(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 May 2015 (28.05.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/077447 A2

(51) International Patent Classification:
A01H 5/00 (2006.01) *C12N 15/82* (2006.01)

(21) International Application Number:
PCT/US2014/066599

(22) International Filing Date:
20 November 2014 (20.11.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/907,617 22 November 2013 (22.11.2013) US

(71) Applicant: **BANGLADESH JUTE RESEARCH INSTITUE** [BD/BD]; Manik Mia Avenue, Dhaka, 1207 (BD).

(72) Inventor; and
(71) Applicant *(for US only)*: **ALAM, Maqsudul** [BD/US]; 3138 Wailae Avenue, Apt. 605, Honolulu, HI 96816 (US).

(72) Inventors: **ISLAM, Mohammed, Shahidul**; Bangladesh Jute Research Institute, Manik Mia Avenue, Dhaka, 1207 (BD). **AHMED, Borhan**; Bangladesh Jute Research Institute, Manik Mia Avenue, Dhaka, 1207 (BD). **HAQUE, Mohammed, Samiul**; Bangladesh Jute Research Institute, Manik Mia Avenue, Dhaka, 1207 (BD). **ALAM, Mohammed, Monjurul**; Bangladesh Jute Research Institute, Manik Mia Avenue, Dhaka, 1207 (BD).

(74) Agents: **GORDON, Dana, M.** et al.; Foley Hoag LLP, 155 Seaport Boulevard, Boston, MA 02210-2600 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

WO 2015/077447 A2

(54) **Title**: NUCLEOTIDE SEQUENCE ENCODING WUSCHEL-RELATED HOMEOBOX4 (WOX4) PROTEIN FROM CORCHORUS OLITORIUS AND CORCHORUS CAPSULARIS AND METHODS OF USE FOR SAME

(57) **Abstract**: The present invention discloses isolated polynucleotides encoding WUSCHEL-related homeobox4 proteins from two species of jute plants, namely, the *Corchorus olitorius ("C. olitorius")* and *Corchorus capsularis ("C. capsularis")*, and corresponding polypeptides derived therefrom. The disclosed polynucleotide sequences encode WUSCHEL-related homeobox4 polypeptides (WOX4), which possess catalytic activities in enhancing fiber production in jute. The present invention also relates to the plants having a modulated expression of a nucleic acid encoding a WOX4 polypeptide, which have enhanced fiber yield relative to corresponding wild type plants or other control plants. Vectors, expression constructs and host cells comprising and/or consisting of the nucleotide sequences of the protein are also provided. Also disclosed are methods for producing the proteins and methods for modifying the proteins in order to improve their desirable characteristics. The proteins of the invention can be used in a variety of ways, including inducing, initiating, improving, or enhancing plant growth, plant height, fiber and seed yield.

Nucleotide Sequence Encoding WUSCHEL-Related homeobox4 (WOX4) Protein from *Corchorus olitorius* and *Corchorus capsularis* and Methods of Use for Same

Related Application

This application claims the benefit of U.S. Provisional Application No. 61/907,617, filed on November 22, 2013; the entire contents of which is hereby incorporated by reference.

Field of invention

The present invention relates generally to the field of molecular biology and concerns a method for enhancing fiber yield-related traits by modulating expression in a plant of a nucleic acid encoding WUSCHEL-related homeobox4 protein (WOX4) polypeptide. More particularly, the present invention provides the WUSCHEL-related homeobox4 (WOX4) homologues isolated from two species of jute plants, namely, *C. olitorius* and *C. capsularis,* and its application in the production of bast fiber in the plant of jute, as well as a transgenic jute plant thereof.

Background of the invention

Jute is an eco-friendly and biodegradable natural fiber. It is a renewable resource with high biomass production per unit area of land. More than 100 jute species (including wild relatives) produce natural bast fiber. Jute is produced from plants in the genus *Corchorus*, which was once classified with the family Tiliaceae, more recently with Malvaceae. However, only two of these species, *C. olitorius* and *C. capsularis,* produce high quality fiber suitable for use in industrial purposes. Being a natural fiber, jute can be used in various ways, supplementing or replacing synthetics, and it has been receiving increasing attention from industry. As there is an increasing global demand for jute fiber, further improvement of jute fiber production is necessary. Thus, there is a significant interest in studying fiber biosynthesis and exploring the molecular biology involved in biosynthesis of this fiber.

Fiber of jute is an extraxylary fiber which is composed of and/or comprises two types of fiber: (i) primary phloem fiber that develops from procambium in the protophloem region through cell division and modification, and (ii) secondary phloem fiber that develops from cambium by the activity of fusiform and ray initials (Maiti and Mitra, 1972. Bull Bot Soc Bengal 26:79–85). These procambium and cambium tissues communicate cell-to-cell mediated by the tracheary element differentiation inhibitory factor (TIDF) and the TDIF

receptor (TDR) membrane protein kinase, which promotes proliferation of procambial cells and suppresses their xylem differentiation (Figure 1; Hirakawa et al., 2010. Plant Cell, 22:2618-2629; Etchells et al., 2013. Development 140, 2224-2234).

The WUSCHEL-related HOMEOBOX (WOX) gene family performs related functions during initiation and/or maintenance of various embryonic, meristematic, and organ initial cells (Haecker et al., 2004). Among the WUSCHEL-related HOMEOBOX (WOX) gene family proteins, WOX4 acts as a key regulator of TDIF signaling pathway (Hirakawa et al. 2010) and expressed preferentially in the procambium and cambium (Schrader et al., 2004; Ji et al., 2010 and Hirakawa et al. 2010). For example, TDIF-TDR induces the transcription of master transcription factor WUSCHEL-related HOMEOBOX4 (WOX4) that promotes the maintenance of procambium/cambium stem cells in *Arabidopsis* and in Tomato.

WUSCHEL-related HOMEOBOX4 (WOX4) polypeptide catalyzes the initiation of bast fiber in plant. However, there are not many characterization reports or existing technologies provided in the prior art relating to this polypeptide. U.S. Patent No. 2011/0283420 A1 (incorporated by reference) has disclosed wuschel related homeobox 1-like (WOX1-like) polypeptide for enhanced yield-related traits in plants. In another E.P. Patent No. 1451301 B1 disclosed the use of wuschel gene in promotion of somatic embryogenesis in plants. Recently, some wuschel gene homologs were disclosed in U.S. Patent No. 2010/0100981 A1 (incorporated by reference).

In view of the fact that WUSCHEL-related homeobox4 protein could play an important role in the biosynthesis pathway of jute fiber, it is desirable for industry to provide a genetic approach relating to the biosynthesis of the fiber in the plant by exploring and utilizing the molecular biology and genetic information of WUSCHEL-related homeobox4 (WOX4). Besides, because the fiber biosynthesis pathway and genetic make-up of each species of plant typically varies, a species-specific approach is also preferable in order to optimize yield of fiber from jute plants, and obtain compatible results to enable use in industry.

Summary of the invention

One aspect of the present invention, amongst others, is to provide a gene encoding protein, derived from *C. olitorius* and *C. capsularis,* which is involved in catalyzing their initiation of bast fiber formation, and has a WUSCHEL-related homeobox4 (WOX4)-like sequence. More specifically, the present invention provides a gene that has the ability to induce phloem fiber, thus providing a protein encoded thereby, and uses thereof.

Another object of the present invention is to provide the molecular biology and genetic information of WUSCHEL-related homeobox4 (WOX4) to be exploited/utilized for improving the production, growth, strength and yield of fiber in the plants of *C. olitorius* and *C. capsularis*, as well as in other bast fiber-producing plants.

Still another object of the present invention is to obtain a transgenic plant of *C. olitorius* and *C. capsularis* with increased fiber production by regulating the biosynthesis of WUSCHEL-related homeobox4 (WOX4) in the plant.

Yet another object of the present invention is to provide isolated polynucleotides having specific nucleotide sequences, which may facilitate the performance of the disclosed method, and provide access to transgenic *C. olitorius* and *C. capsularis* plants.

Further object of the present invention is to provide a potential commercially feasible way to increase the production of fiber for jute fiber-based products.

The present invention provides a gene isolated from *C. olitorius* encoding a protein that has the amino acid sequence as set forth in SEQ ID NO: 3, which is involved in catalyzing the initiation of phloem fiber formation and has a WUSCHEL-related homeobox4 (WOX4)-like sequence. The present invention further provides a gene encoding a protein that has an amino acid sequence modified by the addition or deletion of one or a plurality of amino acids and/or replacement with other amino acids in the amino acid sequence as set forth in SEQ ID NO: 3, which is involved in catalyzing the initiation of formation of phloem fiber and has a WUSCHEL-related homeobox4 (WOX4)-like sequence. The present invention further provides a gene that hybridizes to the nucleic acid as set forth in SEQ ID NO: 1, specifically its DNA or a portion thereof, and encodes a protein involved in catalyzing the initiation of formation of phloem fiber and has a WUSCHEL-related homeobox4 (WOX4)-like sequence.

The present invention also provides a gene isolated from *C. capsularis* encoding a protein that has the amino acid sequence as set forth in SEQ ID NO: 6, which is involved in catalyzing the initiation of formation of phloem fiber and has a WUSCHEL-related homeobox4 (WOX4)-like sequence. The present invention further provides a gene encoding a protein that has an amino acid sequence modified by the addition or deletion of one or a plurality of amino acids and/or replacement with other amino acids in the amino acid sequence as set forth in SEQ ID NO: 6, which is involved in catalyzing the initiation of formation of phloem fiber and has a WUSCHEL-related homeobox4 (WOX4)-like sequence. The present invention further provides a gene that hybridizes the nucleic acid as

set forth in SEQ ID NO: 4, specifically its DNA or a portion thereof, and encodes a protein involved in catalyzing the initiation of formation of phloem fiber and has a WUSCHEL-related homeobox4 (WOX4)-like sequence.

According to one of the preferred embodiments of the present invention, the plant of *C. olitorius* used is variety O-4, and the plant of *C. capsularis* used is variety CVL-1.

Another embodiment of the present invention discloses a recombinant gene construct comprising a polypeptide having nucleotide sequence set forth in SEQ ID NO: 2 and/or SEQ ID NO: 5, wherein the polynucleotide is expressible in a host cell to produce a homologue of WUSCHEL-related homeobox4 (WOX4) in the plant of *C. olitorius* and *C. capsularis*, respectively.

Further embodiment of the present invention is a transformant comprising a recombinant gene construct capable of expressing a polynucleotide having nucleotide sequence set forth in SEQ ID NO: 2 and/or SEQ ID NO: 5 to produce a homologue of WUSCHEL-related homeobox4 (WOX4) protein.

The present invention also provides a method for producing a protein involved in the catalytic activity for phloem fiber initiation, and has a WUSCHEL-related homeobox4 (WOX4)-like sequence, comprising culturing and/or cultivating the above host.

The present invention also provides a method for inducing initiation of phloem fiber of plants or plant cells; furthermore, it also discloses a method comprising and/or consisting of introducing the above gene into plants or plant cells, and driving the expression of said genes.

One skilled in the art will readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The embodiments described herein are not intended as limitations on the scope of the invention.

These and other features, aspects, and advantages of the present invention will be better understood with reference to the following description and claims.

Brief description of the drawings

Figure 1 displays TDIF-TDR signaling pathway. TDIF-TDR signaling diverges into two pathways, one is WOX4 involves in phloem cell proliferation, which promotes fiber cell initiation and another is a hypothetical factor X that inhibits xylem differentiation of vascular stem cells.

Figure 2 displays the phylogenetic tree comparing SEQ ID NO. 3 from *C. olitorius* and SEQ ID NO. 6 from *C. capsularis* along with other amino acid sequences, which produce WUSCHEL-related homeobox4 (WOX4) protein.

Detailed description of the invention

The invention can be more fully understood from the following detailed description and the accompanying drawings which form a part of this application.

The definitions and/or methods provided herein define the present invention and guide those of ordinary skill in the art in the practice of the present invention. Except where otherwise stated, terms are to be understood according to conventional usage by those of ordinary skill in the relevant art. To the extent to which any of the definitions and/or methods is found to be inconsistent with any of the definitions and/or methods provided in any patent or non-patent reference incorporated herein or in any reference found elsewhere, it is understood that the said definition and/or method which has been expressly provided/adopted in this application will be used herein. The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. Hence, "comprising A or B" means including A, or B, or A and B. It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below.

The present invention relates to isolated polynucleotides encoding WUSCHEL-related homeobox4 protein extracted from *C. olitorius* and *C. capsularis*, and corresponding polypeptides derived therefrom. More particularly, the present invention provides WUSCHEL-related homeobox4 homologues, and their application in enhanced fiber yield in *C. olitorius* and *C. capsularis*, both being species of jute plants, as well as related transgenic *C. olitorius* and *C. capsularis* plants. The genomic sequences of the invention encode the enzymes were identified primarily by comparison of nucleotide sequences of *C. olitorius* and *C. capsularis* genomic DNA and the nucleotide sequences of known enzyme genes of other plants. Prior to this invention, the nucleotide sequences of these *C. olitorius* and *C. capsularis* genes, the reading frames, the positions of exons and introns, the

structure of the enzymes, and their potential usefulness in the development of high fiber yield potential jute plants were not known.

Analysis of the genome sequence of commercially cultivated jute species, *C. olitorius* and *C. capsularis,* reveals that both species have single genes coding for enzymes which possess preferred catalytic activities for enhancing fiber production. The nucleotide sequences were initially annotated by software programs, such as Augustus, Semi-HMM-based Nucleic Acid Parser (SNAP), Geneid (Genome BioInformatics Research Lab), which can identify putative coding regions, introns, and splice junctions. Further automated and manual curation of the nucleotide sequences was performed to refine and establish precise characterization of the coding regions and other gene features.

Over 30,096 cDNAs from *C. olitorius* and 37,031 cDNAs from *C. capsularis* were partially or fully sequenced. From them a single cDNA was developed from each of the species, encoding new enzymes, with putative roles, preferred catalytic activities in enhancing fiber production in jute.

Open reading frames (ORFs) are analyzed following full or partial sequencing of clones of cDNA libraries derived from *C. olitorius* and *C. capsularis* mRNA and are further analyzed using sequence analysis software, and by determining homology to known sequences in databases (public/private).

In the context of this disclosure, a number of terms used throughout the specification have the indicated meanings unless expressly indicated to have a different meaning.

As used herein, a “polynucleotide” is a nucleotide sequence such as a nucleic acid fragment. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may comprise and/or consist of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures/combination thereof. An isolated polynucleotide of the present invention may include at least one of 150 contiguous nucleotides (both upstream and downstream) derived from SEQ ID No. 1 and, SEQ ID No. 4, or the complement of such sequences.

"Polypeptide" as used herein, is a single linear chain of amino acids bonded together by peptide bonds, and having usually a sequence greater than 100 amino acids in length.

"Isolated" means altered "by the hand of man" from the natural state. If a composition or substance occurs in nature, it has been "isolated" if it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally

present in a living plant or animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

The term “gene”, as used herein, is defined as the genomic sequences of the plant *C. olitorius* and *C. capsularis*, particularly polynucleotide sequence encoding polypeptide sequence of the WUSCHEL-related homeobox4 enzymes involved in the preferred catalytic activities in enhancing fiber production in jute. The term can further include nucleic acid molecules comprising upstream, downstream, and/or intron nucleotide sequences.

A "coding sequence" or "coding region" refers to a nucleic acid molecule having sequence information necessary to produce a gene product, such as an amino acid or polypeptide, when the sequence is expressed. The coding sequence may comprise and/or consist of untranslated sequences (including introns or 5' or 3' untranslated regions) within translated regions, or may lack such intervening untranslated sequences (e.g., as in cDNA).

The term “oligonucleotide” as used herein, is a short polynucleotide or a portion of polynucleotide, which may preferably comprise 10-1000, most preferably 12 to 50 nucleotides in length. In respect to the embodiment of the present invention, nucleotides contained within the oligonucleotides can be analogs or derivatives of naturally occurring nucleotides.

The term “primer” as used herein is an oligonucleotide capable of binding to a target nucleic acid sequence and priming the nucleic acid synthesis. An amplification oligonucleotide as defined herein may preferably be 10 to 50, most preferably 15 to 25, nucleotides in length. Furthermore, the amplification oligonucleotides of the present invention may be chemically synthesized and such oligonucleotides are not naturally occurring nucleic acids.

The abbreviation used throughout the specification to refer to nucleic acids comprising nucleotide sequences are the conventional one-letter abbreviations. Thus when included in a nucleic acid, the naturally occurring encoding nucleotides are abbreviated as follows: adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U). Also, unless otherwise specified, the nucleic acid sequences presented herein is the 5’→3’direction.

