resistance.

Use of genetic constructs comprising reporter genes (coding sequences which express an activity that is foreign to the host, usually an enzymatic activity and/or a visible signal (e.g., luciferase, GUS, GFP) which may be used for promoter expression analysis in plants and plant tissues are also contemplated. The reporter gene literature is reviewed in Herrera-Estrella *et al.*, 1993, *Nature* 303, 209, and Schrott, 1995, In: Gene Transfer to Plants (Potrykus, T., Spangenbert. Eds) Springer Verlag. Berline, pp. 325-336.

Gene silencing strategies may be focused on the gene itself or regulatory elements which effect expression of the encoded polypeptide. “Regulatory elements is used here in the widest possible sense and includes other genes which interact with the gene of interest.

Genetic constructs designed to decrease or silence the expression of a polynucleotide/polypeptide using the invention may include an antisense copy of the polynucleotide to be silenced. In such constructs the polynucleotide is in an antisense orientation with respect to the promoter and terminator.

An “antisense” polynucleotide is obtained by inverting a polynucleotide or a segment of the polynucleotide so that the transcript produced will be complementary to the mRNA transcript of the gene, e.g.,
Genetic constructs designed for gene silencing may also include an inverted repeat. An ‘inverted repeat’ is a sequence that is repeated where the second half of the repeat is in the complementary strand, e.g.,

5'-GATCTA........TAGATC-3'
3'-CTAGAT........ATCTAG-5'

The transcript formed may undergo complementary base pairing to form a hairpin structure provided there is a spacer of at least 3-5 bp between the repeated regions.

Another silencing approach involves the use of a small antisense RNA targeted to the transcript equivalent to an miRNA (Llave et al., 2002, Science 297, 2053). Use of such small antisense RNA corresponding to polynucleotides described herein is expressly contemplated. The term genetic construct as used herein also includes such small antisense RNAs.

Transformation with an expression construct, as herein defined, may also result in gene silencing through a process known as sense suppression (e.g. Napoli et al., 1990, Plant Cell 2, 279; de Carvalho Niebel et al., 1995, Plant Cell, 7, 347). In some cases sense suppression may involve over-expression of the whole or a partial coding sequence but may also involve expression of non-coding region of the gene, such as an intron or a 5' or 3' untranslated region (UTR). Chimeric partial sense constructs can be used to coordinately silence multiple genes (Abbott et al., 2002, Plant Physiol. 128(3): 844-53; Jones et al., 1998, Planta 204: 499-505). The use of such sense suppression strategies to silence the expression of a polynucleotide described above is also contemplated.

Other gene silencing strategies include the dominant negative approach and the use of ribozyme constructs (McIntyre, 1996, Transgenic Res, 5, 257)

Pre-transcriptional silencing may be brought about through mutation of the gene itself or its regulatory elements. Such mutations may include point mutations, frameshifts, insertions, deletions and substitutions.

The transgenic cells and plants of the invention may be from any species.
In a preferred embodiment the transgenic cells and plants of the invention are derived from forage plant species from a group comprising but not limited to the following genera: *Lolium, Festuca, Dactyliis, Bromus, Trifolium, Medicago, Pheleum, Phalaris, Holcus, Lotus, Plantain and Cichorium.*

In a more preferred embodiment of the invention the transgenic cells and plants are from the genera *Lolium and Trifolium.* Particularly preferred are the species *Lolium perenne and Trifolium repens.*

The following are representative publications disclosing genetic transformation protocols that can be used to genetically transform the following plant species: Rice (Alam et al., 1999, Plant Cell Rep. 18, 572); maize (U. S. Patent Serial Nos. 5, 177, 010 and 5, 981, 840); wheat (Ortiz et al., 1996, Plant Cell Rep. 15, 1996, 877); tomato (U. S. Patent Serial No. 5, 159, 135); potato (Kumar et al., 1996 Plant J. 9, 821); cassava (Li et al., 1996 Nat. Biotechnology 14, 736); lettuce (Michelmore et al., 1987, Plant Cell Rep. 6, 439); tobacco (Horsch et al., 1985, Science 227, 1229); cotton (U. S. Patent Serial No. 5, 846, 797 and 5, 004, 863); grasses (U. S. Patent Nos. 5, 187, 073 and 6, 020, 539); peppermint (Niu et al., 1998, Plant Cell Rep. 17, 165); citrus plants (Pena et al., 1995, Plant Sci.104, 183); caraway (Krens et al., 1997, Plant Cell Rep. 17, 39); banana (U. S. Patent Serial No. 5, 792, 935); soybean (U. S. Patent Nos. 5, 416, 011; 5, 569, 834; 5, 824, 877; 5, 563, 04455 and 5, 968, 830); pineapple (U. S. Patent Serial No. 5, 952, 543); poplar (U. S. Patent No. 4, 795, 855); monocots in general (U. S. Patent Nos. 5, 591, 616 and 6, 037, 522); brassica (U. S. Patent Nos. 5, 188, 958; 5, 463, 174 and 5, 750, 871); and cereals (U. S. Patent No. 6, 074, 877).

Methods are also provided for selecting plants with altered levels of condensed tannin production.

The expression of a polynucleotide, such as a messenger RNA, is often used as an indicator of expression of a corresponding polypeptide. Exemplary methods for measuring the expression of a polynucleotide include but are not limited to Northern analysis, RT-PCR and dot-blot analysis (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987). Polynucleotides or portions of the polynucleotides of the invention are thus useful as probes or primers, as herein defined, in methods for the identification of plants with altered levels of polypeptides involved in condensed tannin production. An altered level in a plant, of a polypeptide involved in condensed tannin biosynthesis may be used as an indicator of altered
condensed tannin biosynthesis in such a plant. The polynucleotides described above may be used as probes in hybridization experiments, or as primers in PCR based experiments, designed to identify such plants.

Alternatively antibodies may be raised against the polypeptides described above. Methods for raising and using antibodies are standard in the art (see for example: Antibodies, A Laboratory Manual, Harlow A Lane, Eds, Cold Spring Harbour Laboratory, 1998). Such antibodies may be used in methods to detect altered expression of polypeptides involved in condensed tannin production in plants. Such methods may include ELISA (Kemeny, 1991, A Practical Guid to ELISA, NY Pergamon Press) and Western analysis (Towbin & Gordon, 1994, J Immunol Methods, 72, 313).

These approaches for analysis of polynucleotide or polypeptide expression and the selection of plants with altered expression are useful in conventional breeding programs designed to produce varieties with altered condensed tannin content.

The invention further provides plants selected by a method of the invention. The selected plants may be from any plant species. In a preferred embodiment of the invention, the selected plants of the invention are from forage plant species from a group comprising but not limited to the following genera: *Lolium, Festuca, Dactyliis, Bromus, Trifolium, Medicago, Phleum, Phalaris, Holcus, Lotus, Plantain and Cichorium.*

In a more preferred embodiment of the invention the selected plants are from the genera *Lolium and Trifolium.* Particularly preferred are the species *Lolium perenne and Trifolium repens.*

The plants of the invention with altered condensed tannin production, may be transgenic plants or those selected based on analysis of polynucleotide/polypeptide expression, as described above. The plants of the invention may be grown and either selfed or crossed with a different plant strain and the resulting hybrids, with the desired phenotypic characteristics, may be identified. Two or more generations may be grown to ensure that the subject phenotypic characteristics are stably maintained and inherited. Plants resulting from such standard breeding approaches also form an aspect of the present invention.

Exemplary methods for measuring condensed tannin content in transgenic plants of the invention or plants selected through methods of the invention are provided in Example 1 below.
BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be better understood with reference to the accompanying drawings in which:

Figure 1 shows the known steps in the biosynthetic pathway for flavonoids. The figure is from Tanner et al., Proanthocyanidin biosynthesis in plants: Purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA J. Biol. Chem., In press (June 2003).

Figure 2 shows the hydrophobicity profile (Persson, B. & Argos, P. (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments J. Mol. Biol. 237, 182-192.) for the polypeptide represented in SEQ ID NO: 11.

Figure 3 shows a multiple sequence alignment of polypeptides, from several species, orthologous to those presented in SEQ ID NOs 2, 3 and 4 and illustrates a region (underlined) containing a consensus sequence, identified by the applicants, which is present in all of the sequences.

Figure 4 shows LC-MS traces for 70% acetone fractions from mature silique from transgenic plants (A) and control plants (B). The transgenic plants were transformed with an RNAi construct containing a sequence corresponding to SEQ ID NO:3. The peak labeled [K-3r, 7r]-1 represents kaempferol dirhamnoside. The numbers -1 and -2 indicate the change state in the mass spectrometer.

Figure 5 shows Proanthocyanidin levels normalized against dimer for transgenic and pHex control plants. The transgenic plants were transformed with an RNAi construct containing a sequence corresponding to SEQ ID NO:3.

EXAMPLES

The invention will now be illustrated with reference to the following non-limiting examples.
Example 1. Identification and isolation of polynucleotides involved in condensed tannin biosynthesis.

