

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. AU 2014352999 B2

(54) Title
Nucleotide sequence encoding WUSCHEL-related homeobox4 (WOX4) protein from Corchorus olitorius and Corchorus capsularis and methods of use for same

(51) International Patent Classification(s)
A01H 5/00 (2006.01) **C07K 14/415** (2006.01)
A01H 5/10 (2006.01) **C12N 15/82** (2006.01)

(21) Application No: **2014352999** (22) Date of Filing: **2014.11.20**

(87) WIPO No: **WO15/077447**

(30) Priority Data

(31) Number **61/907,617** (32) Date **2013.11.22** (33) Country **US**

(43) Publication Date: **2015.05.28**
(44) Accepted Journal Date: **2020.12.24**

(71) Applicant(s)
Bangladesh Jute Research Institute

(72) Inventor(s)
Islam, Mohammed Shahidul;Ahmed, Borhan;Haque, Mohammed Samiul;Alam, Mohammed Monjurul;Alam, Maqsudul

(74) Agent / Attorney
Shelston IP Pty Ltd., L 9 60 Margaret St, Sydney, NSW, 2000, AU

(56) Related Art
Yoshihiro Ohmori, "WUSCHEL-RELATED HOMEOBOX4 Is Involved in Meristem Maintenance and Is Negatively Regulated by the CLE Gene FCP1 in Rice", The Plant Cell Vol. 25: 229–241, January 2013
US 20110004958 A1

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

(19) World Intellectual Property
Organization
International Bureau



(10) International Publication Number

WO 2015/077447 A2

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

(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 INSTITUTE [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

5 **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 10 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 15 well as a transgenic jute plant thereof.

Background of the invention

Any discussion of the prior art throughout the specification should in no way be considered as an admission that such prior art is widely known or forms part of common general knowledge in the field.

20 It is an object of the present invention to overcome or ameliorate at least one of the disadvantages of the prior art, or to provide a useful alternative.

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 25 *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 30 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, 5 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, 10 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 15 (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.

20 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 25 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 30 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

5 According to a first aspect, the present invention relates to an expression vector comprising a polynucleotide having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5.

According to a second aspect, the present invention relates to an isolated polypeptide having at least 95% sequence identity to an amino acid sequence set forth in SEQ ID:

10 NO: 3 or SEQ ID NO: 6.

According to a third aspect, the present invention relates to a transformant comprising an expression vector of the invention, or a polynucleotide having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5,

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

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

a) transfected at least one plant cell with the expression vector of the invention,
20 and

b) cultivating said at least one plant cell under conditions for promoting plant growth and development.

According to a fifth aspect, the present invention relates to a transgenic plant produced by the method according to the invention.

25 According to a sixth aspect, the present invention relates to a method of inducing, initiating, improving, or enhancing plant growth, plant height, fiber, and seed yield in a jute plant, comprising incorporating in to a jute plant the expression vector of the invention.

Unless the context clearly requires otherwise, throughout the description and the claims, the words “comprise”, “comprising”, and the like are to be construed in an inclusive sense as opposed to an exclusive or exhaustive sense; that is to say, in the sense of “including, but not limited to”.

5 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
10 thereof.

Another aspect 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.

15 Still another aspect 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.

20 Yet another aspect 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.

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

25 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
30 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
5 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
10 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.*

5 *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.

15 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

20 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.

25 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

30 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

5 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

10 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
15 "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.

20 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
25 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
30 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

5 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 10 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.

15 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 20 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

25 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 30 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.

25 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.

30 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.

5 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.

10 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 15 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 20 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 25 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 30 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 5 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 10 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.

15 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 comprising extra amino acid 20 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, 25 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 30 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 5 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).

10 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		

15 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,

5 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, 10 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, 15 211, 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%, 20 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

25 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 30 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

5 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 10 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

20 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 25 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 30 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 5 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 10 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 15 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 20 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 25 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.

30 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 5 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 10 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.

15 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 20 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 25 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;

30 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

5 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.

10 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.

15

20

25

30

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

5 *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.

10 15 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.

20 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.

25 30 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
5 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.
10 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
15 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.

20 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
25 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 translatable to
30 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 5 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 10 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 15 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 20 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 25 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% 30 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.

10 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*.

15 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 20 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.

25 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

- 5 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
- 10 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
- 15 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,
- 25 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 5 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.

10 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

5 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 10 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 15 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 ddH₂O 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	CCATGGAAACATGAAGGTGC	682
COL R	1	TGAAACGTCCATCATCTGCCT	
CCA F	4	CCATGGAAACATGAAGGTGC	675
CCA R	4	TTCATGATCTGCCTTCCGGG	

20

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 25 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 5 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 10 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 15 transformed into competent *E. coli* cells (Invitrogen). Plasmids were isolated from putative colonies using QIAprep 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

20 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 25 homeobox4 (WOX4) in fiber biosynthesis was constructed by identifying orthologs from *C. olitorus* and *C. capsularis* protein compared with *Arabidopsis* 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.