As used herein, the term “complementary” and derivatives thereof are used in reference to pairing of nucleic acids by the well-known rules that A pairs with T or U and C pairs with G. Complement can be “partial” or “complete”. In partial complement, only some of the nucleic acid bases are matched according to the base pairing rules; while in complete or

total complement, all the bases are matched according to the pairing rule. The degree of complement between the nucleic acid strands may have significant effects on the efficiency and strength of hybridization between nucleic acid strands as well known in the art. The efficiency and strength of the said hybridization depends upon detection method.

The term "host cell", as used herein, includes any cell type which is susceptible to transformation, transfection, transduction, expression and the like with a nucleic acid construct or expression vector comprising and/or consisting of a polynucleotide of the present invention. Suitable host cell includes fungi and/or plant cells, especially bast fiber producing plant cells.

The term "operably linked" generally denotes herein a configuration in which a control sequence is placed at an appropriate position relative to the coding sequence of the polynucleotide sequence such that the control sequence directs the expression of the coding sequence of a polypeptide. For example, a promoter can be operably-linked with a coding sequence when it affects the expression of that coding sequence, i.e., the coding sequence is under the transcriptional control of the promoter.

A “vector” generally refers to a replicon, such as plasmid, phage, cosmid, yeast or virus to which another nucleic acid segment may be operably inserted so as to bring about the replication or expression of the segment. The term "vector" is also intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid," which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Another type of vector is a viral vector, where additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression

vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The term "nucleic acid construct" or "DNA construct" is sometimes used to refer to a coding sequence or sequences operably linked to appropriate regulatory sequences and inserted into a vector for transforming a cell. This term may be used interchangeably with the term "transforming DNA" or "transgene."

The term "promoter", as used herein, refers to a nucleic acid sequence that functions to direct transcription of a downstream gene. The promoter will generally be appropriate to the host cell in which the target gene is being expressed. The promoter together with other transcriptional and translational regulatory nucleic acid sequences (also termed “control sequences”) is necessary to express a given gene. In general, the transcriptional and translational regulatory sequences include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences.

As used herein, “WUSCHEL polypeptide” or “WUS polypeptide” means a polypeptide having wuschel activity, i.e., involved in the initiation and maintenance of stem cells in plants. Wuschel activity stimulates cell growth, including stem cells. Wuschel is a plant homeodomain protein, comprising an ‘atypical’ (compared to the animal homeodomain motif) helix-loop-helix-turn-helix homeodomain motif com prising extra amino acid residues in the loop and/or turn of the domain. Wuschel proteins may further comprise and/or consist of other conserved motifs, such as the two conserved wuschel C-terminal domains, the (E/R)TLPLFP and A(A/S)LEL(S/T)L domains. The term is also inclusive of fragments, variants, and homologues, with any one of the above stated functions.

The term "homologues", as used herein, refers to a protein encompass peptides, oligopeptides, polypeptides, proteins and enzymes having amino acid substitutions, deletions and/or insertions relative to the unmodified protein in question and having similar biological and functional activity as the unmodified protein from which they are derived.

A deletion refers to removal of one or more amino acids from a protein.

An insertion refers to one or more amino acid residues being introduced into a predetermined site in a protein. Insertions may comprise N-terminal and/or C-terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Generally, insertions within the amino acid sequence will be smaller than N- or C-terminal fusions, of the order of about 1 to 10 residues.

A substitution refers to replacement of amino acids of the protein with other amino acids having similar properties (such as similar hydrophobicity, hydrophilicity, antigenicity, propensity to form or break α-helical structures or β-sheet structures). Amino acid substitutions are typically of single residues, but may be clustered depending upon functional constraints placed upon the polypeptide and may range from 1 to 10 amino acids; insertions will usually be of the order of about 1 to 10 amino acid residues. The amino acid substitutions are preferably conservative amino acid substitutions. Conservative substitution tables are well known in the art (see for example Creighton (1984) Proteins. W.H. Freeman and Company (Eds) and Table 1 below).

Table 1: Examples of conserved amino acid substitutions

Residue	Conservative Substitutions	Residue	Conservative Substitutions
Ala	Ser	Leu	Ile; Val
Arg	Lys	Lys	Arg; Gln
Asn	Gln; His	Met	Leu; Ile
Asp	Glu	Phe	Met; Leu; Tyr
Gln	Asn	Ser	Thr; Gly
Cys	Ser	Thr	Ser; Val
Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp; Phe
His	Asn; Gln	Val	Ile; Leu
Ile	Leu, Val		

Amino acid substitutions, deletions and/or insertions may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis and/or any other synthetic techniques, or by recombinant DNA manipulation. Methods for the manipulation of DNA sequences to produce substitution, insertion or deletion variants of a protein are well known in the art. For example, techniques for making substitution mutations at predetermined sites in DNA are well known to those skilled in the art and include M13 mutagenesis, T7-Gen in vitro mutagenesis (USB, Cleveland, OH), Quick

Change Site Directed mutagenesis (Stratagene, San Diego, CA), PCR-mediated site-directed mutagenesis or other site-directed mutagenesis protocols.

As used herein, a "biologically active portion" may refer to a fragment of WUSCHEL-related homeobox4 protein having a biological activity for catalyzing the initiation, formation, enhancement, or variation in the composition of phloem fiber in the plant of *C. olitorius* or *C. capsularis*. Biologically active portions of a WUSCHEL-related homeobox4 protein include peptides or polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the WUSCHEL-related homeobox4 protein, e.g., the amino acid sequence as set forth in SEQ ID NO: 3 or SEQ ID NO: 6, which include fewer amino acids than the full length WUSCHEL-related homeobox4 proteins, and exhibit at least one activity of a WUSCHEL-related homeobox4 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the WUSCHEL-related homeobox4 protein. A biologically active portion of a WUSCHEL-related homeobox4 protein can be a polypeptide which is, for example, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, or 223 amino acids in length.

The WUSCHEL-related homeobox4 protein may have an amino acid sequence set forth in SEQ ID NO: 3 or SEQ ID NO: 6. In other embodiments, the WUSCHEL-related homeobox4 protein is substantially identical to SEQ ID NO: 3 or SEQ ID NO: 6, and retains the functional activity of the protein of SEQ ID NO: 3 or SEQ ID NO: 6, yet differs in amino acid sequence due to natural allelic variation or mutagenesis. In another embodiment, the WUSCHEL-related homeobox4 protein comprises an amino acid sequence at least about 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to SEQ ID NO: 3 or SEQ ID NO: 6.

The term "domain", as used herein, refers to a set of amino acids conserved at specific positions along an alignment of sequences of evolutionarily related proteins. While amino acids at other positions can vary between homologues, amino acids that are highly conserved at specific positions indicate amino acids which are likely to be essential in the structure, stability or function of a protein. Identified by their high degree of conservation in aligned sequences of a family of protein homologues, they can be used as identifiers to determine if any polypeptide in question belongs to a previously identified polypeptide family.

The term "motif", as used herein, refers to a short conserved region in the sequence of evolutionarily related proteins. Motifs are frequently highly conserved parts of domains, but

may also include only part of the domain, or be located outside of conserved domain (if all of the amino acids of the motif fall outside of a defined domain).

Specialist databases exist for the identification of domains, for example, SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864; Letunic et al. (2002) Nucleic Acids Res 30, 242-244), InterPro (Mulder et al., (2003) Nucl. Acids. Res. 31, 315-318), Prosite (Bucher and Bairoch (1994), A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation. (In) ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds., pp53-61, AAAI Press, Menlo Park; Hulo et al., Nucl. Acids. Res. 32:D134-D137, (2004)), or Pfam (Bateman et al., Nucleic Acids Research 30(1): 276-280 (2002)). A set of tools for in silico analysis of protein sequences is available on the ExPASy proteomics server (Swiss Institute of Bioinformatics (Gasteiger et al., ExPASy: The Proteomics Server for In-depth Protein Knowledge and Analysis, Nucleic Acids Res. 31:3784-3788(2003)). Domains or motifs may also be identified using routine techniques, such as by sequence alignment.

For the purposes of the invention, "transgenic", "transgene" or "recombinant" means with regard to, for example, a nucleic acid sequence, an expression cassette, gene construct or a vector comprising and/or consisting of the nucleic acid sequence or an organism transformed with the nucleic acid sequences, expression cassettes or vectors according to the invention, all those constructions brought about by recombinant methods in which either: (a) the nucleic acid sequences encoding proteins useful in the methods of the invention, or (b) genetic control sequence(s) which is operably linked with the nucleic acid sequence according to the invention, for example a promoter, or (c) a) and b) are not located in their natural genetic environment or have been modified by recombinant methods. The modification may take the form of, for example, a substitution, addition, deletion, inversion or insertion of one or more nucleotide residues. The natural genetic environment is understood as meaning the natural genomic or chromosomal locus in the original plant or the presence in a genomic library. In the case of a genomic library, the natural genetic environment of the nucleic acid sequence is preferably retained, at least in part. The environment flanks the nucleic acid sequence at least on one side and has a sequence length of about 50 bp, preferably of about 500 bp. A naturally occurring expression cassette - for example the naturally occurring combination of the natural promoter of the nucleic acid sequences with the corresponding nucleic acid sequence

encoding a polypeptide useful in the methods of the present invention, as defined above - becomes a transgenic expression cassette when this expression cassette is modified by non-natural, synthetic ("artificial") methods such as, for example, mutagenic treatment. Suitable methods are described, for example, in US 5,565,350 (incorporated by reference) or WO 00/15815 (incorporated by reference).

A transgenic plant for the purposes of the invention is thus understood as including those plants in which the nucleic acids used in the method of the invention are not at their natural locus in the genome of the said plant, and thus it is possible for the nucleic acids to be expressed homologously or heterologously. However, as mentioned, transgenic also means that, while the nucleic acids according to the invention or used in the inventive method are at their natural position in the genome of a plant, the sequence has been modified with regard to the natural sequence, and/or that the regulatory sequences of the natural sequences have been modified. Transgenic is preferably understood as meaning the expression of the nucleic acids according to the invention at an unnatural locus in the genome, where homologous or, preferably, heterologous expression of the nucleic acids takes place.

The term "introduction" or "transformation", as used herein, encompass the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for transfer. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a genetic construct of the present invention and a whole plant regenerated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem). The polynucleotide may be transiently or stably introduced into a host cell and may be maintained in non-integrated form, for example, as a plasmid. Alternatively, it may be integrated into the host genome. The resulting transformed plant cell may then be used to regenerate a transformed plant in a manner known to persons skilled in the art.

The transfer of foreign genes into the genome of a plant is called transformation.

Transformation of plant species is now a fairly routine technique. Advantageously, any of several transformation methods may be used to introduce the gene of interest into a suitable ancestor cell. The methods described for the transformation and regeneration of plants from

plant tissues or plant cells may be utilized for transient or for stable transformation. Transformation methods include the use of liposomes, electroporation and chemicals which increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Transgenic plants, including transgenic crop plants, are preferably produced via Agrobacterium-mediated transformation. An advantageous transformation method is the transformation in planta (Sajib et. al. Plant Cell Tiss. Organ Cult. (2008) 95, 333-34).

Generally after transformation, plant cells or cell groupings are selected for the presence of one or more markers which are encoded by plant-expressible genes co-transferred with the gene of interest, following which the transformed material is regenerated into a whole plant. To select transformed plants, the plant material obtained in the transformation is, as a rule, subjected to selective conditions so that transformed plants can be distinguished from untransformed plants. For example, the seeds obtained in the above-described manner can be planted and, after an initial growing period, subjected to a suitable selection by spraying. A further possibility consists in growing the seeds, if appropriate after sterilization, on agar plates using a suitable selection agent so that only the transformed seeds can grow into plants. Alternatively, the transformed plants are screened for the presence of a selectable marker such as the ones described above.

Following DNA transfer and regeneration, putatively transformed plants may also be evaluated, for instance using Southern analysis, for the presence of the gene of interest, copy number and/or genomic organization. Alternatively or additionally, expression levels of the newly introduced DNA may be monitored using Northern and/or Western analysis, both techniques being well known to persons having ordinary skill in the art.

The generated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1) transformed plant may be selected and homozygous second-generation (or T2) transformants may be selected, and the T2 plants may then further be propagated through classical breeding techniques. The generated transformed organisms may take a variety of forms. For example, they may be chimeras of transformed cells and non-transformed cells; clonal transformants (e.g., all cells transformed to contain the expression cassette); grafts of transformed and untransformed tissues (e.g., in plants, a transformed rootstock grafted to an untransformed scion).

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

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

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added may be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

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

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences can be aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes can be at

least 95% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions can then be compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In one embodiment, the percent identity between two amino acid sequences can be determined using the Needleman and Wunsch (J. Mol. Biol. 48:444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (Comput. Appl. Biosci. 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0 or 2.0U), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

Exemplary computer programs which can be used to determine identity between two sequences include, but are not limited to, the suite of BLAST programs, *e.g.*, BLASTN, BLASTX, and TBLASTX, BLASTP and TBLASTN, publicly accessible at www.ncbi.nlm.nih.gov/BLAST.

Sequence searches are typically carried out using the BLASTN program, when evaluating a given nucleic acid sequence relative to nucleic acid sequences in the GenBank DNA Sequences and other public databases. The BLASTX program is preferred for searching nucleic acid sequences that have been translated in all reading frames against amino acid sequences in the GenBank Protein Sequences and other public databases.

A preferred alignment of selected sequences in order to determine "% identity" between two or more sequences is performed using, for example, the CLUSTAL-W program.

As in setting forth, one embodiment of the present invention are isolated polynucleotides encoding WUSCHEL-related homeobox4 polypeptide found in the plants *C. olitorius and C. capsularis* comprising and/or consisting of nucleotide sequence as set forth in SEQ ID NO 2 and SEQ ID NO 5, respectively. Correspondingly, the respective WUSCHEL-related homeobox4 polypeptides encoded by the nucleotide sequences possess the amino acid sequences set forth in SEQ ID NO 3 and SEQ ID NO 6. According to an embodiment of the present invention, SEQ ID NO 3 refers to the polypeptide sequence of the *C. olitorius*-derived WUSCHEL-related homeobox4 (WOX4) homologue, and SEQ ID NO 6 refers to the polypeptide sequence of the *C. capsularis*-derived WUSCHEL-related homeobox4 (WOX4) homologue. Both these enzymes are present in the biosynthesis pathway of fiber in the plants of *C. olitorius* and *C. capsularis* for catalyzing the initiator molecules for priming the biosynthesis of fiber cell in the plant.

The present invention also provides a gene sequence encoding the WUSCHEL-related homeobox4 (WOX4) homologues from the plants *C. olitorius* and *C. capsularis*.

In one embodiment, the 1250 bp long polynucleotide illustrated in SEQ ID No. 1 is the full length gene isolated from *C. olitorius*. This gene sequence includes at least 150 contiguous nucleotides from both upstream and downstream of the gene. This also provides the intronic sequence of the gene.

In another embodiment, the 1237 bp long polynucleotide illustrated in SEQ ID No. 4 is the full length gene isolated from *C. capsularis*. This gene sequence includes at least 150 contiguous nucleotides from both upstream and downstream of the gene. This also provides the intronic sequence of the gene.

In still another embodiment of the present invention, an isolated polynucleotide encoding a polypeptide comprising nucleotide sequence set forth in SEQ ID NO 2 and/or SEQ ID NO 5 is provided. SEQ ID NO 2 refers to the polynucleotide sequence of the *C. olitorius*-derived WUSCHEL-related homeobox4 (WOX4) homologue sequence and SEQ ID NO 5 refers to the polynucleotide sequence of the *C. capsularis*-derived WUSCHEL-related homeobox4 (WOX4) homologue sequence.

In still another embodiment, an isolated nucleic acid molecule which is capable of encoding a WUSCHEL-related homeobox4 polypeptide, or biologically active fragment thereof, comprises a nucleotide sequence which is at least about 95%, 96%, 97%, 98%,

99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to the entire length of the nucleotide sequence set forth in SEQ ID NO: 2, SEQ ID NO: 5, or any complement thereof.

In one embodiment, the 669 bp long polynucleotide illustrated in SEQ ID No. 2 is the full length cDNA clone encoded WUSCHEL-related homeobox4 (WOX4) protein exhibiting an open reading frame encoding a 222 amino acid polypeptide, as in SEQ ID No. 3, with a calculated molecular mass of about 25.54 kD. Through SMART analysis of SEQ ID No. 3, it reveals presence of homeobox domain in the sequence. This is a DNA-binding factor, which is involved in the transcriptional regulation of key developmental processes of plant. This is WUSCHEL-related homeobox4 (WOX4) protein involved in the vascular cell proliferation of plant.

In one embodiment, the 672 bp long polynucleotide illustrated in SEQ ID No. 5 is the full length cDNA clone encoded WUSCHEL-related homeobox4 (WOX4) protein exhibiting an open reading frame encoding a 223 amino acid polypeptide, as in SEQ ID No. 6, with a calculated molecular mass of about 25.67 kD. Through SMART analysis of SEQ ID No. 6, it reveals presence of homeobox domain in the sequence. This is a DNA-binding factor, which is involved in the transcriptional regulation of key developmental processes of plant. This is WUSCHEL-related homeobox4 (WOX4) protein involved in the vascular cell proliferation of plant.