The applicants identified a mutant (SA1) of cotton (*Gossypium hirsutum*) that produces high levels of CTs in the fiber. In all commercial lines of cotton CTs are either undetectable in fiber or present at very low levels. Since cotton fibers are single-cell trichomes and fiber development in each boll is synchronous, high-CT fibers represent a uniquely uniform CT-producing tissue. The applicants have compared genes expressed during early stages of fiber development in mutant and control lines to identify genes involved in CT biosynthesis.

Plant material

High-CT (SA1) and low-CT (DPL90) lines of cotton (*Gossypium hirsutum*) were grown in the field at USDA-ARS SRRC, New Orleans, La. Fiber was harvested from both lines at 14 days post-anthesis (DPA), removed from the seeds and frozen in liquid nitrogen immediately.

Chemical analysis of fiber CTs

Fiber samples were extracted with aqueous acetone (70% v/v) and the extracts separated into two fractions by LH-20 sephadex chromatography and elution with 50% methanol and 70% acetone. The fractions were examined by HPLC on a C18 column with an acetonitrile/formic acid gradient, with PDA-UV and mass spectrometric detection. Monomeric CT precursors were eluted in the 50% methanol fraction while CT oligomers and higher polymers were eluted in the 70% acetone fraction.

RNA extraction

Total RNA was prepared from fiber samples harvested from SA1 and DPL90 lines at 14 DPA, using Concert™ Plant RNA reagent (Invitrogen) with a modified extraction protocol. Cotton fiber was ground under liquid N₂, and added to Concert™ reagent pre-equilibrated at 37°C (10mL Concert™ reagent/g tissue). The mixture was homogenized for 3 minutes using a polytron homogenizer (speed 20 000 rpm) and incubated 5 minutes at room temperature. RNA was then extracted from the homogenate as described by the manufacturer. The purified RNA
was extracted a second time with Concert\textsuperscript{TM} reagent (4mL Concert\textsuperscript{TM} reagent/mg RNA) and mRNA was purified using Oligotex resin (Qiagen).

**Construction and Screening of Subtracted cDNA libraries**

5 cDNA was prepared from the SA1 and DPL90 fiber mRNA samples and subtraction performed in both directions using the suppression subtractive hybridization (SSH) technique (Gurskaya et al., 1996, Anal Biochem, 240, 90-97; Diachenko et al., 1996, Proc Nat Acad Sci USA, 93, 6025). To reduce the background of non-differentially expressed clones, a mirror orientation selection (MOS) step (Rebrikov et al., 2000, Nucl Acid Res, 28 20) was carried out after subtractive hybridization. The cDNAs were ligated into a plasmid vector (pAL9, Euregene) and transformed into Escherichia coli. Individual clones were plated in 96-well plates; six plates were prepared for the forward subtracted (SA1-specific) library and one plate for the reverse-subtracted (DPL90-specific) library. Differential screening of plated clones was performed using forward-subtracted cDNA and reverse-subtracted cDNA samples as probes.

15 **Chemical analysis and construction of subtracted cDNA libraries**

Fiber samples were harvested from SA1 and DPL90 lines of *Gossypium hirsutum* at 14 DPA. Chemical analysis of fiber samples confirmed a significant difference in proanthocyanidin levels between SA1 and DPL90, and also indicated differences in chemical composition of the predominant proanthocyanadnin oligomers (see supplemental report). RNA was isolated from the SA1 and DPL90 fiber samples with Concert\textsuperscript{TM} Plant RNA reagent (Invitrogen) using an extraction protocol modified for cotton tissue, and subtracted cDNA libraries were constructed using standard SSH and MOS techniques, as described in methods. Both a forward-subtracted library (enriched for genes upregulated in SA1) and a reverse-subtracted library (enriched for genes upregulated in SA1 DPL90) were constructed. Both libraries were screened for differentially expressed genes by hybridization to labeled cDNA samples from SA1 and DPL90 14 DPA fibers.

30 **Sequencing and analysis of genes induced in high-CT (SA1) cotton fiber**

Differentially expressed clones induced in high-CT (SA1) cotton fiber were randomly selected and sequenced in one direction. A total of 163 partial cDNA sequences were obtained. The sequence data was clustered and the Genbank NR database was queried using the BLASTX heuristic search algorithm (Altschul et al., 1990, *J Mol Biol* 215, 403). Most genes (156 in total,
or 96%) showed high similarity to functionally annotated plants genes in the Genbank database (Table 1.). There were seven genes (4%) that appeared orthologous to unknown or putative polypeptides.

Known CT biosynthetic genes are abundant in the differentially expressed cloned from high-CT (SA1) cotton fiber.

Many of the genes with functionally annotated orthologs (65%) showed high sequence similarity to biosynthetic genes thought to be involved in the synthesis of CT precursors (leucoanthocyanadins and catechins) or related flavonoids. A smaller number (17%) appeared orthologous to genes considered unlikely to be involved in flavonoid biosynthesis. The remaining 29 genes (19%) appeared orthologous to genes that were functionally annotated but not highly characterized.

Identification of orthologous sequences and “in-silico” expression characterisation

The high abundance of biosynthetic genes for proanthocyanidin precursors in the SA1-specific library, together with the high levels of proanthocyanidins detected in SA1 fiber indicates that many of the unknown, putative or poorly characterized SA1-specific genes are likely to be novel genes necessary for condensed tannin biosynthesis. Publicly available EST databases and an in house ryegrass EST database was searched for expressed genes orthologous to cotton genes thought likely to perform novel functions in proanthocyanidin biosynthesis. Several cotton genes with likely orthologs expressed in species and tissues known to produce CTs were selected as the best candidates for genes involved in CT production. The orthologues of one such gene appears in libraries prepared from tissues known to produce CTs as shown in table 1 consistent with its role in CT production. The sequence of this polynucleotide from cotton is presented in SEQ ID NO: 2 and is predicted to encode the polypeptide sequence presented in SEQ ID NO:10. Orthologous sequences from Arabidopsis and ryegrass are presented in SEQ ID NOs 3 and 4 respectively and are predicted to encode polypeptides presented in SEQ ID NOs 11 and 12 respectively.

Table 1 shows “in-silico” expression data for sequences orthologous to that presented in SEQ ID NOs 2, 3 and 4.
Table 1. Identification of expressed genes orthologous to polypeptides involved in condensed tannin biosynthesis

The NCBI EST sequence database (2003-May-09), comprising unannotated polynucleotide sequence data for genes expressed in specified plant species and tissues, was searched for sequences orthologous to those presented in SEQ ID NOs 2, 3 and 4. Orthologue accession/ID numbers are listed, together with e values indicating the confidence level for each sequence match. Proposed orthologues in *Arabidopsis thaliana* identified by searching the GenBank, EMBL, DDBJ and PDB annotated polynucleotide sequence databases are also listed, with e values.

<table>
<thead>
<tr>
<th>Arabidopsis orthologue(s)</th>
<th>Arabidopsis seed/silique EST (high tissue)</th>
<th>Tree cambium EST (high tissue)</th>
<th>Lotus EST (high CT tissue)</th>
<th>Fruit / flower EST (high anthocyanin and/or CT tissue)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chitinase (glyc hyd fam 19) 1e-36 At3g16920.1</td>
<td>5e-38 AV561654.1</td>
<td>3e-37 G075P64Y</td>
<td>1e-34 BI420368. (nodule)</td>
<td>1e-26 EST586462</td>
</tr>
</tbody>
</table>

Subsequently published was a sequence F286 from *Gossypium hirsutum* (US Published Patent Application 2004/0049808). The sequence encodes a polypeptide having a deduced amino acid sequence including the partial deduced sequence of SEQ ID NO:10.

Since condensed tannins are sequestered in the plant vacuole, it is likely that the biosynthetic enzymes and transport proteins required for CT production will be targeted to the vacuole or the vacuolar membrane. The polypeptide sequence encoded by candidate polynucleotides were analysed for the presence of potential membrane-spanning regions and N or C-terminal vacuolar targeting signals.

The hydrophobicity profile (Persson, B. & Argos, P. (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments J. Mol. Biol. 237, 182-192.) of the polypeptide presented in SEQ ID NO: 11 is shown in Figure 2. As shown, this polypeptide possesses a hydrophobic N-terminus which could potentially function as a vacuolar targeting signal.
Example 2. Reduction of condensed tannin concentration and polymerization rate by RNAi in transgenic plants

Plant material

Vegetative tissues were harvested from Arabidopsis thaliana, ecotype Columbia.

Extraction of genomic DNA

Genomic DNA was extracted using standard procedures.

Selection of sequences for inserts in RNAi constructs

PCR primers were designed to amplify products from a single exon in genomic sequence corresponding to the sequence in SEQ ID NO: 3. Prior to amplification, the anticipated PCR product sequences were used to perform BLASTN searches (Basic local alignment search tool (1990). Altschul SF; Gish W; Miller W; Myers EW; Lipman DJ J Mol Biol 215: 403-10 ) of Arabidopsis cDNA databases to ensure that they were specific to the sequence in SEQ ID NO: 3, and did not contain stretches of more than 25 nucleotides of perfect homology to any other Arabidopsis sequence.