30 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
5 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
10 practised otherwise than as specifically described and claimed.

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

ACCTCATCATTGGTCTTGGTCTTCAAAGCACCACCTCATATCCCACCTTCTTTCTTTCTTT
TTTAAATTTATTTGGGTTAAATTATTTGTTATTATATAATCAGAGCATTCTCCCCAT
10 CATTCACTCCTCACCATGGAAACATGAAGGTGCATCAGTGGCACGTGGCTATGGGAGCATGAA
CCCTCCCTCTCCCTGGTGCAAGCGCTTACGCCCTCTGCTCCAAGCTCCACCCTCCTCTCCC
CTGATCATACTCCGTCTCCTTTGACCTTAAGACCTTCATTGTCCCAGTGGCCCCGAAA
ACTTGCCCTCTGACGACAAGCGAGATTCTCATTCTCCCCAGGTACTTACAAATATCCATGAAATG
AAACCCCTTTATTTCCAATGATGTATATAATTAGTGGGAAGGATAAGATTGAACTTACGATG
15 ACAATAACAGTGTGAAAGTTGATAATCAAAGTTAAGAATGTGAAGCTGAAACTTTTCTTTT
TTGGTTATGTAGGTGAAACGCACCCAGGGGAACCGGGTGGATCCGACGCAAGAGCAGATAGGG
ATATTGGAGATGCTGTATAGAGGTGGGATGCGAACTCCAAATGCACAGCAAATAGAACAGATCACTG
CACAGTTAGGCAAGTACGGAAGATCGAAGGCAAAACGTTCTATTGGTCCAAAACCACAAAGC
ACCGCAAAGCAAAAGCAGAACGCTAACAGTCTGGTCTAGCCATTCTCCCAGAAACTCTGCTCCC
20 ATTACCAACATAACTTGGACTCTAGGTAAGTTCAAACCAACAAACCCCTTCTTGTATATAT
AACGGTTAGTTTACTTCTTATAAACAGCAATTAACTTAAATGTTTGTATATATA
TAGGGGGAAAGTAATGGAGAGAGAGGGAGGATAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCATTG
AGTACTTAGAAGAAGAAAGCAGATCATCGTCAGTCAAGAGGAGGAAACAGAACTCTGGAGCT
TTCCCATTGCACCCGAAGGCAGATGATGGACGTTCAACTTGAAAAACAAGGAAAAGGAAGC
25 TTAACCCAAAACCAAAAGACTGCTACAAAACCCAAACTCTGGTCCATTATGAAATGATAAACAC
TATGCTTGATGATCCATGATGATGATGATAATGAAGCTGA

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

ATGGGAAACATGAAGGTGCATCAGTTGGCACGTGGCTTATGGGAGCATGAACC
 CTCCCTCTCCCTGGTTGCAAGCGCTTACGCCCTTGTGCTCCCAAGCTCCACCC
 10 TCCTCTTCCCCTGATCATACTTCCGTCTCCTCTTCGACCTTAAGACCTTCATTG
 TCCCAGGAAAGTGGCCCCCGAAAACCTTGCCCTCTGACGACAAGCGAGATTCTCA
 TTCTCCCCAGGTGGAAACGCACCCAGGGGAACGCGGTGGAATCCGACGCAAG
 AGCAGATAGGGATATTGGAGATGCTGTATAGAGGTGGGATGCGAACTCCAAAT
 GCACAGCAAATAGAACAGATCACTGCACAGTTAGGCAAGTACGGGAAGATCGA
 15 AGGCAAAAACGTTTCTATTGGTTCCAAAACCACAAAGCACCGCAAAGGCAA
 AGCAGAACGCTAACAGTCTTGGTCTAGCCATTCTCCCAGAAACTCTGCTCCCA
 TTACCACCATAACTTGACTCTAGGGGGAAAGTAATGGAGAGAGAGGAGGAGGAT
 AGTCCATATAAGAGAAAGTGTAGGAGCTGGTCATTGAGTACTTAGAAGAAGA
 AAGCAGATCATCGTCGAGTCAAGAGGAGGAAAACAGAAACTCTGGAGCTT
 20 TCCCATTGCACCCGGAAGGCAGATGA

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

MGNMKVHQLARGLWEHEPSLSLGCKRLRPLAPKLHPSSPDHTSVSSFDLKFIRPESGPRKLCPSD
 DKRDSHSPQVETHPGGTRWNPTQEIQIGILEMLYRGGMRTPNAQQIEQITAQLGKYGKIEGKNVFYWF
 30 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
 5 ORGANISM : Jute, *C. capsularis*