In accordance with the preferred embodiment of the present invention, the isolated polynucleotide illustrated in SEQ ID NO 2 can be obtained by PCR amplification of the conserved region of the gene using total RNA isolated from the plant of *C. olitorius* and SEQ ID NO 5 can be obtained by PCR amplification of the conserved region of this gene using total RNA isolated from the plant of *C. capsularis*. As set forth in the preceding description, the plant of *C. olitorius* applied is O-4 variety and *C. capsularis* applied is CVL-1 variety.

Another embodiment of the present invention, a recombinant gene construct comprising a polynucleotide having nucleotide sequence set forth in SEQ ID NO 2 and/or SEQ ID NO 4 is disclosed, wherein the polynucleotide is expressible in a host cell, and is translable to produce homologue of WUSCHEL-related homeobox4 (WOX4) protein in the plants of *C. olitorius and C. capsularis*. The procedure for amplifying, cloning and sequencing the WUSCHEL-related homeobox4 (WOX4) from the plants of *C. olitorius* and *C. capsularis* is further detailed in Example 2. Preferably, the recombinant gene construct further

comprises a promoter region operably-linked to enhance expression of the polynucleotide template. Under the transcriptional control of the specific promoter, the expression of the coding region within the recombinant gene constructs containing polynucleotide of SEQ ID NO 2 and/or SEQ ID NO 4 can then be enhanced, leading to higher yield of the WUSCHEL-related homeobox4 (WOX4) protein.

According to an embodiment of the invention, the modulated expression is increased expression or activity, e.g. over-expression of a WUSCHEL-related homeobox4 polypeptide encoding nucleic acid molecule, e.g., of a nucleic acid molecule encoding SEQ ID NO 2 and SEQ ID NO 5. Methods for increasing expression of nucleic acids or genes, or gene products, are well documented in the art and examples are provided in the definitions section.

The invention also provides a method for the production of transgenic plants having enhanced fiber yield relative to control plants, comprising introduction and expression in a plant of any nucleic acid encoding a WUSCHEL-related homeobox4 polypeptide as defined herein above.

More specifically, the present invention provides a method for the production of transgenic plants having enhanced fiber yield in comparison to the null control plants, which method comprises:

(i) introducing and expressing in a plant or plant cell a WUSCHEL-related homeobox4 polypeptide-encoding nucleic acid or a genetic construct comprising and/or consisting of a WUSCHEL-related homeobox4 polypeptide-encoding nucleic acid; and

(ii) cultivating the plant cell under conditions promoting fiber cell growth and development.

Another aspect of the invention relates to isolated polynucleotide which encodes a WUSCHEL-related homeobox4 polypeptide, and is derived from the plant *C. olitorius*, comprising a nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence set forth in of SEQ ID NO 2, or a complement thereof; and

b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence of set forth in SEQ ID NO: 2, or a complement thereof.

In certain embodiments, the plant of *C. olitorius* is variety O-4.

Another aspect of the invention relates to an isolated which encodes a WUSCHEL-related homeobox4 polypeptide and is derived from the plant C. capsularis, comprising a nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 5, or a complement thereof; and

b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 5, or a complement thereof.

In certain embodiments, the plant of *C. capsularis* is variety CVL-1.

Another aspect of the invention relates to an isolated WUSCHEL-related homeobox4 polypeptide comprising an amino acid sequence set forth in SEQ ID: NO: 3, or biologically-active fragment thereof, said polypeptide comprises a function selected from the group consisting of catalyzing the initiation, formation, enhancement, and variation, to thereby modify the composition of phloem fiber in the plant of *C. olitorius*.

In certain embodiments, the plant of *C. olitorius* is variety O-4.

In certain embodiments, said polypeptide comprises at least 95% sequence identity to the amino acid sequence set forth in SEQ ID NO: 3.

Another aspect of the invention relates to an isolated WUSCHEL-related homeobox4 polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 6, or biologically-active fragment thereof, said polypeptide comprises one or more functions selected from the group consisting of catalyzing the initiation, formation, enhancement, and variation, to thereby modify the composition of phloem fiber in the plant of *C. capsularis*.

In certain embodiments, the plant of *C. capsularis* is variety CVL-1.

In certain embodiments, said polypeptide comprises at least 95% sequence identity to the amino acid sequence set forth in SEQ ID NO: 6.

Another aspect of the invention relates to a recombinant gene construct comprising a polynucleotide selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof; and

b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof, wherein the polynucleotide is expressible in a host cell to produce

a homologue of WUSCHEL-related homeobox4 polypeptide in the plants of *C. olitorius* and *C. capsularis*.

In certain embodiments, said construct further comprises a promoter region operably-linked to

a) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof; or

b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof, wherein said promoter enhances the transcription or expression of the nucleic acid molecule.

Another aspect of the invention relates to a transformant comprising a recombinant gene construct capable of expressing a polynucleotide selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof; and

b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof, wherein said transformant produces a homologue of WUSCHEL-related homeobox4 polypeptide.

Another aspect of the invention relates to a method for producing a plant or transgenic plant having increased or enhanced fiber yield relative to control plants, comprising:

a) introducing into a plant cell a recombinant gent construct comprising a polynucleotide selected from the group consisting of: i) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof; and ii) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof,

b) cultivating the plant cell under conditions for promoting plant growth and development; and

c) expressing a polypeptide selected from the group consisting of : i) a polypeptide comprising an amino acid sequence set forth in SEQ ID: NO: 3 or SEQ ID NO: 6, or a biologically-active fragment thereof; and ii) a polypeptide having at least 95% sequence identity to the amino acid sequence set forth in SEQ ID NO: 3 or SEQ ID NO: 6.

The sequences provided by the present invention can also be used as preparatory materials for the rational modification or design of novel enzymes with characteristics that enable the enzymes to perform better in demanding processes.

The present disclosure includes as contained in the appended claims, as well as that of the foregoing description. Although this invention has been described in its preferred form with a degree of particularity, it is understood that the present disclosure of the preferred form has been made only by way of example and that numerous changes in the details of construction and the combination and arrangements of parts may be resorted to without departing from the scope of the invention and claims.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.

Examples

The following examples are intended to further illustrate the invention, without any intent for the invention to be limited to the specific embodiments described therein.

Example 1 Designing and synthesis of primers

The primers used in the study were either designed from the manually curated transcriptome and the "gene models" predicted from the genomic sequences of *C. olitorius* and *C. capsularis*, by choosing the sequences manually with complete ORFs or using databases where similar genes have been successfully isolated from other plants. Comparative bioinformatic analysis of the nucleotide sequences obtained from transcriptome were carried out using NCBI BLAST, BLASTP, RPS-BLAST, BLASTX and PSI-BLAST to identify homologues of the related genes and for the proper identification of gene. Nucleotide sequence alignments were performed through clustalW version 1.82 whenever multiple sequences were found from the "gene pool". The alignment was then edited. Gene specific primers (both forward and reverse) were selected manually or through Primer 3 plus tool and the primers were custom synthesized.

All oligonucleotides used in this study were synthesized and HPLC purified by the supplier and procured from Integrated DNA Technologies (IDT). Stock solution of about 100 pmol were prepared in autoclaved ddH_2O and stored at about -20°C, in aliquots for use.

Oligonucleotides Sequences used as primers for PCR

Primer name	SEQ ID NO	Oligonucleotide sequence	Amplified product from cDNA
COL F	1	CCATGGGAAACATGAAGGTGC	682
COL R	1	TGAAACGTCCATCATCTGCCT	
CCA F	4	CCATGGGAAACATGAAGGTGC	675
CCA R	4	TTCATGATCTGCCTTCCGGG	

Example 2 Amplification, cloning and sequencing of WUSCHEL-related homeobox4 from *C. olitorius* and *C. capsularis*

Total RNA was isolated from three days old seedlings grown on MS medium as previously described by Chomczynski P and Sacchi N, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. (Anal Biochem 1987, 162: 156-

159). The quality or the integrity of the RNA was checked by agarose gel electrophoresis and was quantified using Thermo Scientific Nano Drop 2000 as per standard procedures. cDNA first strand was synthesised using SuperScript III reverse transcriptase (Invitrogen) following the manufacturer's instructions. The gene was amplified from the cDNA by PCR using the gene specific primers. The PCR reaction (50 µL) contained 1 µL of cDNA, 20 pmoles of each primers, 5 µL of 10X PCR Buffer, 5 µL of 2.5 mM dNTP mix and 1.0 unit of PfuTaq DNA polymerase. PCR was carried out in Thermal Cycler (Applied Biosystems) using the following conditions: initial denaturation for about 5 min at approximately 95°C followed by 35 cycles of denaturation at approximately 95°C for about 30 sec, annealing at approximately 59-61°C for about 30 sec and extension at approximately 72°C for about 1 min, with a final extension at approximately 72°C for about 7 min. The PCR product was analyzed by 1% agarose gel using 1X TAE buffer and the amplicon was eluted from the gel using QIAGEN gel extraction kit following the manufacturer's instructions. The purified PCR product was ligated into pCR®8/GW/TOPO® TA cloning kit (Invitrogen) and transformed into competent *E. coli* cells (Invitrogen). Plasmids were isolated from putative colonies using QIAprip Spin Miniprep Kit (QIAGEN) following the manufacturer's instructions. The presence of the insert was checked by using the gene specific primers and positive plasmids were subjected to Sequencing.

Example 3 Analysis of the sequence

The nucleotide sequence and the amino acid sequence were analyzed by BLASTN and BLASTP programs respectively. The sequences reported from other plants were aligned with ClustalW. Phylogenetic analysis was carried out using the Neighbour Joining (NJ).

Example 4 Pathway construction of fiber biosynthesis

Automatic metabolic pathway reconstruction showcasing role of WUSCHEL-related homeobox4 (WOX4) in fiber biosynthesis was constructed by identifying orthologs from *C. olitorus* and *C. capsularis* protein compared with *Arbidopsis* genome. WUSCHEL-related homeobox4 (WOX4) catalyzed enzymatic reactions encoded within *C. olitorus* and *C. capsularis* genome were constructed using enzymatic reactions available in Resnet-Plant 3.0 database for Pathway Studio as well as from metabolic pathway databases.

Incorporation by reference

All of the U.S. patents, U.S. published patent applications, and published PCT applications that cited herein are hereby incorporated by reference.

Equivalents

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto; the invention may be practised otherwise than as specifically described and claimed.

SEQ ID NO : 1
LENGTH : 1250 bp
TYPE : DNA
FEATURE NAME/KEY : Intron, exon including 150 bp 5' UTR and 150 bp 3' UTR
ORGANISM : Jute, *C. olitorius*

ACCTCATCATTTGGTCTTGGTTCTTCAAAGCACCACCTCATATCCCACTTTCTTTTTCTTTTCTTTT
TTTTAAAATTTTTATTTTTTGGGTTAAATTTATTTTTGTTATTATATAATCAGAGCATTCTCCCCAT
CATTCACTCCTTCACCATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAA
CCCTCCCTCTCCCTTGGTTGCAAGCGCTTACGCCCTCTTGCTCCCAAGCTCCACCCTTCCTCTTCCC
CTGATCATACTTCCGTCTCCTCTTTCGACCTTAAGACCTTCATTCGTCCCGAAAGTGGCCCCCGAAA
ACTTTGCCCTTCTGACGACAAGCGAGATTCTCATTCTCCCCAGGTACTTACAAATATCCATGAAATG
AAACCCTTTTATTTTTCCAATGATGTATATAATTAGTGGGGAAGGATAAGATTTTGAACTTACGATG
AACAATAACAGTGTGGAAAGTTGATAATCAAAGGTTAAGAATGTGAAGCTGAAACTTTTTTCTTTTT
TTTGGTTATGTAGGTGGAAACGCACCCAGGGGGAACGCGGTGGAATCCGACGCAAGAGCAGATAGGG
ATATTGGAGATGCTGTATAGAGGTGGGATGCGAACTCCAAATGCACAGCAAATAGAACAGATCACTG
CACAGTTAGGCAAGTACGGGAAGATCGAAGGCAAAAACGTTTTCTATTGGTTCCAAAACCACAAAGC
ACGCGAAAGGCAAAAGCAGAAGCGTAACAGTCTTGGTCTTAGCCATTCTCCCAGAAACTCTGCTCCC
ATTACCACCATAACTTTGGACTCTAGGGTAAGTTCAAACCAACAAAACCCTTTCTTTGTATATATAT
AACGGTTAGTTTTTAGTTTTTACTTCTTATAAACAGCAATTAACATTAATGTTTTTGTTTATATATA
TAGGGGGAAGTAATGGAGAGAGAGGAGGATAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCATTTG
AGTACTTAGAAGAAGAAAGCAGATCATCATCGTCGAGTCAAGAGGAGGAAAACAGAACTCTGGAGCT
TTTCCCATTGCACCCGGAAGGCAGATGATGGACGTTTCAACTTTGAAAAACAAGGAAAAAGGGAAGC
TTAACCCAAAACCAAAAAGACTGCTACAAAACCCAAAACTCTGTTCCCATTTATGAAATGATAAACA
TATGCTTTGATGATCCATGATGATGATGATGATAATGAAGCTGA

SEQ ID NO : 2
LENGTH : 669 bp
TYPE : DNA
ORGANISM : Jute, *C. olitorius*
FEATURE NAME/KEY : CDS
LOCATION : (1) ….. (669)

ATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAACC
CTCCCTCTCCCTTGGTTGCAAGCGCTTACGCCCTCTTGCTCCCAAGCTCCACCCT
TCCTCTTCCCCTGATCATACTTCCGTCTCCTCTTTCGACCTTAAGACCTTCATTCG
TCCCGAAAGTGGCCCCCGAAAACTTTGCCCTTCTGACGACAAGCGAGATTCTCA
TTCTCCCCAGGTGGAAACGCACCCAGGGGGAACGCGGTGGAATCCGACGCAAG
AGCAGATAGGGATATTGGAGATGCTGTATAGAGGTGGGATGCGAACTCCAAAT
GCACAGCAAATAGAACAGATCACTGCACAGTTAGGCAAGTACGGGAAGATCGA
AGGCAAAAACGTTTTCTATTGGTTCCAAAACCACAAAGCACGCGAAAGGCAAA
AGCAGAAGCGTAACAGTCTTGGTCTTAGCCATTCTCCCAGAAACTCTGCTCCCA
TTACCACCATAACTTTGGACTCTAGGGGGGAAGTAATGGAGAGAGAGGAGGAT
AGTCCATATAAGAGAAAGTGTAGGAGCTGGTCATTTGAGTACTTAGAAGAAGA
AAGCAGATCATCATCGTCGAGTCAAGAGGAGGAAAACAGAACTCTGGAGCTTT
TCCCATTGCACCCGGAAGGCAGATGA

SEQ ID NO : 3
LENGTH : 222
TYPE : PRT
ORGANISM : Jute, *C. olitorius*

MGNMKVHQLARGLWEHEPSLSLGCKRLRPLAPKLHPSSSPDHTSVSSFDLKTFIRPESGPRKLCPSD
DKRDSHSPQVETHPGGTRWNPTQEQIGILEMLYRGGMRTPNAQQIEQITAQLGKYGKIEGKNVFYWF
QNHKARERQKQKRNSLGLSHSPRNSAPITTITLDSRGEVMEREEDSPYKRKCRSWSFEYLEEESRSS
SSSQEEENRTLELFPLHPEGR*

SEQ ID NO : 4
LENGTH : 1237 bp
TYPE : DNA
FEATURE NAME/KEY : Intron, exon including 150 bp 5' UTR and 150 bp 3' UTR
ORGANISM : Jute, *C. capsularis*