Amplification of inserts

Inserts for the RNAi constructs were amplified from Arabidopsis thaliana genomic DNA using High Fidelity Expand Taq DNA polymerase (Roche Molecular Biochemicals) according to the manufacturers instructions. The primers used to amplify an insert corresponding to sequence in SEQ ID NO: 3 were as follows:

attB1-TACGTTAGCGTTTTCAAGCGG (SEQ ID NO: 18)

attB2- ACCAGAACTCGGAGGACG (SEQ ID NO: 19)

Polynucleotide sequence of amplified insert is shown below:

TACGTTAGCGTTTTCAACCGCCGATTTTGCGGTGGATGACGCCACCGAAGAAGCATT
ACCCTCGCGCCACGATGTTTTCTCGGAAAATGGAAGCCAACACACAAGGACTG
CTGCAAAAACGAACTCGGCTTGGGAGGACCACATAAACGTTCTTACGCGGATCAA
ATCTGCAACACGGGATTGATAACGATGAGATGAAACACATTTGTTCTCATTATTTG
Cloning of sequences corresponding to the sequence in SEQ ID NO: 3 into RNAi vector

The attB-flanked PCR products amplified above were recombined into attP sites of the GATEWAY donor vector pDONR201 (Invitrogen) to generate a pENTRY clone with the target sequence flanked by attL sites.

The pEntry clone containing the attL-flanked target sequence was recombined with the attR sites of pTKO2 (HortResearch NZ Ltd) to generate the final destination vector with an inverted repeat of target sequence either side of an intron and under control of CaMV 35S promoter.

The destination vector is capable of replication in both Agrobacterium tumefaciens and E.coli, and is maintained by Spectinomycin (100µg/ml) selection. The right and left T-DNA borders delineate the region transferred to the plant genome during Agrobacterium-mediated transformation and include the CaMV35S driven inverted repeat of the target gene and the pNos-NptII-Nos3' region for kanamycin selection of transgenic plants.

Production of transgenic Arabidopsis plants

Transformation of Arabidopsis with the construct described above was by the floral dipping method of Clough and Bent (Clough and Bent, 1998, Plant J 16, 735).

Healthy Arabidopsis plants were grown under long days in pots in soil. Once flowering, the first bolts were clipped to encourage proliferation of many secondary bolts. Plants were the transformed approximately 4-6 days after clipping.

The destination vector described above was transformed into Agrobacterium tumefaciens strain GV3101 by standard procedures and cultured for 16 hrs at 28°C in LB broth with antibiotics for selection of the destination vector. This culture was then added to 200 ml LB. The Agrobacterium cells were then pelleted (by standard procedures) and resuspend to OD600 = 0.8 in 5% Sucrose solution, 200 ml for each two or three small pots to be dipped.
Prior to floral dipping, Silwet L-77 (Setre Chemical Company) was added to a concentration of 0.05% (500 ul/L) in the Agrobacterium solution. The above-ground parts of plants were dipped in Agrobacterium solution for 20 to 30 seconds, with gentle agitation. Dipped plants were placed under a cover for 16 to 24 hours to maintain high humidity. Plants were watered and grown normally until seeds became mature then watering ceased.

Dry seed was harvested and plated to select for T₁ transformants using antibiotic or herbicide selectable marker on 0.5X MS/0.8% tissue culture Agar plates with 50 ug/ml Kanamycin.

Putative transformants were transferred to soil and grown to produce T₂ seed.

Chemical analysis of flavonoid content in Arabidopsis plants transformed with RNAi construct targeting gene At3g16920 (Seq. ID 2)

Plant growth and silique collection

Arabidopsis plants transformed with an RNAi construct containing an insert sequence corresponding to the sequence in SEQ ID NO: 2. (T₁ generation) and control plants (transformed with the RNAi vector only) were grown in a controlled environment under standard long day conditions (22 °C, 70% relative humidity, and a 14 hour photoperiod at an irradiance of 650 μmol m⁻² s⁻¹). Mature siliques were collected from the primary inflorescences of approximately 300 plants transformed with the RNAi construct and approximately 300 control plants, at the onset of yellowing of the most mature silique. All green siliques below the 4th silique from the top of each primary inflorescence were collected, frozen and pooled.

Chemical analysis of flavonoids

Flavonoids were extracted from the silique pools by standard protocols, purified by liquid chromatography using an LH20 column, and characterized by HPLC/PDA/ESI-MS. The ESI-MS chromatograms are shown in figure 4. Peaks for epicatechin (EC) monomer, dimer and higher oligomers are indicated and in figure 4A. The designations -1 and -2 indicate the ion charge state. The peak labelled [K-3r,7r]-1 represents kaempferol dirhamnoside. The numbers in the upper left of the chromatograms are the m/z ratios of the single-ion channels. The peak area for epicatechin monomer, dimer and higher oligomers is listed in table 2. The steady-state concentration of each flavonoid species may be considered proportional to the peak areas. However quantification of pentamer and hexamer levels is impaired by interference from co-
eluting components and the pentamer and hexamer peak areas are therefore not considered reliable.

Total proanthocyanidin levels were approximately 2-fold lower in plants transformed with the RNAi construct compared to the control (table 2), indicating a partial disruption of CT accumulation. To compare the relative levels of epicatechin monomer and oligomers in both RNAi and control plants, each of the peak areas was normalized against the value for dimer. Normalised peak area values are listed in table 3 and plotted in figure 5. The monomer/dimer value is more than 2-fold higher in the RNAi plants compared to controls, whereas trimer/dimer and tetramer/dimer values show no change. The relative abundance of monomer in plants transformed with the RNAi construct indicates a slower rate of polymerization in the RNAi plants, and implies that the target gene is involved in CT polymerization. The low rate of polymerization in the RNAi plants may be due to incomplete gene silencing, or functional redundancy.

Table 2

70% acetone fractions, peak areas

<table>
<thead>
<tr>
<th></th>
<th>Chitinase At3g16920</th>
<th>Control (vector only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC monomer</td>
<td>215736</td>
<td>297409</td>
</tr>
<tr>
<td>EC Dimer</td>
<td>246706</td>
<td>887749</td>
</tr>
<tr>
<td>EC Trimer</td>
<td>89297</td>
<td>340022</td>
</tr>
<tr>
<td>EC Tetramer</td>
<td>4666</td>
<td>12507</td>
</tr>
<tr>
<td>EC Pentamer</td>
<td>400207</td>
<td>771943</td>
</tr>
<tr>
<td>EC Hexamer</td>
<td>64479</td>
<td>143792</td>
</tr>
<tr>
<td>Total</td>
<td>1021091</td>
<td>2453422</td>
</tr>
</tbody>
</table>

Table 3

70% acetone fractions, peak areas normalised against dimer

<table>
<thead>
<tr>
<th></th>
<th>Chitinase At3g16920</th>
<th>Control (vector only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC monomer</td>
<td>0.874465964</td>
<td>0.33501474</td>
</tr>
<tr>
<td>EC Dimer</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>EC Trimer</td>
<td>0.361957147</td>
<td>0.38301592</td>
</tr>
<tr>
<td>EC Tetramer</td>
<td>0.018913</td>
<td>0.014088</td>
</tr>
<tr>
<td>EC Pentamer</td>
<td>1.622202</td>
<td>0.869551</td>
</tr>
<tr>
<td>EC Hexamer</td>
<td>0.26135968</td>
<td>0.16197371</td>
</tr>
</tbody>
</table>
The above examples illustrate practice of the invention. It will be appreciated by those skilled in the art that numerous variations and modifications may be made without departing from the spirit and scope of the invention.

**Brief description of sequences in sequence listing**

SEQ ID NO:1 (polypeptide, consensus)

SEQ ID NO:2 (polynucleotide, *Gossipium hirsutum*, EGHSA1S005D10/5814R)

SEQ ID NO:3 (polynucleotide, *Arabidopsis thaliana*, AF412071/ AT3g16920)

SEQ ID NO:4 (polynucleotide, *Lolium perenne*, CLP0001001520-cE2)

SEQ ID NO:5 (polynucleotide, *Arabidopsis thaliana*, AY034935/ At1g05850)

SEQ ID NO:6 (polynucleotide, *Lolium perenne*, FLPB002449F07-b0FSP)

SEQ ID NO:7 (polynucleotide, *Musa acuminata*, Z99959)

SEQ ID NO:8 (polynucleotide, *Zea mays*, AY105990)

SEQ ID NO:9 (polynucleotide, *Oryza sativa*, AP003865)

SEQ ID NO:10 (polypeptide, *Gossipium hirsutum*, EGHSA1S005D10/5814R)

SEQ ID NO:11 (polypeptide, *Arabidopsis thaliana*, AT3g16920)

SEQ ID NO:12 (polypeptide, *Lolium perenne*, CLP0001001520-cE2)

SEQ ID NO:13 (polypeptide, *Arabidopsis thaliana*, At1g05850)

SEQ ID NO:14 (polypeptide, *Lolium perenne*, FLPB002449F07-b0FSP_20011203_fgenesh_l_pep)

SEQ ID NO:15 (polypeptide, *Musa acuminata*, Z99959)

SEQ ID NO:16 (polypeptide, *Zea mays*, AY105990)

SEQ ID NO:17 (polypeptide, *Oryza sativa*, BAC55635.1)

SEQ ID NO:18 (PRIMER)
SEQ ID NO:19 (PRIMER)

SEQ ID NO:20 \textit{(Arabidopsis thaliana)}
CLAIMS:

1. A method of altering condensed tannin production in a plant comprising transformation of a plant cell or plant with a genetic construct comprising a polynucleotide that alters the expression of a polypeptide comprising S X1 X2 C C N X3 T I X4 D X5 F X6 X7 X8 X9 F E X10 L F X11 X12 R N (SEQ ID NO:1) involved in condensed tannin production in the plant wherein X1=K, E, I or R, X2=Y or F, X3=H, Q, D or G, X4=S or T, X5=Y or F, X6=Q, E or K, X7=T, V or A, X8=Y or H, X9=Q, K or H, X10=D, N, Q or E, X11=A, S or P and X12=K, R or H and wherein the polynucleotide encodes the polypeptide or a fragment thereof or the polynucleotide is complementary to said polynucleotide encoding the polypeptide or fragment.