TCACAAGTCAACCTCACCTCATTTGGTCTGGTCTTCAAAGCACCACCTCATATCCCACTTCC
 TTTTCAATTAAATTTGGGTTAAATTATGGTTATATAATCAGAGCATTCTCCCCAT
 CATTCACTCCTCACCATGGAAACATGAAGGTGCATCAGTGGCACGTGGCTTATGGGAGCATGAA
 10 CCCTCCCTCTCCCTGGTTGCAAGCGTTACGCCCTTGCTCCAAGCTCCACCCTCCTCTTCCC
 CTGATCATACTTCCGTCTCCTTTGACCTTAAGACCTTCATTGTCCGAAAGTGGCCCCGGAA
 ACTTGCCTCTGACGACAAGCGAGATTCTCATTCTGCCAGGTACTTAAAAATTAAATATCCATGA
 AATAATTGCCCCAAGGATAAGTTGAACCTTAATGCATAATAACAGTGTGAAACTTAATAGGTT
 AAGAATATTGAAGAACCTCTATATATGAAGCTGAAACTTTGTGGTGTAGGTGAAACG
 15 CACCCAGGGGAACGCCGTGGAATCCGACGCAAGAGCAGATAGGGATACTGGAGATGCTGTATAGAG
 GTGGGATGCGAACTCAAATGCACAGCAAATAGAACAGATCACTGCACAGCTAGGCAAGTACGCCA
 GATCGAAGGCAAAACGTTCTATTGGTCAAAACCACAAAGCACGCCAGGCAAAAGCAGAAG
 CGTAACAGTCTGGTCTAGCCATTCTCCAGAAACTCAGCTCCATTACCACTATAACTTGGACA
 CTAGGGTAAGTTCAAACCAACAAACCCCTTCTGTGTATATATATAACGGTTAGTTTAGTTAGTT
 20 TTTACTTCTTATAAACAGAGAAAATTAAACAATGTTGTGTTATATATATAGGGGAAAGTAATGG
 AAAGAGAGGAGGATAGTCCATATAAGAGAAAGTGTAGGAGCTGGCTTTGAGTACTTAGAAGAAGA
 AAGCAGATCATCGTCAGTCAAGAGGAGGAAACAGAACTCTGGAGCTTTCCATTGCACCCG
 GAAGGCAGATCATGAAGGGGTTCAACTTCAACTTCAACTTCAACTTCAAAATGAAGGAAA
 AGGAAAGCTTAACCCAAAACCAAAAGACTGCTACAAACCCAAAACTCTGTTCCCATTATGAAAT
 25 GATAAACTTATGCTTGATGATCGATCCATG

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

ATGGGAAACATGAAGGTGCATCAGTGGCACGTGGCTTATGGGAGCATGAACCCCTCCCTCCCTTG
 35 GTTGCAAGCGCTTACGCCCTTGCTCCAAGCTCCACCCTTCCCTGATCATACTTCCGT
 CTCCTCTTCGACCTTAAGACCTTCATTGTCCGAAAGTGGCCCCGGAAACTTGGCCCTTGAC

GACAAGCGAGATTCTCATTCTGCCAGGTGGAAACGCACCCAGGGGAACCGCGTGGATCCGACGC
AAGAGCAGATAGGGATACTGGAGATGCTGTATAGAGGTGGATGCGAACTCCAAATGCACAGCAAAT
AGAACAGATCACTGCACAGCTAGGCAAGTACGGCAAGATCGAAGGCACAAACGTTCTATTGGTTC
CAAAACCACAAAGCAGCGAAAGCAGAAGCGTAACAGTCTGGTCTTAGCCATTCTCCCA
5 GAAACTCAGCTCCCATTACCACTATAACTTGGACACTAGGGGGAAAGTAATGGAAAGAGAGGAGGA
TAGTCCATATAAGAGAAAGTGTAGGAGCTGGTCTTTGAGTACTTAGAAGAAGAAAGCAGATCATCA
TCGTCGAGTCAAGAGGAGGAAAACAGAACTCTGGAGCTTTCCATTGCACCCGGAAGGCAGATCAT
GA

10

SEQ ID NO : 6
LENGTH : 223
TYPE : PRT
ORGANISM : Jute, *C. capsularis*

5

MGNMKVHQLARGLWEHEPSLSLGCKRLRPLAPKLHPSSPDHTSVSSFDLKFIRPESGPRKLCPSD
DKRDSHSRQVETHPGGTRWNPTQEIQIGILEMLYRGGMRTPNAQQIEQITAQLGKYGKIEGKNVFYWF
QNHKARERQKQKRNSLGLSHSPRNSAPITTITLDTRGEVMEREEDSPYKRKCRSWSFEYLEEESRSS
SSSQEEENRTLELFPLHPEGRS*