TCACAAGTCAACCTCACCTCATCATTTGGTCTTGGTTCTTCAAAGCACCACCTCATATCCCACTTCC
TTTTTCAATTTTTAATTTTTTTTGGGGTTAAATTTATTTGGTTATATAATCAGAGCATTCTCCCCAT
CATTCACTCCTTCACCATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAA
CCCTCCCTCTCCCTTGGTTGCAAGCGCTTACGCCCTCTTGCTCCCAAGCTCCACCCTTCCTCTTCCC
CTGATCATACTTCCGTCTCCTCTTTCGACCTTAAGACCTTCATTCGTCCCGAAAGTGGCCCCCGGAA
ACTTTGCCCTTCTGACGACAAGCGAGATTCTCATTCTCGCCAGGTACTTAAAAATTAATATCCATGA
AATAATTTGTGGGGAAGGATAAGTTTTGAACTTAATGCATAATAACAGTGTGGAAACTTAATAGGTT
AAGAATATTTGAAGAACTTCTATATATATGAAGCTGAAACTTTTTTGTGGTGTTGTAGGTGGAAACG
CACCCAGGGGGAACGCGGTGGAATCCGACGCAAGAGCAGATAGGGATACTGGAGATGCTGTATAGAG
GTGGGATGCGAACTCCAAATGCACAGCAAATAGAACAGATCACTGCACAGCTAGGCAAGTACGGCAA
GATCGAAGGCAAAAACGTTTTCTATTGGTTCCAAAACCACAAAGCACGCGAAAGGCAAAAGCAGAAG
CGTAACAGTCTTGGTCTTAGCCATTCTCCCAGAAACTCAGCTCCCATTACCACTATAACTTTGGACA
CTAGGGTAAGTTCAAACCAACAAAACCCTTTCTGTGTATATATATATATAACGGTTAGTTTTTAGTT
TTTACTTCTTATAAACAGAGAAAATTAACAATGTTTGTGTTTATATATATATAGGGGGAAGTAATGG
AAAGAGAGGAGGATAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCTTTTGAGTACTTAGAAGAAGA
AAGCAGATCATCATCGTCGAGTCAAGAGGAGGAAAACAGAACTCTGGAGCTTTTCCCATTGCACCCG
GAAGGCAGATCATGAAGGGGGTTTCAACTTTCAACTTTCAACTTTCAACTTTCAAAATGAAGGGAAA
AGGGAAGCTTAACCCAAAACCAAAAAGACTGCTACAAAACCCAAAACTCTGTTCCCATTTATGAAAT
GATAAACTTATGCTTTGATGATCGATCCATG

SEQ ID NO : 5
LENGTH : 672
TYPE : DNA
ORGANISM : Jute, *C. capsularis*
FEATURE NAME/KEY : CDS
LOCATION : (1) ….. (672)

ATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAACCCTCCCTCTCCCTTG
GTTGCAAGCGCTTACGCCCTCTTGCTCCCAAGCTCCACCCTTCCTCTTCCCCTGATCATACTTCCGT
CTCCTCTTTCGACCTTAAGACCTTCATTCGTCCCGAAAGTGGCCCCCGGAAACTTTGCCCTTCTGAC

GACAAGCGAGATTCTCATTCTCGCCAGGTGGAAACGCACCCAGGGGGAACGCGGTGGAATCCGACGC
AAGAGCAGATAGGGATACTGGAGATGCTGTATAGAGGTGGGATGCGAACTCCAAATGCACAGCAAAT
AGAACAGATCACTGCACAGCTAGGCAAGTACGGCAAGATCGAAGGCAAAAACGTTTTCTATTGGTTC
CAAAACCACAAAGCACGCGAAAGGCAAAAGCAGAAGCGTAACAGTCTTGGTCTTAGCCATTCTCCCA
GAAACTCAGCTCCCATTACCACTATAACTTTGGACACTAGGGGGGAAGTAATGGAAAGAGAGGAGGA
TAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCTTTTGAGTACTTAGAAGAAGAAAGCAGATCATCA
TCGTCGAGTCAAGAGGAGGAAAACAGAACTCTGGAGCTTTTCCCATTGCACCCGGAAGGCAGATCAT
GA

SEQ ID NO : 6

LENGTH : 223

TYPE : PRT

ORGANISM : Jute, *C. capsularis*

MGNMKVHQLARGLWEHEPSLSLGCKRLRPLAPKLHPSSSPDHTSVSSFDLKTFIRPESGPRKLCPSD
DKRDSHSRQVETHPGGTRWNPTQEQIGILEMLYRGGMRTPNAQQIEQITAQLGKYGKIEGKNVFYWF
QNHKARERQKQKRNSLGLSHSPRNSAPITTITLDTRGEVMEREEDSPYKRKCRSWSFEYLEEESRSS
SSSQEEENRTLELFPLHPEGRS*

What is claimed is:

1. An isolated polynucleotide which encodes a WUSCHEL-related homeobox4 polypeptide and is derived from the plant *C. olitorius*, comprising a nucleic acid molecule selected from the group consisting of:
   a) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO 2, or a complement thereof; and
   b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2, or a complement thereof.
2. The isolated polynucleotide encoding according to claim 1, wherein the plant of *C. olitorius* is variety O-4.
3. An isolated polynucleotide which encodes a WUSCHEL-related homeobox4 polypeptide and is derived from the plant *C. capsularis*, comprising a nucleic acid molecule selected from the group consisting of:
   a) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 5, or a complement thereof; and
   b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 5, or a complement thereof.
4. The isolated polynucleotide according to claim 3, wherein the plant of *C. capsularis* is variety CVL-1.
5. An isolated WUSCHEL-related homeobox4 polypeptide comprising an amino acid sequence set forth in SEQ ID: NO: 3, or biologically-active fragment thereof, said polypeptide comprises a function selected from the group consisting of catalyzing the initiation, formation, enhancement, and variation, to thereby modify the composition of phloem fiber in the plant of *C. olitorius*.
6. The isolated polypeptide according to claim 5, wherein the plant of *C. olitorius* is variety O-4.
7. The isolated polypeptide according to claim 5, wherein said polypeptide comprises at least 95% sequence identity to the amino acid sequence set forth in SEQ ID NO: 3.
8. An isolated WUSCHEL-related homeobox4 polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 6, or biologically-active fragment thereof, said polypeptide comprises one or more functions selected from the group consisting of

catalyzing the initiation, formation, enhancement, and variation, to thereby modify the composition of phloem fiber in the plant of *C. capsularis*.

9. The isolated polypeptide according to claim 8, wherein the plant of *C. capsularis* is variety CVL-1.

10. The isolated polypeptide according to claim 8, wherein said polypeptide comprises at least 95% sequence identity to the amino acid sequence set forth in SEQ ID NO: 6.

11. A recombinant gene construct comprising a polynucleotide selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof; and

b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof,

wherein the polynucleotide is expressible in a host cell to produce a homologue of WUSCHEL-related homeobox4 polypeptide in the plants of *C. olitorius* and *C. capsularis*.

12. The recombinant gene construct according to claim 11, further comprising a promoter region operably-linked to a nucleic acid molecule set forth in a) or b), wherein said promoter enhances the transcription or expression of the nucleic acid molecule.

13. A transformant comprising a recombinant gene construct capable of expressing a polynucleotide selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof; and

b) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof,

wherein said transformant produces a homologue of WUSCHEL-related homeobox4 polypeptide.

14. A method for producing a plant or transgenic plant having increased or enhanced fiber yield relative to control plants, comprising:

a) introducing into a plant cell a recombinant gent construct comprising a polynucleotide selected from the group consisting of:

i) a nucleic acid molecule comprising a nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof; and

ii) a nucleic acid molecule comprising a nucleotide sequence having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5, or a complement thereof,

b) cultivating the plant cell under conditions for promoting plant growth and development; and

c) expressing a polypeptide selected from the group consisting of :

i) a polypeptide comprising an amino acid sequence set forth in SEQ ID: NO: 3 or SEQ ID NO: 6, or a biologically-active fragment thereof; and

ii) a polypeptide having at least 95% sequence identity to the amino acid sequence set forth in SEQ ID NO: 3 or SEQ ID NO: 6.

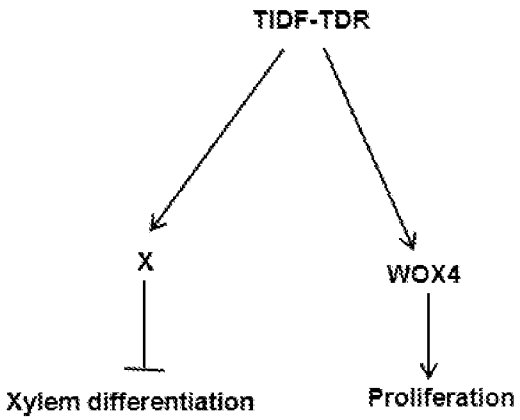

Figure 1

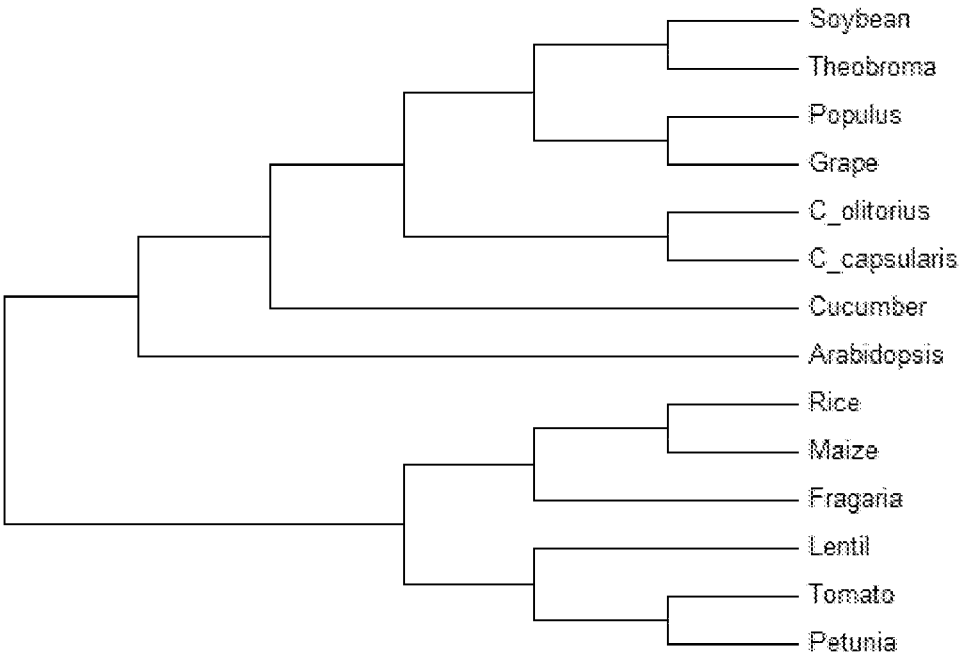

Figure 2

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 May 2015 (28.05.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/077447 A3

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/10* (2006.01)
C07K 14/415 (2006.01)

(21) International Application Number:
PCT/US2014/066599

(22) International Filing Date:
20 November 2014 (20.11.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/907,617 22 November 2013 (22.11.2013) US

(71) Applicant: **BANGLADESH JUTE RESEARCH INSTITUE** [BD/BD]; Manik Mia Avenue, Dhaka, 1207 (BD).

(72) Inventor; and
(71) Applicant *(for US only)*: **ALAM, Maqsudul** [BD/US]; 3138 Wailae Avenue, Apt. 605, Honolulu, HI 96816 (US).

(72) Inventors: **ISLAM, Mohammed, Shahidul**; Bangladesh Jute Research Institute, Manik Mia Avenue, Dhaka, 1207 (BD). **AHMED, Borhan**; Bangladesh Jute Research Institute, Manik Mia Avenue, Dhaka, 1207 (BD). **HAQUE, Mohammed, Samiul**; Bangladesh Jute Research Institute, Manik Mia Avenue, Dhaka, 1207 (BD). **ALAM, Mohammed, Monjurul**; Bangladesh Jute Research Institute, Manik Mia Avenue, Dhaka, 1207 (BD).

(74) Agents: **GORDON, Dana, M.** et al.; Foley Hoag LLP, 155 Seaport Boulevard, Boston, MA 02210-2600 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(88) Date of publication of the international search report:
13 August 2015

(54) Title: NUCLEOTIDE SEQUENCE ENCODING WUSCHEL-RELATED HOMEOBOX4 (WOX4) PROTEIN FROM CORCHORUS OLITORIUS AND CORCHORUS CAPSULARIS AND METHODS OF USE FOR SAME

(57) Abstract: The present invention discloses isolated polynucleotides encoding WUSCHEL-related homeobox4 proteins from two species of jute plants, namely, the *Corchorus olitorius ("C. olitorius")* and *Corchorus capsularis ("C. capsularis")*, and corresponding polypeptides derived therefrom. The disclosed polynucleotide sequences encode WUSCHEL-related homeobox4 polypeptides (WOX4), which possess catalytic activities in enhancing fiber production in jute. The present invention also relates to the plants having a modulated expression of a nucleic acid encoding a WOX4 polypeptide, which have enhanced fiber yield relative to corresponding wild type plants or other control plants. Vectors, expression constructs and host cells comprising and/or consisting of the nucleotide sequences of the protein are also provided. Also disclosed are methods for producing the proteins and methods for modifying the proteins in order to improve their desirable characteristics. The proteins of the invention can be used in a variety of ways, including inducing, initiating, improving, or enhancing plant growth, plant height, fiber and seed yield.

WO 2015/077447 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US14/66599

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N15/82; C07K 14/415; A01H 5/10 (2015.01)
CPC - C12N 15/8261, 15/8273; C07K 14/415
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)
IPC(8): C12N15/82; C07K 2/00, 14/415; C07H 21/04; A01H 5/00, 5/10; C12N 5/10, 15/63 (2015.01)
CPC: C12N 15/8261, 15/8271, 15/8273; C07K 14/415

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); Google Scholar; ScienceDirect; Dialog ProQuest; olitorius, recombinant, link*, homeobox, 'homeobox4,' yield, polynucleotide, polypeptide, fiber, increase, 'Wuschel,' capsularis

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011/0283420 A1 (HATZFELD, Y et al.); November 17, 2011; abstract; SEQ ID NOS: 437, 438; paragraphs [0075], [0089], [0090], [0091], [0096], [0123], [0164], [0280]	1-14
A	US 2011/0004958 A1 (ALONI, R et al.); January 6, 2011; paragraphs [0016], [0019]	5-12
A	OHMORI, Y et al. WUSCHEL-Related HOMEOBOX4 Is Involved In Meristem Maintenance And Is Negatively Regulated By The CLE Gene FCP1 In Rice. The Plant Cell. 31 January 2013; abstract.	1-14

☐ Further documents are listed in the continuation of Box C. ☐

* 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	Date of mailing of the international search report
05 March 2015 (05.03.2013)	23 MAR 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105164262 A
(43)申请公布日 2015.12.16

(21)申请号 201480003587.X
(22)申请日 2014.11.20
(30)优先权数据
61/907617 2013.11.22 US
(85)PCT国际申请进入国家阶段日
2015.06.18
(86)PCT国际申请的申请数据
PCT/US2014/066599 2014.11.20
(87)PCT国际申请的公布数据
WO2015/077447 EN 2015.05.28
(71)申请人 孟加拉朱特研究所
地址 孟加拉国达卡
(72)发明人 M. 阿拉姆 M.S. 伊斯拉姆
B. 艾哈迈德 M.S. 哈克
M.M. 阿拉姆
(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 李波 黄希贵
(51)Int.Cl.
C12N 15/29(2006.01)
C07K 14/415(2006.01)
C12N 15/82(2006.01)
A01H 5/00(2006.01)

权利要求书2页 说明书17页 附图1页

(54)发明名称

来自长蒴黄麻和圆果种黄麻的编码WUSCHEL相关的同源框4(WOX4)蛋白的核苷酸序列和其使用方法

(57)摘要

本发明公开了分离的多核苷酸，其编码来自两个黄麻植物物种即长蒴黄麻(“C.olitorius”)和圆果种黄麻(“C.capsularis”)的WUSCHEL相关的同源框4蛋白，和源自其的相应多肽。公开的多核苷酸序列编码WUSCHEL相关的同源框4多肽(WOX4)，其具有增强黄麻中纤维生产的催化活性。本发明也涉及调制表达编码WOX4多肽的核酸的植物，其相对于相应的野生型植物或其他对照植物具有改善的纤维产生。也提供了载体、表达构建体和宿主细胞，其包括和/或由所述蛋白质的核苷酸序列组成。也公开了产生蛋白质的方法和修饰所述蛋白质以便改善它们期望特征的方法。本发明的蛋白质可以各种方式使用，包括诱导、启动、改善或增强植物生长、植物高度、纤维和种子产生。

CN 105164262 A

1. 分离的多核苷酸，其编码WUSCHEL相关的同源框4多肽并且源自植物长蒴黄麻，其包括选自下述的核酸分子：

a) 包括SEQ ID NO 2中阐释的核苷酸序列的核酸分子，或其互补物；和

b) 包括与SEQ ID NO:2中阐释的核苷酸序列具有至少95％序列同一性的核苷酸序列的核酸分子，或其互补物。

2. 根据权利要求1所述分离的多核苷酸编码，其中所述长蒴黄麻的植物是品种O-4。

3. 分离的多核苷酸，其编码WUSCHEL相关的同源框4多肽并且源自植物圆果种黄麻，其包括选自下述的核酸分子：

a) 包括SEQ ID NO:5中阐释的核苷酸序列的核酸分子，或其互补物；和

b) 包括与SEQ ID NO:5中阐释的核苷酸序列具有至少95％序列同一性的核苷酸序列的核酸分子，或其互补物。

4. 根据权利要求3所述分离的多核苷酸，其中所述圆果种黄麻的植物是品种CVL-1。

5. 分离的WUSCHEL相关的同源框4多肽，其包括SEQ ID:NO:3中阐释的氨基酸序列，或其生物活性片段，所述多肽包括选自催化长蒴黄麻的植物中启动、形成、增强和改变从而修饰韧皮部纤维组成的功能。