2. A method of altering condensed tannin production in a plant comprising transformation of a plant cell or plant with a genetic construct comprising a sequence selected from SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a functionally equivalent variant thereof, wherein said genetic construct encodes a polypeptide involved in condensed tannin production in plants.

3. A method as claimed in claim 2 wherein said sequence comprises SEQ ID NO:2 or a variant SEQ ID NO:2 having at least 70% homology with SEQ ID NO:2.

4. A method as claimed in claim 2 wherein said sequence comprises SEQ ID NO:3 or a variant of SEQ ID NO:3 having at least 70% homology with SEQ ID NO:3.

5. A method as claimed in claim 2 wherein said sequence comprises SEQ ID NO:4 or a variant of SEQ ID NO:4 having at least 70% homology with SEQ ID NO:4.

6. A method as claimed in claim 2 wherein the genetic construct comprises a sequence selected from SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a naturally occurring polynucleotide orthologue thereof wherein said genetic construct encodes a polypeptide involved in condensed tannin production in plants.

7. A method as claimed in claim 2 wherein said genetic construct comprises a sequence selected from SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a functionally equivalent variant wherein said variant has at least 90% homology with one of SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4.
8. A transgenic plant having altered condensed tannin content comprising a polynucleotide that alters the expression of a polypeptide comprising S X₁ X₂ C C N X₃ T I X₄ D X₅ F X₆ X₇ X₈ X₉ F E X₁₀ L F X₁₁ X₁₂ R N (SEQ ID NO:1) involved in condensed tannin production in the plant wherein X₁=K, E, I or R, X₂=Y or F, X₃=H, Q, D or G, X₄=S or T, X₅=Y or F, X₆=Q, E or K, X₇=T, V or A, X₈=Y or H, X₉=Q, K or H, X₁₀=D, N, Q or E, X₁₁=A, S or P and X₁₂=K, R or H and wherein the polynucleotide encodes the polypeptide or a fragment thereof or polynucleotide is complementary to said polynucleotide encoding the polypeptide or fragment.

9. A transgenic plant having altered condensed tannin content comprising a genetic construct comprising a sequence selected from SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a functionally equivalent variant thereof, wherein said genetic construct encodes a polypeptide involved in condensed tannin production in plants.

10. A plant as claimed in claim 9 or claim 10 wherein said sequence comprises SEQ ID NO:2 or a variant thereof having at least 70% homology with SEQ ID NO:2.

11. A plant as claimed in claim 10 said sequence comprises SEQ ID NO:3 or a variant thereof having at least 70% homology with SEQ ID NO:3.

12. A plant as claimed in claim 10 said sequence comprises SEQ ID NO:4 or a variant thereof having at least 70% homology with SEQ ID NO:4.

13. A plant as claimed in claim 9 or claim 10 wherein the genetic construct comprises a sequence selected from SEQ ID NO: 2, SEQ ID NO:3 and SEQ ID NO:4 or a naturally occurring polynucleotide orthologue thereof wherein said genetic construct includes a polypeptide involved in condensed tannin production in plants.

14. A plant as claimed in claim 9 or claim 10 wherein said genetic construct comprises a sequence selected from SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a functionally equivalent variant wherein said variant has at least 90% homology with one of SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4.

15. A plant as claimed in any one of claims 8-14 wherein said plant is a forage plant.
16. A plant as claimed in claim 15 wherein said plant is a plant from a genus selected from *Lolium, Festuca, Dactylis, Bromus, Trifolium, Medicago, Pheleum, Phalaris, Holcus, Lotus, Plantain and Cichorium*.

17. A plant as claimed in claim 16 wherein said plant is selected from the species *Lolium perenne* and *Trifolium repens*.

18. A plant as claimed in claim 17 wherein the plant of the species *Lolium perenne*.

19. A method for selecting plants with altered condensed tannin content comprising the steps of (a) contacting a sample from a plant with a polynucleotide selected from SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 and orthologues thereof, or a fragment of one of these, as a probe or primer (b) determining presence/content of complementary sequences to assess altered expression of a polypeptide involved in condensed tannin production in a plant; and (c) selection of a plant showing such altered expression.

20. A method for selecting plants with altered condensed tannin content comprising the steps of: (a) contacting a plant sample or plant with an antibody raised against a polypeptide encoded by a nucleotide sequence comprising SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or a naturally occurring orthologue thereof to assess altered expression of a polypeptide involved in condensed tannin production in a plant; and (b) selection of a plant shown such altered expression.

21. A plant selected by the method of claim 20 or claim 21.

22. A method as claimed in claim 1 or claim 2 wherein the genetic construct comprises a polynucleotide encoding the polypeptide or a functionally equivalent fragment thereof and the transformed plant overexpresses the polypeptide.

23. A plant as claimed in claim 8 or claim 9 wherein the genetic construct comprises a polynucleotide encoding the polypeptide or a functionally equivalent fragment thereof and the transformed plant overexpresses the polypeptide.

24. A plant as claimed in claim 8 or claim 9 wherein the genetic construct comprises a polynucleotide which causes a reduction of the expression of the polypeptide.
FIGURE 2
<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>5814Rcottonc &amp; SLDSLPFLPFL PLSLMPFLFPL</td>
<td>MEAKWLVL</td>
<td>SVAAMLVAL</td>
<td>ANCQESLKPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EGHSA18005D1 &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>MEAKWLVL</td>
<td>SVAAMLVAL</td>
<td>ANCQESLKPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP142071pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>MEAKWLVL</td>
<td>SVAAMLVAL</td>
<td>ANCQESLKPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AY034935pep &amp; MVTIRGSGTVE LLLVLLAVSF</td>
<td>LALVANGED. KTIK.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AY105990pep &amp; MRRKTRNKIV WTLALAAAV WLLVGTVTAL VLTAGTWKAK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLP00100152 &amp; MRRKTRNKIV WTLALAAAV WLLVGTVTAL VLTAGTWKAK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP003865pep &amp; MRRKTRNKIV WTLALAAAV WLLVGTVTAL VLTAGTWKAK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Z99959pep &amp; EVASLFLNCRL RDQQWSQDEL SIVGKMRAWS AVLLVVAVAF AAAAESKSK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>MEAKWLVL</td>
<td>SVAAMLVAL</td>
<td>ANCQESLKPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td>100</td>
<td>5814Rcottonc &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EGHSA18005D1 &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP142071pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AY034935pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AY105990pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLP00100152 &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP003865pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>101</td>
<td>150</td>
<td>5814Rcottonc &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AY105990pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLP00100152 &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FLPB002449F &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP003865pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>151</td>
<td>200</td>
<td>5814Rcottonc &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AY105990pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLP00100152 &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FLPB002449F &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP003865pep &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus &amp; SLDLPFLPFL PLSLMPFLFPL</td>
<td>SKYCCNHTIS</td>
<td>DYPQYQYQED</td>
<td>LFAKRN. TPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5814Rcottonc</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGSHA1S005D1</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF412071pep</td>
<td>GALPVYWYN</td>
<td>YGQTGEALKV</td>
<td>DLLSHPEYLE</td>
<td>NNATLAFAQAA</td>
<td>IWRWMTPPKK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY034935pep</td>
<td>GALPYYWFN</td>
<td>YGAAGEALKA</td>
<td>DLLNHPEYIE</td>
<td>QNATLAFAQAA</td>
<td>IWRWMTPIKR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY105990pep</td>
<td>GALPVYWYN</td>
<td>YGAAGDGKIA</td>
<td>DLLIHPEYLE</td>
<td>QNATLAFMAA</td>
<td>MWRWMTPIKK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLP000100152</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP03865pep</td>
<td>GALPVYWYN</td>
<td>YGIIGGQIKQ</td>
<td>DLLNHPELLE</td>
<td>QNATLAFEAA</td>
<td>IWRWMTPMKR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z099599pep</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5814Rcottonc</th>
<th>~~~~~~~~~~~~~</th>
<th>~~~~~~~~~~~~~</th>
<th>~~~~~~~~~~~~~</th>
<th>~~~~~~~~~~~~~</th>
<th>~~~~~~~~~~~~~</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGSHA1S005D1</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
</tr>
<tr>
<td>AF412071pep</td>
<td>HLPSAHDFVF</td>
<td>GKWPTKNDT</td>
<td>AAAPGFGFA</td>
<td>TINVLYGDQI</td>
<td>CNSGFDNDEM</td>
</tr>
<tr>
<td>AY034935pep</td>
<td>AQSASHDIFV</td>
<td>GNWPTKNDT</td>
<td>LSATGFSGS</td>
<td>TNIVLYGEYT</td>
<td>CGQGSIDPM</td>
</tr>
<tr>
<td>AY105990pep</td>
<td>SQPSAHDAFV</td>
<td>GNWPTKNDT</td>
<td>LSRKPGFGGA</td>
<td>TMNILYGESI</td>
<td>CGKGYVDAM</td>
</tr>
<tr>
<td>CLP000100152</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
</tr>
<tr>
<td>AP03865pep</td>
<td>KQPSAHDFVF</td>
<td>GNWPTKNDT</td>
<td>LSKRYPFQGA</td>
<td>TMNIVLYDLI</td>
<td>CGQGSIDKM</td>
</tr>
<tr>
<td>Z099599pep</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
</tr>
<tr>
<td>Consensus</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5814Rcottonc</th>
<th>~~~~~~~~~~~~~</th>
<th>~~~~~~~~~~~~~</th>
<th>~~~~~~~~~~~~~</th>
<th>~~~~~~~~~~~~~</th>
<th>~~~~~~~~~~~~~</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGSHA1S005D1</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
</tr>
<tr>
<td>AF412071pep</td>
<td>NNIVSHYLYY</td>
<td>LDLIGVGREE</td>
<td>AGPHKSLCA</td>
<td>DQEPFSSSH</td>
<td>APPSSGSSS~</td>
</tr>
<tr>
<td>AY034935pep</td>
<td>NNIISLYYF</td>
<td>LDLMIGRED</td>
<td>AGRNDELSCA</td>
<td>EQKFNPSST.</td>
<td>VPSSS~-------</td>
</tr>
<tr>
<td>AY105990pep</td>
<td>NVIISHQYY</td>
<td>LDLMGVREH</td>
<td>SG..DNPDCIA</td>
<td>EQAPFNPSSP</td>
<td>TDDQQQQQG</td>
</tr>
<tr>
<td>CLP000100152</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
</tr>
<tr>
<td>AP03865pep</td>
<td>NVIIVSHQHY</td>
<td>LDLMGVGSK</td>
<td>AG..DNLDCA</td>
<td>DQVAPFNPSK</td>
<td>NLDSS~-------</td>
</tr>
<tr>
<td>Z099599pep</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
</tr>
<tr>
<td>Consensus</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
<td>~~~~~~~~~~~~~</td>
</tr>
</tbody>
</table>