10

What is claimed is:

1. An expression vector comprising a polynucleotide having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2, or SEQ ID NO: 5.
2. The expression vector of claim 1, wherein said polynucleotide has at least 98% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5.
3. The expression vector of claim 1, wherein said polynucleotide has at least 99% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5.
4. The expression vector of claim 1, wherein said polynucleotide has 100% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5.
5. An isolated polypeptide having at least 95% sequence identity to an amino acid sequence set forth in SEQ ID: NO: 3, or SEQ ID NO: 6.
6. The isolated polypeptide according to claim 5, wherein said polypeptide has at least 98% sequence identity to the amino acid sequence set forth in SEQ ID NO: 3 or SEQ ID NO: 6.
7. The isolated polypeptide according to claim 5, wherein said polypeptide has at least 99% sequence identity to the amino acid sequence set forth in SEQ ID NO: 3 or SEQ ID NO: 6.
8. The isolated polypeptide according to claim 5, wherein said polypeptide has 100% sequence identity to the amino acid sequence set forth in SEQ ID NO: 3 or SEQ ID NO: 6.
9. A transformant comprising an expression vector of any one of claims 1 to 4, or a polynucleotide having at least 95% sequence identity to the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 5,
wherein said transformant produces a homologue of WUSCHEL-related homeobox4 polypeptide.
10. A method for producing a transgenic plant having increased or enhanced fiber yield relative to control plants, comprising:

a) transfected at least one plant cell with the expression vector of any one of claims 1 to 4, and

b) cultivating said at least one plant cell under conditions for promoting plant growth and development.

11. A transgenic plant produced by the method according to claim 10.

12. A method of inducing, initiating, improving, or enhancing plant growth, plant height, fiber, and seed yield in a jute plant, comprising incorporating in to a jute plant the expression vector of any one of claims 1 to 4.

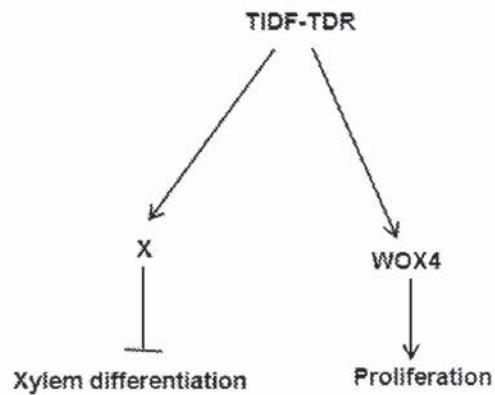


Figure 1

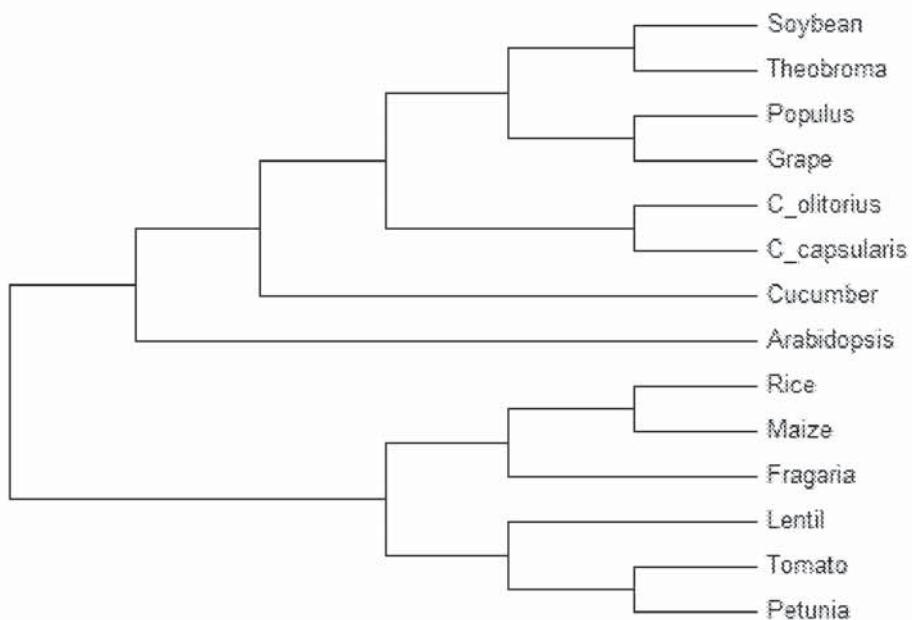


Figure 2