6. 根据权利要求5所述的分离的多肽，其中所述长蒴黄麻的植物是品种O-4。

7. 根据权利要求5所述的分离的多肽，其中所述多肽包括与SEQ ID NO:3中阐释的氨基酸序列至少95％的序列同一性。

8. 分离的WUSCHEL相关的同源框4多肽，其包括SEQ ID NO:6中阐释的氨基酸序列，或其生物活性片段，所述多肽包括选自催化圆果种黄麻的植物中启动、形成、增强和改变从而修饰韧皮部纤维组成的一个或多个功能。

9. 根据权利要求8所述的分离的多肽，其中所述圆果种黄麻的植物是品种CVL-1。

10. 根据权利要求8所述分离的多肽，其中所述多肽包括与SEQ ID NO:6中阐释的氨基酸序列至少95％的序列同一性。

11. 重组基因构建体，其包括选自下述的多核苷酸：

a) 包括SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列的核酸分子，或其互补物；和

b) 包括与SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列具有至少95％序列同一性的核苷酸序列的核酸分子，或其互补物，

其中所述多核苷酸在宿主细胞中是可表达的以在长蒴黄麻和圆果种黄麻的植物中产生WUSCHEL相关的同源框4多肽的同系物。

12. 根据权利要求11所述的重组基因构建体，进一步包括可操作地连接至a)或b)中阐释的核酸分子的启动子区域，其中所述启动子增强所述核酸分子的转录或表达。

13. 转化株，其包括能够表达多核苷酸的重组基因构建体，所述多核苷酸选自下述：

a) 包括SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列的核酸分子，或其互补物；和

b) 包括与SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列具有至少95％序列同一性的核苷酸序列的核酸分子，或其互补物，

其中所述转化株产生WUSCHEL相关的同源框4多肽的同系物。

14.产生相对于对照植物具有增加或改善的纤维产率的植物或转基因植物的方法，包括：

a) 在植物细胞中引入包括多核苷酸的重组基因构建体，所述多核苷酸选自：

i) 包括 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列的核酸分子，或其互补物；和

ii) 包括与 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列具有至少 95％序列同一性的核苷酸序列的核酸分子，或其互补物，

b) 在促进植物生长和发育的条件下培养植物细胞；和

c) 表达选自下述的多肽：

i) 包括 SEQ ID:NO:3 或 SEQ ID NO:6 中阐释的氨基酸序列的多肽，或其生物活性片段；和

ii) 与 SEQ ID NO:3 或 SEQ ID NO:6 中阐释的氨基酸序列具有至少 95％序列同一性的多肽。

来自长蒴黄麻和圆果种黄麻的编码 WUSCHEL 相关的同源框 4(WOX4) 蛋白的核苷酸序列和其使用方法

[0001] 相关申请

[0002] 本申请要求 2013 年 11 月 22 日提交的美国临时申请号 61/907,617 的益处 ;其全部内容通过引用并入本文。

技术领域

[0003] 本发明一般涉及分子生物学领域并且涉及通过调节植物中编码 WUSCHEL 相关的同源框 4 蛋白 (WOX4) 多肽的核酸的表达,增强纤维产率相关的特性的方法。更具体地,本发明提供了分离自两个黄麻植物物种,即,长蒴黄麻和圆果种黄麻的 WUSCHEL 相关的同源框 4(WOX4) 同系物,和其在产生黄麻植物的韧皮纤维中的应用,以及其转基因黄麻植物。

背景技术

[0004] 黄麻是环境友好的和生物可降解的天然纤维。其是可再生资源,每单位土地面积具有高的生物质产生。超过 100 个黄麻物种 (包括野生近亲缘类) 产生天然韧皮纤维。黄麻由黄麻属植物产生,其曾经归类为椴树科,最近归类为锦葵科。但是,这些物种中的仅仅两个,长蒴黄麻和圆果种黄麻,产生适于工业目的使用的高质量纤维。作为天然纤维,黄麻可用于各种途径,补充或替换人工合成材料,并且已经接收增加来自工业的注意。随着全球对黄麻纤维需求的增加,需要进一步改善黄麻纤维产生。因此,研究纤维生物合成和开发涉及该纤维生物合成的分子生物学非常令人感兴趣。

[0005] 黄麻的纤维是木质部外纤维,其由两种类型的纤维组成和 / 或包括两种类型的纤维 :(i) 在原生韧皮部区域通过细胞分裂和修饰,由原形成层发育的主要韧皮部纤维,和 (ii) 由形成层通过纺锤状活性和射线起始发育的第二韧皮部纤维 (Maiti 和 Mitra,1972. Bull Bot Soc Bengal26:79 - 85)。这些原形成层和形成层组织通过导管元件分化抑制因子 (TIDF) 和 TDIF 受体 (TDR) 膜蛋白质激酶介导细胞与细胞通信,其促进原形成层细胞的增殖和抑制它们的木质部分化 (图 1 ;Hirakawa 等,2010). Plant Cell,22:2618-2629 ;Etchells 等,2013. Development 140,2224-2234)。

[0006] WUSCHEL 相关的同源框 (WOX) 基因家族在启动和 / 或保持各种胚芽、分生组织和器官初始细胞期间进行相关的功能 (Haecker 等,2004)。在 WUSCHEL 相关的同源框 (WOX) 基因家族蛋白质中,WOX4 用作 TDIF 信号传导途径的关键调节剂 (Hirakawa 等 2010) 并且优选在原形成层和形成层中表达 (Schrader 等,2004 ;Ji 等,2010 和 Hirakawa 等 2010)。例如,TDIF-TDR 诱导主转录因子 WUSCHEL 相关的同源框 4(WOX4) 的转录,其促进保持拟南芥属和西红柿中的原形成层 / 形成层干细胞。

[0007] WUSCHEL 相关的同源框 4(WOX4) 多肽催化植物中韧皮纤维的启动。但是,还未有许多与该多肽相关的表征报道或存在现有技术提供的技术。美国专利号 2011/0283420A1(通过参考并入) 已经公开了 wuschel 相关的同源框 1 样 (WOX1 样) 多肽,用于改善植物中相关的特征的产率。在另一欧洲专利号 1451301B1 中,公开了 wuschel 基因在促进植物中体

细胞胚芽发生中的用途。最近，美国专利号 2010/0100981A1（通过参考并入）中公开了一些 wuschel 基因同系物。

[0008] 鉴于 WUSCHEL 相关的同源框 4 蛋白可在黄麻纤维的生物合成途径中其重要作用的事实，期望通过开发和利用 WUSCHEL 相关的同源框 4(WOX4) 的分子生物学和遗传信息，工业提供植物中与纤维的生物合成相关的遗传方法。另外，因为植物每个物种的纤维生物合成途径和遗传组成通常不同，所以物种特异性方法也是优选的以便优化从黄麻植物的纤维产率，并且获得兼容的结果，以确保工业中的使用。

发明内容

[0009] 本发明的一个方面等是提供基因编码蛋白质，源自长蒴黄麻和圆果种黄麻，其参与催化它们韧皮纤维形成的启动，并且具有 WUSCHEL 相关的同源框 4(WOX4) 样序列。更具体而言，本发明提供了具有诱导韧皮部纤维能力的基因，因此提供了从而编码的蛋白质，以及其用途。

[0010] 本发明的另一目的提供了待开发 / 使用而用于改善长蒴黄麻和圆果种黄麻植物以及其他产生韧皮纤维的植物中纤维的生产、生长、强度和产率的 WUSCHEL 相关的同源框 4(WOX4) 的分子生物学和遗传信息。

[0011] 仍本发明的另一目的是通过调节 WUSCHEL 相关的同源框 4(WOX4) 在植物中的生物合成，获得具有增加的纤维产生的长蒴黄麻和圆果种黄麻的转基因植物。

[0012] 本发明的仍另一目的是提供分离的多核苷酸，其具有特定的核苷酸序列，其可利于公开方法的表现，并且提供对转基因长蒴黄麻和圆果种黄麻植物的利用。

[0013] 本发明的进一步目的是提供潜在的商业上可行的方式，以增加用于黄麻纤维基产品的纤维的生产。

[0014] 本发明提供了分离自长蒴黄麻的基因，其编码具有如 SEQ ID NO :3 中阐释的氨基酸序列的蛋白质，其参与催化韧皮部纤维形成的启动并且具有 WUSCHEL 相关的同源框 4(WOX4) 样序列。本发明进一步提供编码蛋白质的基因，所述蛋白质具有通过添加或缺失一个或多个氨基酸和 / 或用如 SEQ ID NO :3 中阐释的氨基酸序列的其他氨基酸替换而修饰的氨基酸序列，其参与催化韧皮部纤维的形成的启动并且具有 WUSCHEL 相关的同源框 4(WOX4) 样序列。本发明进一步提供了与 SEQ ID NO :1 中阐释的核酸杂交的基因，尤其其 DNA 或其一部分，并且编码参与催化韧皮部纤维的形成的启动并且具有 WUSCHEL 相关的同源框 4(WOX4) 样序列的蛋白质。

[0015] 本发明也提供了分离自圆果种黄麻的基因，其编码具有如 SEQ ID NO :6 中阐释的氨基酸序列的蛋白质，其参与催化韧皮部纤维的形成的启动并且具有 WUSCHEL 相关的同源框 4(WOX4) 样序列。本发明进一步提供了编码蛋白质的基因，所述蛋白质具有通过添加或缺失一个或多个氨基酸和 / 或用如 SEQ ID NO :6 中阐释的氨基酸序列的其他氨基酸替换而修饰的氨基酸序列，其参与催化韧皮部纤维的形成的启动并且具有 WUSCHEL 相关的同源框 4(WOX4) 样序列。本发明进一步提供了与 SEQ ID NO :4 中阐释的核酸杂交的基因，尤其其 DNA 或其一部分，并且编码参与催化韧皮部纤维的形成的启动并且具有 WUSCHEL 相关的同源框 4(WOX4) 样序列的蛋白质。

[0016] 根据本发明的一种优选的实施方式，使用的长蒴黄麻的植物是品种 O-4，和使用的

圆果种黄麻的植物是品种 CVL-1。

[0017] 本发明的另一实施方式公开重组基因构建体包括具有 SEQ ID NO :2 和 / 或 SEQ ID NO :5 中阐释的核苷酸序列的多肽，其中多核苷酸在宿主细胞中是可表达的，以分别在长蒴黄麻和圆果种黄麻的植物中产生 WUSCHEL 相关的同源框 4(WOX4) 的同系物。

[0018] 本发明的进一步实施方式是包括重组基因构建体的转化株，所述重组基因构建体能够表达具有 SEQ ID NO :2 和 / 或 SEQ ID NO :5 中阐释的核苷酸序列的多核苷酸，以产生 WUSCHEL 相关的同源框 4(WOX4) 蛋白的同系物。

[0019] 本发明也提供了产生蛋白质的方法，所述蛋白质参与韧皮部纤维启动的催化活性，并且具有 WUSCHEL 相关的同源框 4(WOX4) 样序列，所述方法包括培养和 / 或栽培上述宿主。

[0020] 本发明也提供了诱导植物或植物细胞的韧皮部纤维启动的方法 ；此外，其也公开了包括和 / 或由下述组成的方法 ：将上述基因引入植物或植物细胞，并且驱动所述基因的表达。

[0021] 本领域技术人员容易认识到，本发明非常适于进行目标并且获得提到的目的和优势，以及其中固有的那些。本文所述的实施方式不旨在限制本发明的范围。

[0022] 参考下述说明书和权利要求，将更容易理解本发明的这些和其他特征、方面和优势。

附图说明

[0023] 图 1 显示 TDIF-TDR 信号传导途径。TDIF-TDR 信号传导分成两个途径，一个是参与韧皮部细胞增殖的 WOX4，其促进纤维细胞启动和另一个是假设因子 X，其抑制脉管干细胞的木质部分化。

[0024] 图 2 显示系统发生树，其比较来自长蒴黄麻的 SEQ ID NO.3 和来自圆果种黄麻的 SEQ ID NO.6，以及产生 WUSCHEL 相关的同源框 4(WOX4) 蛋白的其他氨基酸序列。

[0025] 发明详述

[0026] 从下述形成本申请一部分的详细说明书和附图可更充分理解本发明。

[0027] 本文提供的定义和 / 或方法限定本发明并且在本发明的实践中指导本领域技术人员。除非以其他方式指明，应根据相关领域技术人员的常规的用途理解术语。在任何定义和 / 或方法与本文并入的任何专利或非专利参考文献提供的或其他地方出现的任何参考文献的任何定义和 / 或方法不一致的情况下，应理解，本文使用在本申请中已经清楚提供 / 改编的所述定义和 / 或方法。单数术语“一个 (a/an)”和“所述”包括复数参照物，除非上下文清楚得另外指出。类似地，词语“或”旨在包括“和”，除非上下文清楚得另外指出。因此，“包括 A 或 B”意思是包括 A 或 B，或 A 和 B。进一步理解，为核酸或多肽提供的所有基础尺寸或氨基酸尺寸，和所有分子量或分子量值是近似的，并且提供为了说明。尽管与本文所述的那些类似或相同的方法和材料可用于本公开的实践或测试，但是下面描述了适当的方法和材料。

[0028] 本发明涉及分离的多核苷酸和源自其的相应多肽，所述多核苷酸编码提取自长蒴黄麻和圆果种黄麻的 WUSCHEL 相关的同源框 4 蛋白。更具体地，本发明提供了 WUSCHEL 相关的同源框 4 同系物，和它们在改善都为黄麻植物物种的长蒴黄麻和圆果种黄麻中纤维产

生的用途，以及相关的转基因长蒴黄麻和圆果种黄麻植物。主要通过比较长蒴黄麻和圆果种黄麻基因组 DNA 的核苷酸序列和其他植物已知酶基因的核苷酸序列鉴定本发明编码酶的基因组序列。在本发明之前，f 这些长蒴黄麻和圆果种黄麻基因的核苷酸序列、阅读框、外显子和内含子的位置、酶的结构，以及它们在开发高纤维产率潜能的黄麻植物中的潜在用途是未知的。

[0029] 商业上栽培的黄麻物种，长蒴黄麻和圆果种黄麻的基因组序列的分析揭示两个物种都具有编码具备优选用于增强纤维产生的催化活性的酶的单个基因。起初由软件程序，比如 Augustus，半 HMM 基核酸分析程序 (SNAP)、Geneid(Genome BioInformatics Research Lab) 注释核苷酸序列，其可鉴定推定编码区域，内含子，和拼接接头。进一步进行自动和手动管理核苷酸序列，以优化和确立编码区域和其他基因特征的精确表征。

[0030] 部分或完全测序来自长蒴黄麻的超过 30,096cDNA 和来自圆果种黄麻的 37,031cDNA。从它们中，开发了来自每个物种编码新酶的单个 cDNA，具有推定的作用，优选的增强黄麻中纤维产生的催化活性。

[0031] 在对源自长蒴黄麻和圆果种黄麻 mRNA 的 cDNA 文库的克隆完全或部分测序之后分析开放阅读框 (ORF)，并且使用序列分析软件进一步分析，并且通过测定与数据库中已知序列的同源性（公共的 / 私人的）。

[0032] 本公开的上下文中，遍及说明书使用的许多术语具有指示的意思，除非明确指出具有不同的意思。

[0033] 如本文所使用，“多核苷酸”是比如核酸片段的核苷酸序列。多核苷酸可以是单链或双链的 RNA 或 DNA 的聚合物，其任选地包含合成的、非天然的或改变的核苷酸碱基。为 DNA 聚合物形式的多核苷酸可包括和 / 或由 cDNA、基因组 DNA、合成 DNA，或其混合物 / 组合的一个或多个区段组成。本发明分离的多核苷酸可包括源自 SEQ ID NO.1 和 SEQ ID NO.4 的至少一条 150 个连续核苷（上游和下游二者），或这些序列的互补序列。

[0034] 如本文所使用，“多肽”是通过肽键结合在一起的线性单链氨基酸，并且通常长度具有大于 100 个氨基酸的序列。

[0035] “分离的”意思是“通过手动”改变天然状态。如果组合物或物质在自然中存在，如果已经被或去除从其初始环境，或二者，其已经是被“分离的”。例如，天然存在于活植物或动物中的多核苷酸或多肽不是“分离的”，但是从其天然状态的共存材料分开的相同多核苷酸或多肽是“分离的”，如本文所采用的术语。