**FIGURE 3**
FIGURE 5
<220>
<221> MISC_FEATURE
<222> (2) . . . (2)
<223> X is K, E, I or R

<220>
<221> MISC_FEATURE
<222> (3) . . . (3)
<223> X is Y or F

<220>
<221> MISC_FEATURE
<222> (7) . . . (7)
X is H, Q, D or G

MISC_FEATURE
(10)..<10)
X is S or T

MISC_FEATURE
(12)..<12)
X is Y or F

MISC_FEATURE
(14)..<14)
X is Q, E or K

MISC_FEATURE
(15)..<15)
X is T, V or A

MISC_FEATURE
(16)..<16)
X is Y or H

MISC_FEATURE
(17)..<17)
X is Q, K or H
<220>

<221> MISC_FEATURE

<222> (20) .. (20)

<223> X is D, N, Q or E

<220>

<221> MISC_FEATURE

<222> (23) .. (23)

<223> X is A, S or P

<220>

<221> MISC_FEATURE

<222> (24) .. (24)

<223> X is K, R or H

<400> 1

Ser Xaa Xaa Cys Cys Asn Xaa Thr Ile Xaa Asp Xaa Phe Xaa Xaa Xaa
1  5  10  15

Xaa Phe Glu Xaa Leu Phe Xaa Xaa Arg Asn
20  25

<210> 2

<211> 337

<212> DNA

<213> Gossypium hirsutum

<400> 2

ccatcctaat acgactcact atagggtctg agcgccgcgcc cgaggcaggtgc gaccaccaca 60
ttttgttctcg tatcctttgat aatgggaagcc aatgggctgg ttctcttcttc atgtggcgga 120
atgtgtgtgg cactggccaa ctggcaagaa tgttggtaagc cattggtgaa gatggtaaag 180
ggcaagaagc tgtgtgacaa agggtagggaa tgtaaaggt gtgcacagta ttggtgcaac 240
catcatttt ctgattacct ccaaaacctac cagtttgagg accttttttg acagcgttac 300
agcgggttag caccgaggttg tgggttcttg gattacc 337

<210> 3
<211> 1204
<212> DNA
<213> Arabidopsis thaliana

<400> 3
gacgaagaaa aagacaaaaa aaaaatcacc atggtttctg aaccgtctatt ctgccctatttg 60
cctctcaggg ttgcaattggt tgtgctttcag aaccggaacc ttgctcaacgc tgaagattcc 120
 gagccttctga gttcaacgag aaagcctctg gtagaagatg tgaagagtt aaattatatgc 180
 gacaagagat gggaatgcga aagtagtgggt cagattttggt ggacaccatac atctctccgat 240
ttttttgaag cctatcattt cggagatca ttttcaagct gtaacagtctc agtggtctac 300
 ggcggcttgcggt ttgggatttt ccggtttcttc attactcgcgct ctgcccagat caaccgctttct 360
gttccggtatatctgc gaaactttcag gggatgaagg aagtctgctgc gttctttggtc 420
cacctgagaa acacaaacttc atgcgtggtac ggggctgtgc caggccgccacc attagcttttg 480
 ggaattgttca aacaaacagga gatgagtcct catcagccttt acctatcaca tttcttctcc 540
tggacaccttc ttcgctcctct taccaggtggtc gaggagtttc acgcgggttg 600
tggactatcattt gtagcttgaag attcgcgtggcg tcggatcgc gccatcgaat aacccgggag 660
gatcgcttctc agaagaacagt tacgtagtgc tctcaagcgc cgatattgcag gtggatgacg 720
cacgcacaag aacgattcctgc gatgttttccg tcggaaatgt gaagcccaacac 780
 aagaatgaca ctgctgtcagaa aacgacttcg gcgttggagg cacagctataa cgcttttcttac 840
ggcgatcaca ccggtacagcg gaggattggt gtagatcgat acgatgagat tgaacaccatact gttcttcctcat 900
 tctcttgatatcttcagctt gatagttgtta ggacggagaa aagcttggcttc tcacagagaag 960
cgcgtcatcg cacatcagaa accttttctct tctcttctctct tgcctctcccc gatgttcggt 1020
tctcttttct ctgtagaagg tgtttataaccct ctggcttgcggt gacttttgaa aagaaagtgt 1080
 gacttttggt tgtatagtttgc tttgctgtggt tgggttttata tatttttgaa aataatactgt 1140
 agacgaagcag tcatatatgtt cccacaccatt tctcttttctc aaaaatattaaa tatatatattgc 1200
atgc

<210> 4
<211> 693
<212> DNA
<213> Lolium perenne

<400> 4
ccttccttcc ccocgtccgg caacgcttctgc tggcaacctc gcttcctcttt caccacccatc 60
gccaatttca cctctcttac ggcaccaccg agatctcacc cggccagcta cggcatctcc 120
aagatccgaa ccacaccacg ctagcaagat aatcaagacg caagatccag agcagttcga 180
gctgcggcgg ctcctgcggc gcgtctcaga gcaaacaacc aagatgaagc gcaagaagccg 240
gaacaagac tgtgctgaag tgtgtgctgac gggggtggcg gcggtgtgctg tcggcggccac 300
gatgcccgttc atctgcaagc cgagggcagg gaaggtgaag ctcaagaggct gcaggggaa 360
gatgtgcaac aaggggtggg agtgtgctcgc cagcaagtaac tgctgcaacag agccaactac 420
cgaaccttc aaggtatatc agtttgagaa cctcttccac aagcgeaact aacccggtggc 480
caaagcggtgc ggcctctggg acactcgagc ctctcataac gcagcgcgcca tgtgcagggc 540
ccagggcttc tcggccaccc gcgggtgcagc gatcggagat aggagatctc gcgcctcttc 600
gcgcacagtc gcgcgcgaaga caatctgtgg ttctcgggtg gcaccccggtg ctccgctgac 660
gtggggtgccg tgcctacaacc acagagctgac acc 693

<210> 5
<211> 1322
<212> DNA
<213> Arabidopsis thaliana

<400> 5
aagaatctct ctcttttttt ctcgccgaacg aagcacaag aggtatatttt gtctggtttgaa 60
ttcacaagct atgttttctcg aacaatatgt ctagatctgac ttgctgcttg 120
tagctggtat ttgtccataat gtgtgaagct atgggtgagaa tcatggcatac 180
atattttgttc tgcgtggctgt atccatcttg gccttggtttg ccaatgagac gcgcacacag 240
attaagatga aagaatgtgag ggagaataag gtgtgcacgc aaggtgagga atgcagctgg 300
tggtctaatc atcgtgctaa ccagacgata tcagattact tctagttttt ctctcttgag 360
caactcttttt ccacagaccc caaccccatt gcctcatgtg ctggttcttttg gcactcaccg 420
tctttccattc tctctcttttc cttctttggg tgtgttatgctt gcgtgctgctg 480
tctatggcct tcggaacagt ggtgcttttt ctcggtcatgc tgcggagaca aacgctctgt 540
ggtctatggag ttggacaagc agggtgcttttc tgggtgctgac tgtgctacaacc cgaaggaattt 600
gaccccagagc aacatctcgtc tgcaggtgcc gtaaugacttc agagcctgtcgc 660
gctgaaactc acggagccgg tgtgcttacc attactgtaa acctcacta cggagacgct 720
ggggaagcccc tgaagacctg ctcttggggc cacccctgtg acatgagc aacgcgcacc 780
tctgctttcc tgcagctgaca atggagatgg atgacgccaa tcagagacgc tgcggcctca 840
gctcagacact cttctggttg aaactgagaa cctacaaaga acgacacttt gttcaagcgt 900
ggcccgactt ttggcagcag ctagaacgctc ctctacgggag agtacacatg tggtaaaggt 960
tccatggatc caatgaacaa cataatcctca cactatttat actttccttgac cctcatggtt 1020
atggagaagag aagacgaggc accaaacgat gagctcagct gcgcgaacac gaaacccttctc 1080
aaccccttcaac ctgtacccct cttctccctct ctgtaagtct ttaatgaatat cggccgttgct 1140
ctcaagtctc tctgtgtatca tgttacagtaa taagttgaaat tgcctgtgtt gaaaaaagaa 1200
ggaagtgattt ttggtccctc tagaacttat aatgcaaacac acattacatt ttatatatat 1260
ctgtatgtaa aacataatat atticgtgtaa ttggaagaat gtgtgcaaaa aaaaaaaaa 1320
aa 1322

<210>  6
<211>  685
<212>  DNA
<213>  Lolium perenne

<400>  6
gtccgttttg aaagttata gaagtcactg attgacgact tctgatgtta tttgccaggt 60
gcagagctac tccctaaaca tatattacag ctaacctcttg ttgagcatat ttttttttt 120
tatgtctctca ttttagtttg tcaacccctct ctgcatgtatat cagcatctgtg gtgggtgttt 180
gtgcctctgct cagagttgatt tcctagaatcc gaaattaggg tttgtgtgtga gtagattctc 240
ccaggggtta gcgctccagca agccctataac ccaacactgttt actgcctccatt gttctataca 300
atcctataat ttggtgtcgg ggaagggggtg agcgtacatag aggtctctggc gcgcacagct 360
ccgaggaatg gcgcacactc cttgctgccc gttatctttgc cgccggttggt gcgcacacccg 420
cggggctcag attggggtggc gcgcggcagatg aagccctggtgt gagctccagaa gcgcggcggc 480
tgggcgatgg agtggttttgg cctgggggaag agtcctctga agttggtgtgc cttgaagtgg 540
tggtgatgg cccggtgca gcagacacct gtcgcgcagtggc actccacacc ggtgctgcac 600
gccttgtgct gcgcctctct cctctctgcct cgcgcctcgc cgcgcgcggc gcgcacaccc 660
agcagaaatgc cgcgctctgac gcg 685

<210>  7
<211>  715
<212>  DNA
<213>  Musa acuminata

<220>
misc_feature
(620) ..(620)
n is a or g or c or t/u, or unknown

misc_feature
(635) ..(635)
n is a or g or c or t/u, or unknown

misc_feature
(703) ..(703)
n is a or g or c or t/u, or unknown

acgagttgcgtcctctccttcgaagagatacgatcctgtctgtcaagcttctcgtctgtcctgatcaagaac 60
acgagttgcgtcctctccttcgaagagatacgatcctgtctgtcctgatcaagaac 60
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
tcctctcttcggagagatgasggggagatgtggtgccggtgtcctgctgtggtgcgtgtgtagctgtcctgatcaagaac 120
ctgagtgtgct gccgtgcccgg gccgtgctccc cccctgacag cctgccgcca cggacaacca 60
catctctccc ctgtggacca acgtgtagtg ccaacctgctt ttcacccttc tctcaetata 120
cctcacaact ctctccatcg attaatctcc tcccccctccaa caaccgcaag atcttcagct 180
acgctctctct ccgctgcggt tgtatatccgc tttcatccaga tccagccctg aagccccagc 240
cggcttcgag atccagctag ttctctatca agcatcgcgcc gggagcgaca gggggtctgc 300
cggcggaga gagaagtagaag cgcaagacgc ggaacaagag atacgtatgg acgctgggcc 360
tggctgcagt ggaattttcg tgggagccga cggattgccct ggtgtcctacag gcggggagct 420
gagagccgaa gataaaagag tgccaggaga agatcctgtaa caaggggtgg gatgctgcgg 480
ggacgaagtca ctgctgcaac gacacatcc caacagtttc caaggtgtac aagtgcgaga 540
accttctgc gaagcggcaac acccccttcgc cggacgccct cgggttctctg gactacagg 600
cctctcatac ccgggcgccc cttctcgggc cccagggggt ttgctacacc gcgggcaagc 660
agatgcagat ggtgagctcc tgcgcctttcc tggagcgacgt cgccggaacag acttcgtcttg 720
ggtagccccgt ggcgacggcg gggcccacgg cggctgggggt tgtgtacaac cagcagatga 780
gccccacacg caacctctgc gagaagacct aacccagcga ccctctgctg cgggccgctg 840
agtaacgcgg ccggagcgcgc atctcctgtct actggaacat caactacggc gctgcccgttg 900
agcgagatcga ccggacctcg ctccaacacc cagagttactt ggcagcagac gcgcagcttg 960
catcctgacgc gccgagttcg ggcttggtatga cggccgtaca gaagagcccaac cggctgggcc 1020
aacacgccct cgggacccct ccagagccga cccagctccc aaccgcctgc 1080
cctggtttccg cgccacccatgc acatacctct aaccgcagtc ccagccgcggc aagggaacgc 1140
tcgacgcctat gaaagttata atctcctgcac accagttata cctctgtacct atggggtcgc 1200
ggcgtgacca cttctgagacc aaccgtgatt gcgcgcagac gcgcacggccc acccctccca 1260
gccgcagcga taggacagaa gacagcagact caggaagacta agacagactc atgggcccct 1320
tgcactgtat caacagttccca cagccgatat tggagttaga cagctttcag tttgtgtaga 1380
ggtgcctgcac tgtggattgt ccaggggtttg aagagacttag tttcttgctg atggagattt 1440
cagatgtttgt gtttatctac tttcttccag cttcatagata tgttgcatttg gtctttcctttg 1500	ttctgacagc taggttatag acatgttttta agagaactgca acatccccag acccctttttc 1560
cgccttggaaa aaaaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1619

<210> 9
<211> 951
<212> DNA
<213> Oryza sativa
<400> 9
tagaggaaggt cgcgtgcggc gcggcgccgg cgtatcttct gcgcgtgtcct gctggtgaggg 60
ctacttcttg ggccggaggg gaggagggcag aagagccgca gcggccagca gaggccgag 120
gcgccgaca aggggtgggaa tgtcaacgcca acgcgccgag cgtctctcacc 180
gacactctca aggcgtactcag agtccgaggag ctctctcccc acgcacacgg cgcctgctct 240
gcccacgccc cggggtctcgg ggacactcacta cccgccccgc gcgtctgttgc 300
ccccgccccg tggcacctacac gcgggggaag gaggtgggga tgaagggaggt cgccgccttc 360
cctgaccacg cgggccccac gaacctctgt ggatatagtg tgcgaactgg tgtccatttg 420
ggtctctgcccc ccatagaattg agccaaagcc aataagctattg tgaacaacgc 480
aacgaaattg atctctgtgt tgaaggggtc gatattattg tgcgaggccgc tctttctgtt 540
tactgaacct tcaacatattg tattatcgtt caggttatca aacaggacact gttgaaaccac 600
cagagtttctttgagaag ctagctgctctg cctccatgctggagttgagggc 660
aptttacactga agagagcacta ccccagcccc caatgtgctgt tggggtgtcg ctgaaacca 720
acaaataagct acacattttggtt cttgctctgt gcggctcaat gagatcccttg 780
tatggcgtac tcatctctggg cagggcctcc gtccgataaga tgaacgctcat tgtatccac 840
taccaacatttt acacattttcag cggggttgtt ggtgtggtga caagctggaga taaccttgct 900
tgtgctgacc agttgtgcttt caatccatcc tcccaagatgt tggactctcttt a 951

<210> 10
<211> 85
<212> PRT
<213> Gossypium hirsutum

<400> 10
Met Glu Ala Lys Trp Leu Val Leu Phe Ser Val Ala Ala Met Leu Val 1 5 10 15
Ala Leu Ala Asn Cys Gln Glu Ser Leu Lys Pro Leu Val Val Lys Met Val 20 25 30
Lys Gly Lys Lys Leu Cys Asp Lys Gly Trp Glu Cys Lys Gly Trp Ser 35 40 45
Lys Tyr Cys Cys Asn His Thr Ile Ser Asp Tyr Phe Gln Thr Tyr Gln 50 55 60
Phe Glu Asp Leu Phe Ala Lys Arg Asn Thr Pro Val Ala His Ala Val 65 70 75 80
Gly Phe Trp Asp Tyr