[0036] 如本文所使用，术语“基因”定义为植物长蒴黄麻和圆果种黄麻的基因组序列，尤其是编码 WUSCHEL 相关的同源框 4 酶多肽序列的多核苷酸序列，所述酶参与黄麻中增强纤维生长的优选的催化活性。术语可进一步包括核酸分子，包括上游、下游和 / 或内含子核苷酸序列。

[0037] “编码序列”或“编码区域”指当序列表达时，具有足以以产生基因产物，比如氨基酸或多肽的序列信息的核酸分子。编码序列可包括和 / 或由下述组成 :翻译区域中的非翻译序列（包括内含子或 5’ 或 3’ 非翻译区），或可缺少这种中间非翻译序列（例如，如在 cDNA 中）。

[0038] 如本文所使用，术语“寡核苷酸”，是短多核苷酸或一部分多核苷酸，其长度可优选地包括 10-1000，最优选地 12 至 50 个核苷。在本发明的实施方式中，寡核苷酸中包含的核

苷可以是天然存在的核苷的类似物或衍生物。

[0039] 如本文所使用，术语“引物”是能够结合靶核酸序列和启动核酸合成的寡核苷酸。如本文定义的扩增寡核苷酸可优选地长度是 10 至 50，最优选地 15 至 25 个核苷。此外，本发明的扩增寡核苷酸可以是化学合成的并且这样的寡核苷酸不是天然存在的核酸。

[0040] 遍及说明书使用的缩写指包括核苷酸序列的核酸是常规的单字符缩写。因此当包括在核酸中时，天然存在的编码核苷的缩写如下：腺嘌呤 (A)、鸟嘌呤 (G)、胞嘧啶 (C)、胸苷嘧啶 (T) 和尿嘧啶 (U)。而且，除非另外指出，本文存在的核酸序列是 5 □→ 3 □方向。

[0041] 如本文所使用，使用术语“互补的”和其衍生词指核酸通过熟知的规则配对，A 与 T 或 U 配对和 C 与 G 配对。互补可以是“部分的”或“完全的”。在部分互补中，仅仅一些核酸碱基根据碱基对规则匹配；而在完全的或全部互补中，所有的碱基根据配对规则匹配。核酸链之间的互补程度可对核酸链之间的杂交的效率和强度有显著作用，如本领域熟知的。所述杂交的效率和强度取决于探测方法。

[0042] 如本文所使用，术语“宿主细胞”包括易于用核酸构建体或表达载体转化、转染、转导、表达等的任何细胞类型，所述核酸构建体或表达载体包括和 / 或由本发明的多核苷酸组成。适当的宿主细胞包括真菌和 / 或植物细胞，尤其生产韧皮纤维的植物细胞。

[0043] 术语“可操作地连接”在本文通常指其中控制序列相对于多核苷酸序列的编码序列布置在适当位置的构造，使得控制序列引导多肽的编码序列的表达。例如，当启动子影响编码序列的表达时，即该编码序列在启动子的转录控制下，启动子可以可操作地连接编码序列。

[0044] “载体”一般指复制子，比如质粒、噬菌体、黏粒、酵母或病毒，另一核酸区段可可操作地插入其中，以便使得该区段复制或表达。术语“载体”也旨在指能够运输其已经连接的另一核酸的核酸分子。一种类型的载体是“质粒”，其指其中可连接另外的 DNA 区段的环形双链 DNA 环。另一类型的载体是病毒载体，其中另外 DNA 区段可连接至病毒基因组。某些载体能够在引入它们的宿主细胞中自主复制（例如，具有细菌复制起点的细菌载体和附加体哺乳类的载体）。当引入宿主细胞时，其他载体可整合至宿主细胞的基因组，并且从而与宿主基因组一起复制。而且，某些载体能够引导它们可操作连接的基因的表达。这样的载体在本文称为“重组表达载体”（或简单地称为“表达载体”）。一般而言，在重组 DNA 技术中使用的表达载体通常为质粒的形式。在本说明书中，“质粒”和“载体”可交替使用，因为质粒是最常使用的载体形式。但是，本发明旨在包括这样的其他形式的表达载体，比如病毒载体（例如，复制缺陷逆转录病毒、腺病毒和腺伴随病毒），其其相同的功能。

[0045] 术语“核酸构建体”或“DNA 构建体”有时用于指编码序列或可操作地连接至适当的调节序列并且插入载体用于转化细胞的序列。该术语可与术语“转化 DNA”或“转基因”交替使用。

[0046] 如本文所使用，术语“启动子”指用于引导下游基因转录的核酸序列。启动子一般适于其中表达靶基因的宿主细胞。启动子与其他转录和翻译调节核酸序列（也称为“控制序列”）一起对于表达给定基因是必要的。一般而言，转录和翻译调节序列包括但不限于启动子序列、核糖体结合位点、转录起始序列和终止子序列、翻译起始序列和终止子序列，以及增强子或活化序列。

[0047] 如本文所使用，“WUSCHEL 多肽”或“WUS 多肽”意思是具有 wuschel 活性，即，参与

植物中启动和保持干细胞的多肽。Wuschel 活性刺激细胞生长，包括干细胞。Wuschel 是植物同源域蛋白质，包括‘不规则’（相比动物同源域基序）螺旋 - 环 - 螺旋 - 转角 - 螺旋同源域基序，在结构域的环和 / 或转角包括额外的氨基酸残基。Wuschel 蛋白质可进一步包括和 / 或由其他保守基序组成，比如两个保守 wuschel C- 末端结构域，(E/R)TLPLFP 和 A(A/S)LEL(S/T)L 结构域。术语也包括具有上述功能任何一种的片段、变体和同系物。

[0048] 如本文所使用，术语“同系物”指蛋白质包括肽、寡肽、多肽、蛋白质和酶，其相对于讨论的未修饰蛋白质具有氨基酸替换、缺失和 / 或插入并且具有与它们从中衍生的未修饰的蛋白质类似的生物和功能活性。

[0049] 缺失指从蛋白质去除一个或多个氨基酸。

[0050] 插入指一个或多个氨基酸残基引入蛋白质中的预定位点。插入可包括 N- 末端和 / 或 C- 末端融合以及序列内插入单个或多个氨基酸。一般而言，在氨基酸序列内插入比 N- 或 C- 末端融合小约 1 至 10 个残基的级别。

[0051] 置换指用具有类似特性（比如类似的疏水性、亲水性、抗原性、偏好，以形成或破坏 α - 螺旋结构或 β - 折叠结构）的其他氨基酸替换蛋白质的氨基酸。氨基酸置换通常为单个残基，但是取决于对多肽所起的作用功能约束可聚簇并且范围可为 1 至 10 个氨基酸；插入通常为约 1 至 10 个氨基酸残基的级别。氨基酸置换优选地为保守氨基酸置换。保守置换表是本领域熟知的（见例如 Creighton(1984)Protein.W.H.Freeman 和 Company(Eds) 和下面表 1)。

[0052] 表 1 :保守氨基酸置换的例子

[0053]

残基	保守替换	残基	保守替换
Ala	Ser	Leu	Ile ;Val
Arg	Lys	Lys	Arg ;Gln
Asn	Gln ;His	Met	Leu ;Ile
Asp	Glu	Phe	Met ;Leu ;Tyr
Gln	Asn	Ser	Thr ;Gly
Cys	Ser	Thr	Ser ;Val
Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp ;Phe
His	Asn ;Gln	Val	Ile ;Leu
Ile	Leu, Val		

[0054] 氨基酸置换、缺失和 / 或插入可容易使用本领域熟知的肽合成技术，比如固相肽

合成和/或任何其他合成技术，或通过重组DNA操作进行。操作DNA序列以产生蛋白质的置换、插入或缺失变体的方法是本领域熟知的。例如，在DNA的预定位点形成置换突变的技术是本领域技术人员熟知的并且包括M13诱变、T7-Gen体外诱变(USB，Cleveland，OH)、Quick Change位点定向诱变(Stratagene，San Diego，CA)、PCR介导的位点定向诱变或其他位点定向诱变方案。

[0055] 如本文所使用，“生物活性部分”可指WUSCHEL相关的同源框4蛋白的具有催化长蒴黄麻或圆果种黄麻的植物中启动、形成、增强或改变韧皮部纤维组成的生物活性的片段。WUSCHEL相关的同源框4蛋白的生物活性部分包括肽或多肽，其包括与WUSCHEL相关的同源框4蛋白足够相同或源自WUSCHEL相关的同源框4蛋白的氨基酸序列的氨基酸序列，例如，SEQ ID NO:3或SEQ ID NO:6中阐释的氨基酸序列，其包括比全长WUSCHEL相关的同源框4蛋白更少氨基酸，并且展示WUSCHEL相关的同源框4蛋白的至少一种活性。典型地，生物活性部分包括具有WUSCHEL相关的同源框4蛋白的至少一种活性的结构域或基序。WUSCHEL相关的同源框4蛋白的生物活性部分可以是，例如，长度为210、211、212、213、214、215、216、217、218、219、220、221、222或223个氨基酸的多肽。

[0056] WUSCHEL相关的同源框4蛋白可具有SEQ ID NO:3或SEQ ID NO:6中阐释的氨基酸序列。在其他实施方式中，WUSCHEL相关的同源框4蛋白基本上与SEQ ID NO:3或SEQ ID NO:6相同，并且保持SEQ ID NO:3或SEQ ID NO:6蛋白质的功能活性，但是氨基酸序列由于天然等位基因变化或诱变而不同。在另一实施方式中，WUSCHEL相关的同源框4蛋白包括与SEQ ID NO:3或SEQ ID NO:6具有至少约95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更多同一性的氨基酸序列。

[0057] 如本文所使用，术语“结构域:指沿着进化上亲缘的序列比对在特定位置保守的一组氨基酸。尽管在其他位置上的氨基酸可能在同系物之间不同，但是在特定位置高度保守的氨基酸指示可能对于蛋白质的结构、稳定性或功能是必须的氨基酸。通过在蛋白质同系物家族的对比序列中它们高度的保守鉴定，它们可用作标识符，以确定任何讨论的多肽是否属于先前鉴定的多肽家族。

[0058] 如本文所使用，术语“基序“指进化上亲缘蛋白质的序列中短的保守区域。基序通常是结构域的高度保守部分，但是也可包括结构域的仅仅一部分，或位于保守结构域的外部(如果基序的所有的氨基酸落在限定的结构域的外部)。

[0059] 为了鉴定结构域而存在的专用数据库，例如，SMART(Schultz等(1998)Proc. Natl.Acad.Sci.USA 95,5857-5864;Letunic等(2002)Nucleic Acids Res 30,242-244)，InterPro(Mulder等，(2003)Nucl.Acids.Res.31,315-318)，Prosite(Bucher和Bairoch(1994)，用于自动序列阐释的生物分子序列基序和其功能的一般轮廓特征。(In)ISMB-94;第二届分子生物学智能系统的国际大会的会议记录。Altman R.，Brutlag D.，Karp P.，Lathrop R.，Searls D.，Eds.，pp53-61，AAAI Press，Menlo Park;Hulo等，Nucl. Acids.Res.32:D134-D137，(2004))，或Pfam(Bateman等，Nucleic Acids Research 30(1):276-280(2002))。用于计算机分析蛋白质序列的一组工具在ExPASy蛋白质组服务器中是可用的(Swiss Institute ofBioinformatics(Gasteiger等，ExPASy:用于深度蛋白质知识和分析的蛋白质组服务器，Nucleic Acids Res.31:3784-3788(2003))。可使用常规技术，

比如通过序列比对来鉴定结构域或基序也。

[0060] 为了本发明的目的，“转基因的”、“转基因”或“重组“意思是例如，核酸序列、表达盒、基因构建体或载体包括和/或由核酸序列组成或生物体用根据本发明的核酸序列、表达盒或载体转化，所有的那些构建体通过重组方法产生，其中：(a) 编码蛋白质的核酸序列用于本发明的方法，或 (b) 遗传控制序列（一条或多条）可操作地连接根据本发明的核酸序列，例如启动子，或 (c)a) 和 b) 不在它们的天然遗传环境中或已经通过重组方法修饰。修饰可为，例如置换、添加、缺失、倒置或插入一个或多个核苷酸残基的形式。天然遗传环境理解为意思是在起始植物中的天然基因组或染色体基因座或存在于基因组文库中。在基因组文库的情况下，优选地至少部分保持核酸序列的天然遗传环境。环境在核酸序列至少一侧的侧翼并且序列长度为约 50bp，优选地为约 500bp。天然存在的表达盒 - 例如天然存在的核酸序列的天然启动子与编码用于如上定义的本发明方法的多肽的相应的核酸序列组合 - 当该表达盒通过非天然合成的（”人工“）方法比如，例如，诱变处理修饰时变成转基因表达盒。适当的方法描述在，例如 US 5,565,350（通过参考并入）或 WO00/15815（通过参考并入）中。

[0061] 为了本发明的目的，转基因植物因此理解为包括其中在本发明的方法中使用的核酸不在它们植物的所述基因组的的天然基因座的那些植物，并且因此核酸同源或异源表达是可能的。但是，如所提到的，转基因也意思是，尽管根据本发明的或在本发明的方法中使用的核酸在它们植物基因组的天然位置，相对于天然序列，序列已经被修饰，和/或天然序列的调节序列已经被修饰。转基因优选地理解为意思是根据本发明的核酸在基因组的非天然基因座的表达，其中核酸发生同源的或，优选地，异源表达。

[0062] 如本文所使用，术语“引入”、或“转化”指本文包括外源性多核苷酸转化至宿主细胞，无论用于转化的方法如何。能够随后克隆繁殖的植物组织，无论通过器官发生或胚芽发生，可用本发明的遗传构建体转化并且从其再生完整植物。取决于可用的克隆繁殖系统，和最适于待转化的具体物种，选择的具体组织不同。示例性组织靶包括叶盘、花粉、胚芽、子叶、下胚轴、雌配子体、愈伤组织、存在的分生组织（例如，顶端分生组织、腋芽和根分生组织），和诱导分生组织（例如，子叶分生组织和下胚轴分生组织）。多核苷酸可瞬时或稳定引入宿主细胞并且可以保持为非整合形式，例如，质粒。可选地，其可整合至宿主基因组。所得转化植物细胞可然后用于以本领域熟知的方式再生转化植物。

[0063] 外源基因转移至植物的基因组称为转化。

[0064] 植物物种的转化现在是相当常规的技术。有利地，任何数个转化方法可用于将感兴趣的基因引入适当的祖先细胞。为从植物组织或植物细胞转化和再生植物描述的方法可用于瞬时或稳定转化。转化方法包括使用脂质体、电穿孔和增加游离 DNA 吸收的化学品、DNA 直接注入植物、粒子枪轰击，和使用病毒或花粉和微投射的转化。转基因植物，包括转基因作物植物优选地经土壤杆菌介导的转化产生。有利的转化方法是在植物学中的转化 (Sajib 等，Plant Cell Tiss.Organ Cult.(2008)95,333-34)。

[0065] 通常，转化之后，选择存在由与感兴趣的基因一起转化，植物可表达的基因编码的一个或多个标记物的植物细胞或细胞组，其后将转化材料再生至完整植物。为了选择转化的植物，转化的获得的植物材料通常经受选择性条件，从而转化的植物可与未转化的植物不同。例如，可种植以上述方式获得的种子，并且在初始生长阶段之后，通过喷雾经受适当

的选择。进一步可能性由如果需要，在消毒之后，在使用适当的选择试剂的琼脂平板上的种子生长组成，从而仅仅转化的种子可生长为植物。可选地，用存在的选择标记比如上述那些筛选转化的植物。

[0066] 在DNA转移和再生之后，也可评估推定转化的植物，例如使用Southern分析，感兴趣的基因的存在、拷贝数和/或基因组组织。可选地或另外，可使用Northern和/或Western分析监测新引入的DNA的表达水平，两种技术是本领域技术人员熟知的。

[0067] 产生的转化植物可通过各种方式繁殖，比如通过克隆繁殖或经典繁殖技术。例如，第一代（或T1）转化植物可自授粉的并且可选择纯合子第二代（或T2）转化株，和T2植物可然后进一步通过经典繁殖技术繁殖。产生的转化生物体可采用各种形式。例如，它们可以是转化细胞和非转化细胞的嵌合体；克隆转化株（例如，包含表达盒的所有细胞转化）；转化的和未转化的组织的嫁接体（例如，植物中，嫁接至未转化的接穗的转化根茎）。

[0068] 如本文所使用，术语“增加的表达”或“过表达”指比初始野生型表达水平更多的任何形式的表达。

[0069] 本领域充分记载了用于增加基因或基因产物表达的方法并且包括，例如，通过适当的启动子驱动的过表达、使用转录增强子或翻译增强子。用作启动子或增强子元件的分离的核酸可在非异源形式的多核苷酸的适当位置（通常上游）引入，以便上调编码感兴趣多肽的核酸的表达。例如，可通过突变、缺失和/或置换在体内改变内源启动子（见，Kmiec，US 5,565,350；Zarling等WO93/22443），或分离的启动子可以以适当的定向和与本发明基因适当的距离引入植物细胞，以便控制基因的表达。

[0070] 如果期望多肽表达，一般期望在多核苷酸编码区域的3’端包括多腺苷酸化区域。多腺苷酸化区域可源自天然基因、各种其他植物基因或T-DNA。待添加的3’端序列可源自，例如，胭脂氨酸合酶或章鱼碱合酶基因，或可选地源自另一植物基因，或次优选地源自任何其他真核基因。

[0071] 内含子序列也可添加至5’非翻译区（UTR）或部分编码序列的编码序列，以增加积聚在胞质溶胶中成熟的信使的量。已经显示，在植物和动物表达构建体的转录单元中包括可拼接的内含子增加在mRNA和蛋白质水平的基因表达至1000倍（Buchman和Berg(1988) Mol.Cell biol.8：4395-4405；Callis等(1987)Genes Dev 1:1183-1200）。当布置在转录单元的5’端附近时，这样的内含子增强基因表达通常是最大的。玉米内含子Adh1-S内含子1、2和6，Bronze-1内含子的使用是本领域已知的。对于一般的信息见：The Maize Handbook，116章，Freeling和Walbot，Eds.，Springer，N.Y.(1994)。

[0072] 为了测定两个氨基酸序列或两个核酸序列的同一性百分数，可为了最佳比较目的而比对序列（例如，可在第一和第二氨基酸或核酸序列的之一或二者都引入空隙，以为了最佳比对，并且为了比较目的可忽略不同的序列）。为了比较目的，比对的参考序列的长度可以是参考序列的长度的至少95％。可然后比较相应的氨基酸位置或核苷酸位置的氨基酸残基或核苷。当第一序列中的位置被第二序列中相应位置的相同氨基酸残基或核苷酸占据时，分子在该位置是相同的（如本文所使用，氨基酸或核酸“同一性”与氨基酸或核酸“同源性”相同）。两条序列之间的同一性百分数是序列共享的相同位置数量的函数，考虑需要为了两条序列的最佳比对引入空隙的数量，和空隙的长度。

[0073] 序列的比较和两条序列之间的同一性百分数的测定可使用数学算法完成。在一

种实施方式中，两条氨基酸序列之间的同一性百分数可使用 Needleman 和 Wunsch(J.Mol.Biol.48:444-453(1970)) 算法测定，其已经并入 GCG 软件包装的 GAP 程序（在 http://www.gcg.com 可用），使用 Blosum 62 矩阵或 PAM250 矩阵，和空隙权重为 16、14、12、10、8、6 或 4 和长度权重为 1、2、3、4、5 或 6。在仍另一优选的实施方式中，两条核苷酸序列之间的同一性百分数可使用 GCG 软件包装中的 GAP 程序（在 http://www.gcg.com 可用）测定，使用 NWSgapdna.CMP 矩阵并且空隙权重为 40、50、60、70 或 80 和长度权重为 1、2、3、4、5 或 6。在另一实施方式中，两条氨基酸或核苷酸序列之间的同一性百分数可使用 E.Meyers 和 W.Miller(Comput)Appl.Biosci.4:11-17(1988)) 测定，其已经并入 ALIGN 程序（版本 2.0 或 2.0U)，使用 PAM120 权重残基表，空隙长度罚分为 12 和空隙罚分为 4。

[0074] 可用于测定两条序列之间同一性的示例性计算机程序包括但不限于 BLAST 程序套装，例如，BLASTN、BLASTX 和 TBLASTX，BLASTP 和 TBLASTN，在 www.ncbi.nlm.nih.gov/BLAST 是公众可用的。

[0075] 当评估给定的核酸序列相对于 GenBank DNA 序列和其他公开数据库的核酸序列时，通常使用 BLASTN 程序进行序列搜索。对于针对 GenBank 蛋白质序列和其他公开数据库中的氨基酸序列搜索在所有阅读框中翻译的核酸序列，优选 BLASTX 程序。

[0076] 为了测定两条或更多条序列之间的“%同一性”，使用例如 CLUSTAL-W 程序优选的进行选择序列的比对比对。

[0077] 如所阐释，本发明的一种实施方式是编码植物长蒴黄麻和圆果种黄麻的 WUSCHEL 相关的同源框 4 多肽的分离的多核苷酸，其包括和 / 或由分别如 SEQ ID NO 2 和 SEQ ID NO 5 中阐释的核苷酸序列组成。相应地，由核苷酸序列编码的各自的 WUSCHEL 相关的同源框 4 多肽具有 SEQ ID NO 3 和 SEQ ID NO 6 中阐释的氨基酸序列。根据本发明的实施方式，SEQ ID NO 3 指源自长蒴黄麻的 WUSCHEL 相关的同源框 4(WOX4) 同系物的多肽序列，和 SEQ ID NO 6 指源自圆果种黄麻的 WUSCHEL 相关的同源框 4(WOX4) 同系物的多肽序列。两种这些酶出现在长蒴黄麻和圆果种黄麻的植物中纤维的生物合成途径，用于催化启动植物中纤维的生物合成细胞的启动分子。

[0078] 本发明也提供了编码来自植物长蒴黄麻和圆果种黄麻的 WUSCHEL 相关的同源框 4(WOX4) 同系物的基因序列。

[0079] 在一种实施方式中，SEQ ID NO.1 中阐释的 1250bp 长多核苷酸是分离自长蒴黄麻的全长基因。该基因序列包括来自基因上游和下游的至少 150 个连续核苷。这也提供了基因的基因内序列。

[0080] 在另一实施方式中，SEQ ID NO.4 中阐释的 1237bp 长多核苷酸是分离自圆果种黄麻的全长基因。该基因序列包括来自基因上游和下游的至少 150 个连续核苷。这也提供了基因的基因内序列。

[0081] 在本发明的仍另一实施方式中，提供了编码多肽的分离的多核苷酸，其包括 SEQ ID NO 2 和 / 或 SEQ ID NO 5 中阐释的核苷酸序列。SEQ ID NO 2 指源自长蒴黄麻的 WUSCHEL 相关的同源框 4(WOX4) 同系物序列的多核苷酸序列和 SEQ ID NO 5 指源自圆果种黄麻的 WUSCHEL 相关的同源框 4(WOX4) 同系物序列的多核苷酸序列。

[0082] 在仍另一实施方式中，能够编码 WUSCHEL 相关的同源框 4 多肽，或其生物活性片段的分离的核酸分子，包括与 SEQ ID NO :2，SEQ ID NO :5 中阐释的核苷酸序列的全长具

有至少约 95％,96％,97％,98％,99％,99.1％,99.2％,99.3％,99.4％,99.5％,99.6％,99.7％,99.8％,99.9％或更高同一性的核苷酸序列,或其任何互补物。

[0083] 在一种实施方式中,SEQ ID NO.2 中阐释的 669bp 长多核苷酸是全长 cDNA 克隆,其编码 WUSCHEL 相关的同源框 4(WOX4) 蛋白,展示编码如 SEQ ID NO.3 中 222 个氨基酸多肽的开放阅读框,计算的分子量为约 25.54kD。通过 SEQ ID NO.3 的 SMART 分析,揭示序列中存在同源框结构域。这是 DNA- 结合因子,其参与植物的关键发育过程的转录调节。这是 WUSCHEL 相关的同源框 4(WOX4) 蛋白,其参与植物的脉管细胞增殖。

[0084] 在一种实施方式中,SEQ ID NO.5 中阐释的 672bp 长多核苷酸是全长 cDNA 克隆,其编码 WUSCHEL 相关的同源框 4(WOX4) 蛋白,展示编码如 SEQ ID NO.6 中 223 个氨基酸多肽的开放阅读框,计算的分子量为约 25.67kD。通过 SEQ ID NO.6 的 SMART 分析,揭示序列中存在同源框结构域。这是 DNA- 结合因子,其参与植物的关键发育过程的转录调节。这是 WUSCHEL 相关的同源框 4(WOX4) 蛋白,其参与植物的脉管细胞增殖。

[0085] 根据本发明的优选的实施方式,SEQ ID NO 2 中阐释的分离的多核苷酸可通过基因的保守区域的 PCR 扩增获得,使用分离自长蒴黄麻的植物的总 RNA 并且 SEQ ID NO 5 可通过该基因的保守区域的 PCR 扩增获得,使用分离自圆果种黄麻的植物的总 RNA。如前述说明书中阐释,使用的长蒴黄麻的植物是 O-4 品种和使用的圆果种黄麻是 CVL-1 品种。

[0086] 本发明的另一实施方式,公开了重组基因构建体,其包括具有 SEQ ID NO 2 和 / 或 SEQ ID NO 4 中阐释的核苷酸序列的多核苷酸,其中多核苷酸在宿主细胞中是可表达的,并且是可翻译的以产生长蒴黄麻和圆果种黄麻的植物中的 WUSCHEL 相关的同源框 4(WOX4) 蛋白的同系物。扩增、克隆和测序来自长蒴黄麻和圆果种黄麻的植物的 WUSCHEL 相关的同源框 4(WOX4) 的程序进一步详述在实施例 2 中。优选地,重组基因构建体进一步包括可操作地连接,以增强多核苷酸模板表达的启动子区域。在具体启动子的转录控制下,可然后改善包含 SEQ ID NO 2 和 / 或 SEQ ID NO 4 的多核苷酸的重组基因构建体中编码区域的表达,导致更高产率的 WUSCHEL 相关的同源框 4(WOX4) 蛋白。

[0087] 根据本发明的实施方式,调制表达是增加的表达或活性,例如编码 WUSCHEL 相关的同源框 4 多肽的核酸分子,例如,编码 SEQ ID NO2 和 SEQ ID NO 5 的核酸分子的过表达。本领域充分记载了增加核酸或基因,或基因产物表达的方法并且例子提供在定义章节。

[0088] 本发明也提供了产生相对于对照植物具有改善的纤维产率的转基因植物的方法,包括在植物中引入和表达如本文上面定义的编码 WUSCHEL 相关的同源框 4 多肽的任何核酸。

[0089] 更具体而言,本发明提供了产生相比空对照植物具有改善的纤维产率的转基因植物的方法,该方法包括 :

[0090] (i) 在植物或植物细胞中引入和表达编码的 WUSCHEL 相关的同源框 4 多肽核酸或遗传构建体,其包括和 / 或由编码 WUSCHEL 相关的同源框 4 多肽的核酸组成 ;和

[0091] (ii) 在促进纤维细胞生长和发育的条件下培养植物细胞。

[0092] 本发明的另一方面涉及分离的多核苷酸,其编码 WUSCHEL 相关的同源框 4 多肽,并且源自植物长蒴黄麻,其包括选自下述的核酸分子 :

[0093] a) 包括 SEQ ID NO 2 中阐释的核苷酸序列的核酸分子,或其互补物 ;和

[0094] b) 包括与 SEQ ID NO :2 中阐释的核苷酸序列具有至少 95％序列同一性的核苷酸

序列的核酸分子，或其互补物。

[0095] 在某些实施方式中，长蒴黄麻的植物是品种 O-4。

[0096] 本发明的另一方面涉及分离物，其编码 WUSCHEL 相关的同源框 4 多肽并且源自植物圆果种黄麻，其包括选自下述的核酸分子：

[0097] a) 包括 SEQ ID NO:5 中阐释的核苷酸序列的核酸分子，或其互补物；和

[0098] b) 包括与 SEQ ID NO:5 中阐释的核苷酸序列具有至少 95％序列同一性的核苷酸序列的核酸分子，或其互补物。

[0099] 在某些实施方式中，圆果种黄麻的植物是品种 CVL-1。

[0100] 本发明的另一方面涉及分离的 WUSCHEL 相关的同源框 4 多肽，其包括 SEQ ID:NO:3 中阐释的氨基酸序列，或其生物活性片段，所述多肽包括选自催化长蒴黄麻的植物中启动、形成、增强和改变从而修饰韧皮部纤维组成的功能。

[0101] 在某些实施方式中，长蒴黄麻的植物是品种 O-4。

[0102] 在某些实施方式中，所述多肽包括与 SEQ ID NO:3 中阐释的氨基酸序列至少 95％的序列同一性。

[0103] 本发明的另一方面涉及分离的 WUSCHEL 相关的同源框 4 多肽，其包括 SEQ ID NO:6 中阐释的氨基酸序列，或其生物活性片段，所述多肽包括选自催化启动、形成、增强和改变从而修圆果种黄麻的植物中饰韧皮部纤维组成的一个或多个功能。

[0104] 在某些实施方式中，圆果种黄麻的植物是品种 CVL-1。

[0105] 在某些实施方式中，所述多肽包括与 SEQ ID NO:6 中阐释的氨基酸序列至少 95％的序列同一性。

[0106] 本发明的另一方面涉及重组基因构建体，其包括选自下述的多核苷酸：

[0107] a) 包括 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列的核酸分子，或其互补物；和

[0108] b) 包括与 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列具有至少 95％序列同一性的核苷酸序列的核酸分子，或其互补物，其中多核苷酸在宿主细胞中是可表达的以在长蒴黄麻和圆果种黄麻的植物中产生 WUSCHEL 相关的同源框 4 多肽的同系物。

[0109] 在某些实施方式中，所述构建体进一步包括启动子区域，其可操作地连接至

[0110] a) 包括 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列的核酸分子，或其互补物；或

[0111] b) 包括与 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列具有至少 95％序列同一性的核苷酸序列的核酸分子，或其互补物，其中所述启动子增强核酸分子的转录或表达。

[0112] 本发明的另一方面涉及包括重组基因构建体的转化株，其能够表达选自下述的多核苷酸：

[0113] a) 包括 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列的核酸分子，或其互补物；和

[0114] b) 包括与 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列具有至少 95％序列同一性的核苷酸序列的核酸分子，或其互补物，其中所述转化株产生 WUSCHEL 相关的同源框 4 多肽的同系物。

[0115] 本发明的另一方面涉及产生相对于对照植物具有增加或改善的纤维产率的植物或转基因植物的方法，其包括：

[0116] a) 在植物细胞中引入包括选自下述多核苷酸的重组基因构建体：i) 包括 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列的核酸分子，或其互补物；和 ii) 包括与 SEQ ID NO:2 或 SEQ ID NO:5 中阐释的核苷酸序列具有至少 95％序列同一性的核苷酸序列的核酸分子，或其互补物，

[0117] b) 在促进植物生长和发育的条件下培养植物细胞；和

[0118] c) 表达选自下述的多肽：i) 包括 SEQ ID:NO:3 或 SEQ ID NO:6 中阐释的氨基酸序列的多肽，或其生物活性片段；和 ii) 与 SEQ ID NO:3 或 SEQ ID NO:6 中阐释的氨基酸序列具有至少 95％序列同一性的多肽。

[0119] 本发明提供的序列可也用作合理修饰或设计新酶的预备材料，所述新酶具有确保酶在要求过程中表现更好的特征。

[0120] 本公开包括如所附的权利要求书中包含的，以及前述说明书中包含的。尽管已经以其一定程度特殊的优选形式描述了本发明，但是应理解，优选形式的公开仅仅作为例子并且可采用部件的构造和组合以及布置细节的许多改变，而不背离本发明和权利要求的范围。

[0121] 本文引用了许多参考文献，其内容通过参考以它们的整体并入。

实施例

[0122] 下述实施例旨在进一步阐释本发明，而决不将本发明限于其中描述的具体的实施方式。

[0123] 实施例 1引物的设计和合成

[0124] 在研究中使用的引物是手动精选的转录组设计并且从长蒴黄麻和圆果种黄麻的基因组序列预测的“基因模型”，通过手动选择具有完全 ORF 的序列或使用类似的基因已经成功从其他植物分离的数据库。使用 NCBI BLAST、BLASTP、RPS-BLAST、BLASTX 和 PSI-BLAST 进行获得自转录组的核苷酸序列的比较生物信息学分析，以鉴定相关基因的同系物并且用于基因的适当鉴定。当多个序列出现在“基因库”中时，通过 clustalW 版本 1.82 进行核苷酸序列比对。然后编辑比对。手动选择基因特异性引物（正向和反向）或通过 Primer 3plus 工具并且定制合成引物。

[0125] 合成在该研究中使用的所有寡核苷酸并且通过供应商 HPLC 纯化以及从整合 DNA 技术 (IDT) 获得。在高压灭菌 ddH2O 中制备约 100pmol 的原料液并且储存在约 -20℃，以等分试样待用。

[0126] 用作 PCR 引物的寡核苷酸序列

[0127]

引物名	SEQ ID NO	寡核苷酸序列	来自 cDNA 的扩增产品
COL F	1	CCATGGGAAACATGAAGGTGC	682
COL R	1	TGAAACGTCCATCATCTGCCT	
CCA F	4	CCATGGGAAACATGAAGGTGC	675
CCA R	4	TTCATGATCTGCCTTCCGGG	

[0128] 实施例 2来自长蒴黄麻和圆果种黄麻的 WUSCHEL 相关的同源框 4 的扩增、克隆和测序

[0129] 从 MS 培养基上生长的三天龄秧苗分离总 RNA，如先前 Chomczynski P 和 Sacchi N 描述，通过酸胍盐硫氰酸盐 - 苯酚 - 氯仿提取的 RNA 分离的单步方法。(Anal Biochem 1987，162 :156-159)。通过琼脂糖凝胶电泳检查 RNA 的质量或完整性并且使用 Thermo Scientific Nano Drop 2000 根据标准程序来量化。使用 SuperScript III 逆转录酶 (Invitrogen) 按照制造商的指导合成 cDNA 第一链。通过 PCR 使用基因特异性引物从 cDNA 扩增基因。PCR 反应 (50 μ L) 包含 1 μ L 的 cDNA，20pmoles 的每个引物，5 μ L 的 10X PCR 缓冲液，5 μ L 的 2.5mMdNTP 混合物和 1.0 单位的 PfuTaq DNA 聚合酶。在 Thermal Cycler(Applied Biosystems) 中使用下述条件进行 PCR :在约 95℃下初始变性约 5min 随后 35 个下述循环 :在约 95℃下变性约 30sec、在约 59-61℃下退火约 30sec 和在约 72℃下延伸约 1min，在约 72℃下最终延伸约 7min。通过 1％琼脂糖凝胶使用 1X TAE 缓冲液分析 PCR 产品并且从凝胶使用 QIAGEN 凝胶提取试剂盒根据制造商的指导洗脱扩增子。纯化的 PCR 产品连接至 pCR®8/GW/TOPO® TA 克隆试剂盒 (Invitrogen) 和转化至感受态大肠杆菌细胞 (Invitrogen)。从推定克隆使用 QIAprip Spin Miniprep 试剂盒 (QIAGEN) 按照制造商指导分离质粒。通过使用基因特异性引物检查插入物的存在并且对阳性质粒进行测序。

[0130] 实施例 3序列分析

[0131] 分别通过 BLASTN 和 BLASTP 程序分析核苷酸序列和氨基酸序列。从其他植物报道的序列用 ClustalW 比对。使用 Neighbour Joining(NJ) 进行系统发生分析。

[0132] 实施例 4纤维生物合成的途径构建

[0133] 显示 WUSCHEL 相关的同源框 4(WOX4) 在纤维生物合成中作用的自动代谢途径重建通过鉴定来自长蒴黄麻和圆果种黄麻的直向同源基因构建的蛋白质与 Arbidopsis 基因组比较。使用酶反应构建长蒴黄麻和圆果种黄麻基因组中编码的 WUSCHEL 相关的同源框 4(WOX4) 催化酶反应，所述酶反应在用于途径工作室的 Resnet-Plant 3.0 数据库以及来自代谢途径数据库可用。

[0134] 通过参考并入

[0135] 本文引用的所有的美国专利、美国公开专利申请，和公开的 PCT 申请通过参考并入本文。

[0136] 等同原则

[0137] 尽管本文已经描述和阐释了本发明的数个实施方式，本领域技术人员容易想到发

挥功能和/或获得本文所述的结果和/或一种或多种优势的各种其他方式和/或结构，并且每个这样的变型和/或修饰视为在本发明的范围内。本领域技术人员不使用过多的常规实验认识到，或能够确定本文所述的本发明具体实施方式的许多等价方式。所以，应理解前述实施方式仅仅作为例子呈现，并且在所附的权利要求和其等价方式的范围内；本发明可以如具体描述和要求的其他方式实践。

[0138] SEQ ID NO:1

[0139] 长度:1250bp

[0140] 类型:DNA

[0141] 特征名/关键字:内含子，外显子——包括150bp 5'UTR和150bp3'UTR

[0142] 生物体:黄麻，长蒴黄麻

[0143] ACCTCATCATTTGGTCTTGGTTCTTCAAAGCACCACCTCATATCCCACTTTCTTTTTCTTTTCTTTTTTTTAAAATTTTTATTTTTTGGGTTAAATTTATTTTTGTTATTATATAATCAGAGCATTCTCCCCATCATTCACTCCTTCACCATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAACCCTCCCTCTCCCTTGGTTGCAAGCGCTTACGCCCTCTTGCTCCCAAGCTCCACCCTTCCTCTTCCCCTGATCATACTTCCGTCTCCTCTTTCGACCTTAAGACCTTCATTCGTCCCGAAAGTGGCCCCCGAAAACTTTGCCCTTCTGACGACAAGCGAGATTCTCATTCTCCCCAGGTACTTACAAATATCCATGAAATGAAACCCTTTTATTTTTCCAATGATGTATATAATTAGTGGGGAAGGATAAGATTTTGAACTTACGATGAACAATAACAGTGTGGAAAGTTGATAATCAAAGGTTAAGAATGTGAAGCTGAAACTTTTTTCTTTTTTTTGGTTATGTAGGTGGAAACGCACCCAGGGGGAACGCGGTGGAATCCGACGCAAGAGCAGATAGGGATATTGGAGATGCTGTATAGAGGTGGGATGCGAACTCCAAATGCACAGCAAATAGAACAGATCACTGCACAGTTAGGCAAGTACGGGAAGATCGAAGGCAAAAACGTTTTCTATTGGTTCCAAAACCACAAAGCACGCGAAAGGCAAAAGCAGAAGCGTAACAGTCTTGGTCTTAGCCATTCTCCCAGAAACTCTGCTCCCATTACCACCATAACTTTGGACTCTAGGGTAAGTTCAAACCAACAAAACCCTTTCTTTGTATATATATAACGGTTAGTTTTTAGTTTTTACTTCTTATAAACAGCAATTAACATTAATGTTTTTGTTTATATATATAGGGGGAAGTAATGGAGAGAGAGGAGGATAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCATTTGAGTACTTAGAAGAAGAAAGCAGATCATCATCGTCGAGTCAAGAGGAGGAAAACAGAACTCTGGAGCTTTTCCCATTGCACCCGGAAGGCAGATGATGGACGTTTCAACTTTGAAAAACAAGGAAAAAGGGAAGCTTAACCCAAAACCAAAAAGACTGCTACAAAACCCAAAACTCTGTTCCCATTTATGAAATGATAAACATATGCTTTGATGATCCATGATGATGATGATGATAATGAAGCTGA

[0144] SEQ ID NO:2

[0145] 长度:669bp

[0146] 类型:DNA

[0147] 生物体:黄麻，长蒴黄麻

[0148] 特征名/关键字:CDS

[0149] 位置:(1)…..(669)

[0150] ATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAACCCTCCCTCTCCCTTGGTTGCAAGCGCTTACGCCCTCTTGCTCCCAAGCTCCACCCTTCCTCTTCCCCTGATCATACTTCCGTCTCCTCTTTCGACCTTAAGACCTTCATTCGTCCCGAAAGTGGCCCCCGAAAACTTTGCCCTTCTGACGACAAGCGAGATTCTCATTCTCCCCAGGTGGAAACGCACCCAGGGGGAACGCGGTGGAATCCGACGCAAGAGCAGATAGGGATATTGGAGATGCTGTATAGAGGTGGGATGCGAACTCCAAATGCACAGCAAATAGAACAGATCACTGCACAGTTAGGCAAGTACGGGAAGATCGAAGGCAAAAACGTTTTCTATTGGTTCCAAAACCACAAAGCACGCGAAAGGCAAAAGCAGAAGCGTAACAGTCTTGGTC

TTAGCCATTCTCCCAGAAACTCTGCTCCCATTACCACCATAACTTTGGACTCTAGGGGGGAAGTAATGGAGAGAGAG
GAGGATAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCATTTGAGTACTTAGAAGAAGAAAGCAGATCATCATCGTC
GAGTCAAGAGGAGGAAAACAGAACTCTGGAGCTTTTCCCATTGCACCCGGAAGGCAGATGA

[0151] SEQ ID NO:3

[0152] 长度:222

[0153] 类型:PRT

[0154] 生物体:黄麻,长蒴黄麻

[0155] MGNMKVHQLARGLWEHEPSLSLGCKRLRPLAPKLHPSSSPDHTSVSSFDLKTFIRPESGPRKLCPSDDK
RDSHSPQVETHPGGTRWNPTQEQIGILEMLYRGGMRTPNAQQIEQITAQLGKYGKIEGKNVFYWFQNHKARERQKQK
RNSLGLSHSPRNSAPITTITLDSRGEVMEREEDSPYKRKCRSWSFEYLEEESRSSSSSQEEENRTLELFPLHPEGR*

[0156] SEQ ID NO:4

[0157] 长度:1237bp

[0158] 类型:DNA

[0159] 特征名/关键字:内含子,外显子——包括150bp 5'UTR和150bp3'UTR

[0160] 生物体:黄麻,圆果种黄麻

[0161] TCACAAGTCAACCTCACCTCATCATTTGGTCTTGGTTCTTCAAAGCACCACCTCATATCCCACTTCCTT
TTTCAATTTTTAATTTTTTTTGGGGTTAAATTTATTTGGTTATATAATCAGAGCATTCTCCCCATCATTCACTCCTT
CACCATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAACCCTCCCTCTCCCTTGGTTGCA
AGCGCTTACGCCCTCTTGCTCCCAAGCTCCACCCTTCCTCTTCCCCTGATCATACTTCCGTCTCCTCTTTCGACCTT
AAGACCTTCATTCGTCCCGAAAGTGGCCCCCGGAAACTTTGCCCTTCTGACGACAAGCGAGATTCTCATTCTCGCCA
GGTACTTAAAAATTAATATCCATGAAATAATTTGTGGGGAAGGATAAGTTTTGAACTTAATGCATAATAACAGTGTG
GAAACTTAATAGGTTAAGAATATTTGAAGAACTTCTATATATATGAAGCTGAAACTTTTTTGTGGTGTTGTAGGTGG
AAACGCACCCAGGGGGAACGCGGTGGAATCCGACGCAAGAGCAGATAGGGATACTGGAGATGCTGTATAGAGGTGGG
ATGCGAACTCCAAATGCACAGCAAATAGAACAGATCACTGCACAGCTAGGCAAGTACGGCAAGATCGAAGGCAAAAA
CGTTTTCTATTGGTTCCAAAACCACAAAGCACGCGAAAGGCAAAAGCAGAAGCGTAACAGTCTTGGTCTTAGCCATT
CTCCCAGAAACTCAGCTCCCATTACCACTATAACTTTGGACACTAGGGTAAGTTCAAACCAACAAAACCCTTTCTGT
GTATATATATATATAACGGTTAGTTTTTAGTTTTTACTTCTTATAAACAGAGAAAATTAACAATGTTTGTGTTTATA
TATATATAGGGGGAAGTAATGGAAAGAGAGGAGGATAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCTTTTGAGTA
CTTAGAAGAAGAAAGCAGATCATCATCGTCGAGTCAAGAGGAGGAAAACAGAACTCTGGAGCTTTTCCCATTGCACC
CGGAAGGCAGATCATGAAGGGGGTTTCAACTTTCAACTTTCAACTTTCAACTTTCAAAATGAAGGGAAAAGGGAAGC
TTAACCCAAAACCAAAAAGACTGCTACAAAACCCAAAACTCTGTTCCCATTTATGAAATGATAAACTTATGCTTTGA
TGATCGATCCATG

[0162] SEQ ID NO:5

[0163] 长度:672

[0164] 类型:DNA

[0165] 生物体:黄麻,圆果种黄麻

[0166] 特征名/关键字:CDS

[0167] 位置:(1)…..(672)

[0168] ATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAACCCTCCCTCTCCCTTGGT

TGCAAGCGCTTACGCCCTCTTGCTCCCAAGCTCCACCCTTCCTCTTCCCCTGATCATACTTCCGTCTCCTCTTTCGA
CCTTAAGACCTTCATTCGTCCCGAAAGTGGCCCCCGGAAACTTTGCCCTTCTGACGACAAGCGAGATTCTCATTCTC
GCCAGGTGGAAACGCACCCAGGGGGAACGCGGTGGAATCCGACGCAAGAGCAGATAGGGATACTGGAGATGCTGTAT
AGAGGTGGGATGCGAACTCCAAATGCACAGCAAATAGAACAGATCACTGCACAGCTAGGCAAGTACGGCAAGATCGA
AGGCAAAAACGTTTTCTATTGGTTCCAAAACCACAAAGCACGCGAAAGGCAAAAGCAGAAGCGTAACAGTCTTGGTC
TTAGCCATTCTCCCAGAAACTCAGCTCCCATTACCACTATAACTTTGGACACTAGGGGGGAAGTAATGGAAAGAGAG
GAGGATAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCTTTTGAGTACTTAGAAGAAGAAAGCAGATCATCATCGTC
GAGTCAAGAGGAGGAAAACAGAACTCTGGAGCTTTTCCCATTGCACCCGGAAGGCAGATCATGA

[0169] SEQ ID NO:6

[0170] 长度:223

[0171] 类型:PRT

[0172] 生物体:黄麻,圆果种黄麻

[0173] MGNMKVHQLARGLWEHEPSLSLGCKRLRPLAPKLHPSSSPDHTSVSSFDLKTFIRPESGPRKLCPSDDK
RDSHSRQVETHPGGTRWNPTQEQIGILEMLYRGGMRTPNAQQIEQITAQLGKYGKIEGKNVFYWFQNHKARERQKQK
RNSLGLSHSPRNSAPITTITLDTRGEVMEREEDSPYKRKCRSWSFEYLEEESRSSSSSQEEENRTLELFPLHPEG
RS*

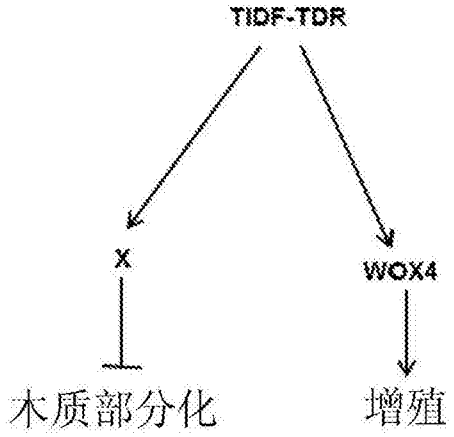
TIDF-TDR
X
WOX4
木质部分化
增殖

图 1

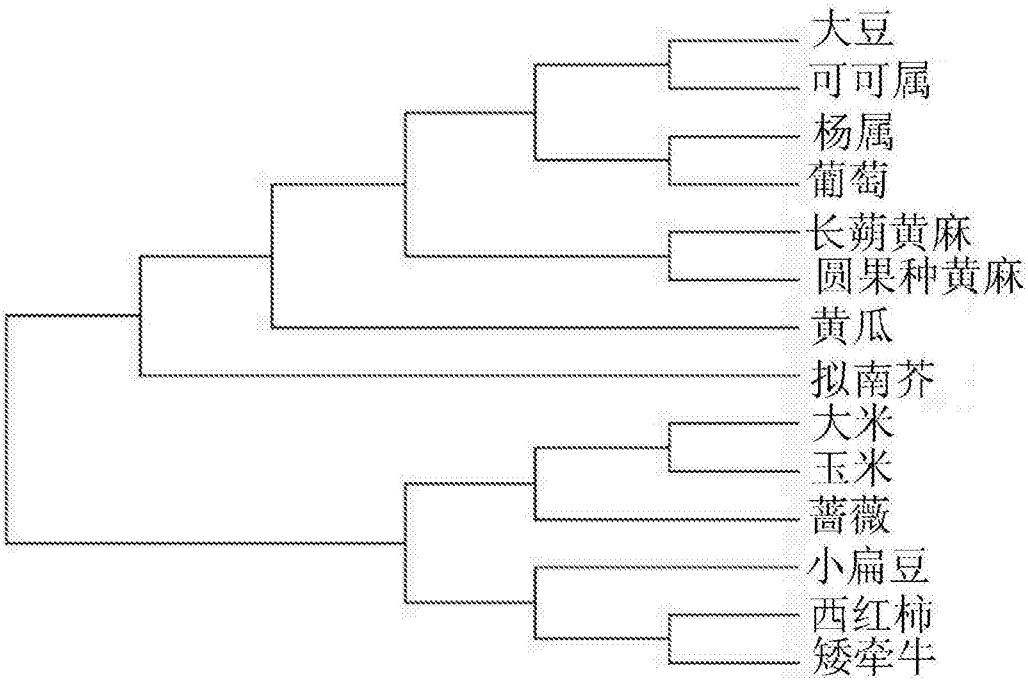
大豆
可可属
杨属
葡萄
长蒴黄麻
圆果种黄麻
黄瓜
拟南芥
大米
玉米
蔷薇
小扁豆
西红柿
矮牵牛

图 2