<210> 11
<211> 333
<212> PRT
<213> Arabidopsis thaliana

<400> 11
Met Val Ser Lys Pro Leu Phe Ser Leu Leu Leu Leu Leu Thr Val Ala Leu
1   5    10   15
Val Val Phe Gln Thr Gly Thr Leu Val Asn Ala Glu Asp Ser Glu Pro
20  25   30

Ser Ser Ser Thr Arg Lys Pro Leu Val Lys Ile Val Lys Gly Lys Lys
35  40   45

Leu Cys Asp Lys Gly Trp Glu Cys Lys Gly Trp Ser Glu Tyr Cys Cys
50  55   60

Asn His Thr Ile Ser Asp Phe Phe Glu Thr Tyr Gln Phe Glu Asn Leu
65  70   75   80

Phe Ser Lys Arg Asn Ser Pro Val Ala His Ala Val Gly Phe Trp Asp
85  90   95

Tyr Arg Ser Phe Ile Thr Ala Ala Ala Glu Tyr Gln Pro Leu Gly Phe
100 105 110

Gly Thr Ala Gly Glu Lys Leu Gln Gly Met Lys Glu Val Ala Ala Phe
115 120 125

Leu Gly His Val Gly Ser Lys Thr Ser Cys Gly Tyr Gly Val Ala Thr
130 135 140

Gly Gly Pro Leu Ala Trp Gly Leu Cys Tyr Asn Lys Glu Met Ser Pro
145 150 155 160

Asp Glu Leu Tyr Cys Asp Asp Tyr Tyr Lys Leu Thr Tyr Pro Cys Thr
165 170 175

Pro Gly Val Ser Tyr His Gly Arg Gly Ala Leu Pro Val Tyr Trp Asn
180 185 190
Arg Asn Tyr Pro Val Ala Lys Ala Val Gly Phe Trp Asp Tyr Gln Ala 85
Phe Ile Thr Ala Ala Ala Met Phe Glu Ala Gln Gly Phe Cys Thr Thr 100
105
Gly Gly Phe Glu Met Gln Met Lys Glu Leu Cys Ala Phe Leu Ala His 115
120
125
Val Gly Ala Lys Thr Ser Cys Gly Phe Gly Val Ala Thr Gly Gly Pro 130
135
140
Leu Ala Trp Gly Leu Cys Tyr Asn His Glu Leu Ala 145
150
155
<210> 13
<211> 321
<212> PRT
<213> Arabidopsis thaliana

<400> 13
Met Val Thr Ile Arg Ser Gly Ser Ile Val Ile Leu Val Leu Leu Ala 1
5
10
15
Val Ser Phe Leu Ala Leu Val Ala Asn Gly Glu Asp Lys Thr Ile Lys 20
25
30
Val Lys Lys Val Arg Gly Asn Lys Val Cys Thr Gln Gly Trp Glu Cys 35
40
45
Ser Trp Trp Ser Lys Tyr Cys Asn Gln Thr Ile Ser Asp Tyr Phe 50
55
60
Gln Val Tyr Gln Phe Glu Gln Leu Phe Ser Lys Arg Asn Thr Pro Ile 65
70
75
80
Ala His Ala Val Gly Phe Trp Asp Tyr Gln Ser Phe Ile Thr Ala Ala 85
90
95
Ala Leu Phe Glu Pro Leu Gly Phe Gly Thr Thr Gly Gly Lys Leu Met 100
105
110
115
120
125
Page 12
Ser Cys Gly Tyr Gly Val Ala Thr Gly Gly Pro Leu Ala Trp Gly Leu
130   135   140

Cys Tyr Asn Arg Glu Met Ser Pro Met Gln Ser Tyr Cys Asp Glu Ser
145    150    155    160

Trp Lys Phe Lys Tyr Pro Cys Ser Pro Gly Ala Glu Tyr Tyr Gly Arg
165   170

Gly Ala Leu Pro Ile Tyr Trp Asn Phe Asn Tyr Gly Ala Ala Gly Glu
180    185    190

Ala Leu Lys Ala Asp Leu Leu Asn His Pro Glu Tyr Ile Glu Gln Asn
195    200    205

Ala Thr Leu Ala Phe Gln Ala Ala Ala Ile Trp Arg Trp Met Thr Pro Ile
210    215    220

Lys Arg Ala Gln Pro Ser Ala His Asp Ile Phe Val Gly Asn Trp Lys
225    230    235    240

Pro Thr Lys Asn Asp Thr Leu Ser Lys Arg Gly Pro Thr Phe Gly Ser
245    250    255

Thr Met Asn Val Leu Tyr Gly Glu Tyr Thr Cys Gly Gln Gly Ser Ile
260    265    270

Asp Pro Met Asn Ile Ile Ser His Tyr Leu Tyr Phe Leu Asp Leu
275    280    285

Met Gly Ile Gly Arg Glu Ala Gly Pro Asn Asp Glu Leu Ser Cys
290    295    300

Ala Glu Gln Lys Pro Phe Asn Pro Ser Thr Val Pro Ser Ser Ser Ser
305    310    315    320

Ser

<210>  14

<211>  118

<212>  PRT

<213>  Lolium perenne

<400>  14
 Ala Glu Ala Arg Arg Glu Lys Gly Gly Asp Lys Ala Cys Asp Thr Gly  
  1  5  10  15

 Trp Glu Cys Ser Gly Ser Arg Phe Cys Cys Asn Gly Thr Ile Thr Asp  
  20  25  30

 Tyr Phe Lys Ala His His Phe Glu Glu Leu Phe Pro Arg Arg Asn Asp  
  35  40  45

 Ser Ile Ala His Ala Thr Gly Phe Trp Ser Tyr Gln Ala Phe Ile Ala  
  50  55  60

 Ala Ala Ala Gln Phe Glu Pro Arg Gly Phe Gly Thr Thr Gly Gly Lys  
  65  70  75  80

 Asp Thr Gly Ile Lys Glu Val Ala Ala Phe Leu Gly His Val Gly Ala  
  85  90  95

 Arg Thr Ser Cys Thr Leu Leu Pro Ser Pro His Leu Lys Ile Arg Ile  
 100 105 110

 Cys Met Asn Asn Gly Gln 115

<210> 15

<211> 172

<212> PRT

<213> Musa acuminata

<400> 15

 Glu Val Ala Ser Leu Phe Asn Cys Leu Arg Arg Asp Gln Ile Trp Ser  
  1  5 10 15

 Asp Gln Glu Leu Ser Ile Val Gly Lys Met Arg Ala Trp Ser Ala Val  
  20  25  30

 Leu Leu Val Val Ala Ala Phe Ala Ala Ala Glu Ala Ser Lys  
  35  40  45

 Ser Lys Pro Lys Val Cys Asn Lys Gly Trp Glu Cys Ser Gly Ser Ile  
  50  55  60

 Tyr Cys Cys Asn Asp Thr Ile Ser Asp Tyr Phe Lys Val Tyr Gln Phe  
  65  70  75  80


Glu Asn Leu Phe Ser Lys Arg Asn Ser Pro Val Ala His Ala Val Gly 85 90 95

Phe Trp Asp Tyr Gln Ser Phe Ile Leu Ala Ala Ser Val Tyr Glu Pro 100 105 110

Leu Gly Phe Gly Thr Thr Gly Asp Lys Val Thr Lys Met Gln Glu Val 115 120 125

Arg Pro Pro Ser Ser Gly His Val Arg Gln Ala Arg Pro Pro Cys Trp 130 135 140

Ile Trp Cys Leu Pro Leu Val Ala His Leu His Gly Asp Tyr Ala Ile 145 150 155 160

Ile His Glu Asn Glu Pro Gln Ala Gln Glu Tyr Cys 165 170

<210> 16
<211> 328
<212> PRT
<213> Zea mays

<400> 16

Met Lys Arg Lys Thr Arg Asn Lys Ile Ile Val Trp Thr Leu Ala Leu 1 5 10 15

 Ala Ala Val Ala Ile Leu Val Gly Gly Thr Ile Ala Leu Val Leu Thr 20 25 30

 Ala Gly Thr Trp Lys Ala Lys Ile Lys Lys Ser Gln Glu Lys Ile Cys 35 40 45

Asn Lys Gly Trp Glu Cys Ser Gly Ser Lys Tyr Cys Cys Asn Asp Thr 50 55 60

Ile Thr Asp Phe Phe Lys Val Tyr Lys Phe Glu Asn Leu Phe Ala Lys 65 70 75 80

Arg Asn Thr Pro Val Ala His Ala Val Gly Phe Trp Asp Tyr Gln Ala 85 90 95

Phe Ile Thr Ala Ala Ala Leu Phe Glu Pro Gln Gly Phe Cys Thr Thr 100 105 110

Gly Gly Lys Gln Met Gln Met Met Glu Leu Cys Ala Phe Leu Gly His

Page 15
Val Gly Ala Lys Thr Ser Cys Gly Tyr Gly Val Ala Thr Gly Gly Pro
115                             120                             125
130                             135                             140
Thr Ala Trp Gly Leu Cys Tyr Asn His Glu Met Ser Pro Asp Gln Thr
145                             150                             155                             160
Tyr Cys Asp Lys Thr Tyr Thr Gln Tyr Pro Cys Val Glu Gly Ala Glu
165                             170                             175
Tyr Tyr Gly Arg Gly Ala Ile Pro Val Tyr Trp Asn Tyr Asn Tyr Gly
180                             185                             190
Ala Ala Gly Asp Gly Ile Lys Ala Asp Leu Leu His His Pro Glu Tyr
195                             200                             205
Leu Glu Gln Asn Ala Thr Leu Ala Phe Met Ala Ala Met Trp Arg Trp
210                             215                             220
Met Thr Pro Ile Lys Ser Gln Pro Ser Ala His Asp Ala Phe Val
225                             230                             235                             240
Gly Asn Trp Lys Pro Thr Lys Asn Asp Thr Leu Ser Lys Arg Leu Pro
245                             250                             255
Gly Phe Gly Ala Thr Met Asn Ile Leu Tyr Gly Glu Ser Ile Cys Gly
260                             265                             270
Lys Gly Tyr Val Asp Ala Met Asn Val Ile Ile Ser His Tyr Gln Tyr
275                             280                             285
Tyr Leu Asp Leu Met Gly Val Gly Arg Glu His Ser Gly Asp Asn Arg
290                             295                             300
Asp Cys Ala Glu Gln Ala Pro Phe Asn Pro Ser Ser Pro Thr Asp Asp
305                             310                             315                             320
Gln Lys Gln Gln Gln Ser Gly Ser
325

<210>  17
<211>  316
<212>  PRT
<213>  Oryza sativa
<400>  17
Met Arg Thr Ser Arg Ala Ala Ala Ala Ala Ala Ser Leu Pro Leu Leu
  1    5      10      15
Leu Leu Val Ala Leu Leu Val Ala Ala Glu Gly Arg Arg His Lys Asp
  20     25                     30
Gly Ser Gly Asp Glu Glu Ala Lys Ala Cys Asp Lys Gly Trp Glu Cys
  35            40                45
Ser Gly Ser Arg Phe Cys Cys Asn Asp Thr Ile Thr Asp Tyr Phe Lys
  50            55                60
 Ala Tyr Gln Phe Glu Glu Leu Phe Ala His Arg Asn Asp Arg Ser Leu
  65       70         75        80
 Ala His Ala Ala Gly Phe Trp Asp Tyr His Ala Phe Ile Thr Ala Ala
  85          90                   95
 Ala Leu Phe Glu Pro Arg Gly Phe Gly Thr Thr Gly Lys Glu Val
100             105        110
Gly Met Lys Glu Val Ala Ala Phe Leu Gly His Val Gly Ala Lys Thr
115            120                 125
 Ser Cys Gly Tyr Ser Val Ala Thr Gly Gly Pro Leu Ala Trp Gly Leu
130           135             140
Cys Tyr Asn His Glu Leu Ser Pro Ser Gln Ser Tyr Cys Asp Asn Ser
145          150       155           160
Asn Glu Leu Tyr Pro Cys Val Glu Gly Val Glu Tyr Gly Arg Gly
165          170              175
 Ala Leu Pro Val Tyr Trp Asn Tyr Asn Tyr Gly Ile Ile Gly Gln Gly
180          185               190
 Ile Lys Gln Asp Leu Leu Asn His Pro Glu Leu Leu Glu Gln Asn Ala
195          200             205
 Thr Leu Ala Phe Glu Ala Ala Ile Trp Arg Trp Met Thr Pro Met Lys
210           215           220
Arg Lys Gln Pro Ser Ala His Asp Val Phe Val Gly Asn Trp Lys Pro
225          230      235        240
 Thr Lys Asn Asp Thr Leu Ser Lys Arg Tyr Pro Gly Phe Gly Ala Thr
245          250             255

Page 17
Met Asn Ile Leu Tyr Gly Asp Leu Ile Cys Gly Gln Gly Ser Ile Asp
260 265 270

Lys Met Asn Val Ile Val Ser His Tyr Gln His Tyr Leu Asp Leu Met
275 280 285

Gly Val Gly Ser Asp Lys Ala Gly Asp Asn Leu Asp Cys Ala Asp Gln
290 295 300

Val Ala Phe Asn Pro Ser Ser Lys Asn Leu Asp Ser
305 310 315

<210>  18
<211>  20
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab
<400>  18
    ttcagaacct ggagagcag  20

<210>  19
<211>  20
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab
<400>  19
    accagaacct ggagagcag  20

<210>  20
<211>  340
<212> DNA
<213> Arabidopsis thaliana

<400>  20
    ttcagaacct ggagagcag ggatgagcag cccaggaaga acatcttacc  60
gtcgccgac gatgttttct ggtgaaatg gaagccaaac aagaatgaca ctgctgcaca 120  
acgaacctcc gggtttgag ccaccataaa cgtttctttac ggcgatcataa tctgcaacag 180  
cggattgat aacgatgaga tgaacaacat tggttctctat tattttgttatt atctttgatct 240  
'gataggttt ggacgagaa aagctgggtcc tcacgagaag ctgctatgcag cagatcaaga 300  
acctttttct tcttcctctct ctgctcctcc gagtcttgtt 340
INTERNATIONAL SEARCH REPORT

A.
CLASSIFICATION OF SUBJECT MATTER

Int. Cl.: A01H 4/00 A01H 5/00 C12N 15/05 C12N 15/29

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)

SEE ELECTRONIC DATABASE FIELD

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

SEE ELECTRONIC DATABASE FIELD

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

Search GENBANK, EMBL DDBJ EST and DDBJ no EST, A. thaliana Database, O. sativa Database and D -GENE with SEQ ID NO 2, 3 and 4

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>
<tbody>
<tr>
<td>X</td>
<td>US 20030106097 A1 (HAIGLER et al.) 5 Jun 2003 (see in particular Claim 8 SEQ ID NO 1 is 88% identical over 195 nucleotides with SEQ ID NO 2 - nct 59 to 253: nct 143 to 337)</td>
<td>8 to 10, 13 to 18 23 and 24</td>
</tr>
<tr>
<td>X</td>
<td>EP 1033405 A2 (CERES INC) 6 Sep 2000 (see in particular Claim 1, SEQ ID 49338* and SEQ ID NO 14874* are 99% identical to SEQ ID NO 3 over 1138 nct (nct 41 to 1181: nct 25 to 1165 and nct 43 to 1183 and 25 to 1165, respectively) and SEQ ID NO 19936* is 97% identical over 444 nct to SEQ ID NO 3 (nct 43 to 486: nct 25 to 467 - * as cited in DGENE)</td>
<td>8, 9, 11, 13 to 18, 23 and 24</td>
</tr>
<tr>
<td>X</td>
<td>US 20020040489 A1 (GORLACH et al.) 4 Apr 2002 (see in particular Claim 1 SEQ ID NO 284* is 99% identical to SEQ ID NO 3 over 561 nct - nct 561 to 1: nct 642 to 1202 - * as cited in DGENE)</td>
<td>8, 9, 11, 13 to 18, 23 and 24</td>
</tr>
</tbody>
</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
  "&" document member of the same patent family

Date of the actual completion of the international search: 26 November 2004

Date of mailing of the international search report: 9 Dec 2004

Name and mailing address of the ISA/AU
AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaaustralia.gov.au
Facsimile No. (02) 6285 3929

Authorized officer
ALISTAIR BESTOW
Telephone No.: (02) 6283 2450
<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>
<tbody>
<tr>
<td>X</td>
<td>WO 2000 056908 A2 (PIONEER HI-BRED INT) 28 Sep 2000 (see in particular Claim 1; Page 74-76 where SEQ ID NO 11 is 88% identical over 393 nucleotides with SEQ ID NO 4 – nct 384 to 776: nct 293 to 685)</td>
<td>8, 9, 12 to 18 23 and 24</td>
</tr>
<tr>
<td>P,X</td>
<td>US 20040049808 A1 (HAGLER et al.) 11 Mar 2004 (see in particular: Claim 8 SEQ ID NO 3 100% identical over 256 nucleotides with SEQ ID NO 2 (nct 1 to 256 : nct 82 to 337) and SEQ ID NO 1 is 88% identical over 195 nucleotides with SEQ ID NO 2 (nct 59 to 253 : nct 143 to 337)</td>
<td>8 to 10, 13 to 18 23 and 24</td>
</tr>
</tbody>
</table>
INTERNATIONAL SEARCH REPORT

Box No. II  Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. □ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 21 (wholly) and 1, 8, 22, 23 and 24 (all partially)
   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
   Claims 1, 8, 22, 23 and 24 were partially searched insofar as they encompassed SEQ ID NO 2, 3 and 4, sequences for which the description lent support. Claim 21 could not be searched as it did not embody the technical features of the invention..

3. □ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box No. III  Observations where unity of Invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. □ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. □ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. □ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. □ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
□ The additional search fees were accompanied by the applicant’s protest.
□ No protest accompanied the payment of additional search fees.
This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EP 1033405</td>
<td>CA 2300692</td>
</tr>
<tr>
<td>US 20020040489</td>
<td>no family member</td>
</tr>
<tr>
<td>WO 2000 056908</td>
<td>AU 37330/00</td>
</tr>
</tbody>
</table>

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX