



(86) Date de dépôt PCT/PCT Filing Date: 2005/06/14  
(87) Date publication PCT/PCT Publication Date: 2005/12/22  
(85) Entrée phase nationale/National Entry: 2006/12/14  
(86) N° demande PCT/PCT Application No.: IL 2005/000627  
(87) N° publication PCT/PCT Publication No.: 2005/121364  
(30) Priorité/Priority: 2004/06/14 (US60/578,833)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01)

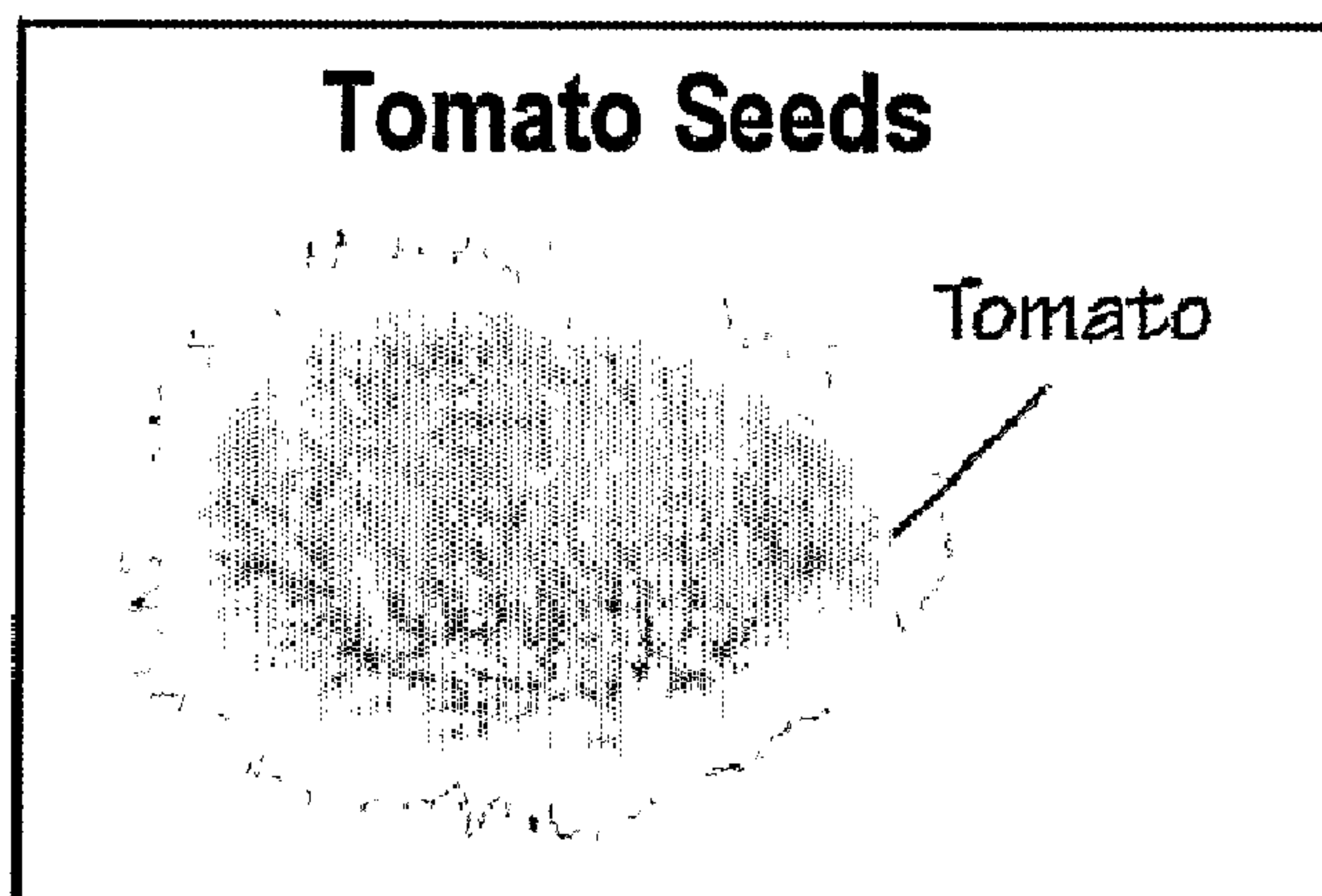
(71) Demandeur/Applicant:  
EVOGENE LTD., IL

(72) Inventeurs/Inventors:  
RONEN, GIL, IL;  
GOLD, EVGENIA, IL;  
YELIN, RODRIGO, IL;  
MEISSNER, RAFAEL, IL;  
KARCHI, HAGAI, IL;  
AYAL, SHARON, IL

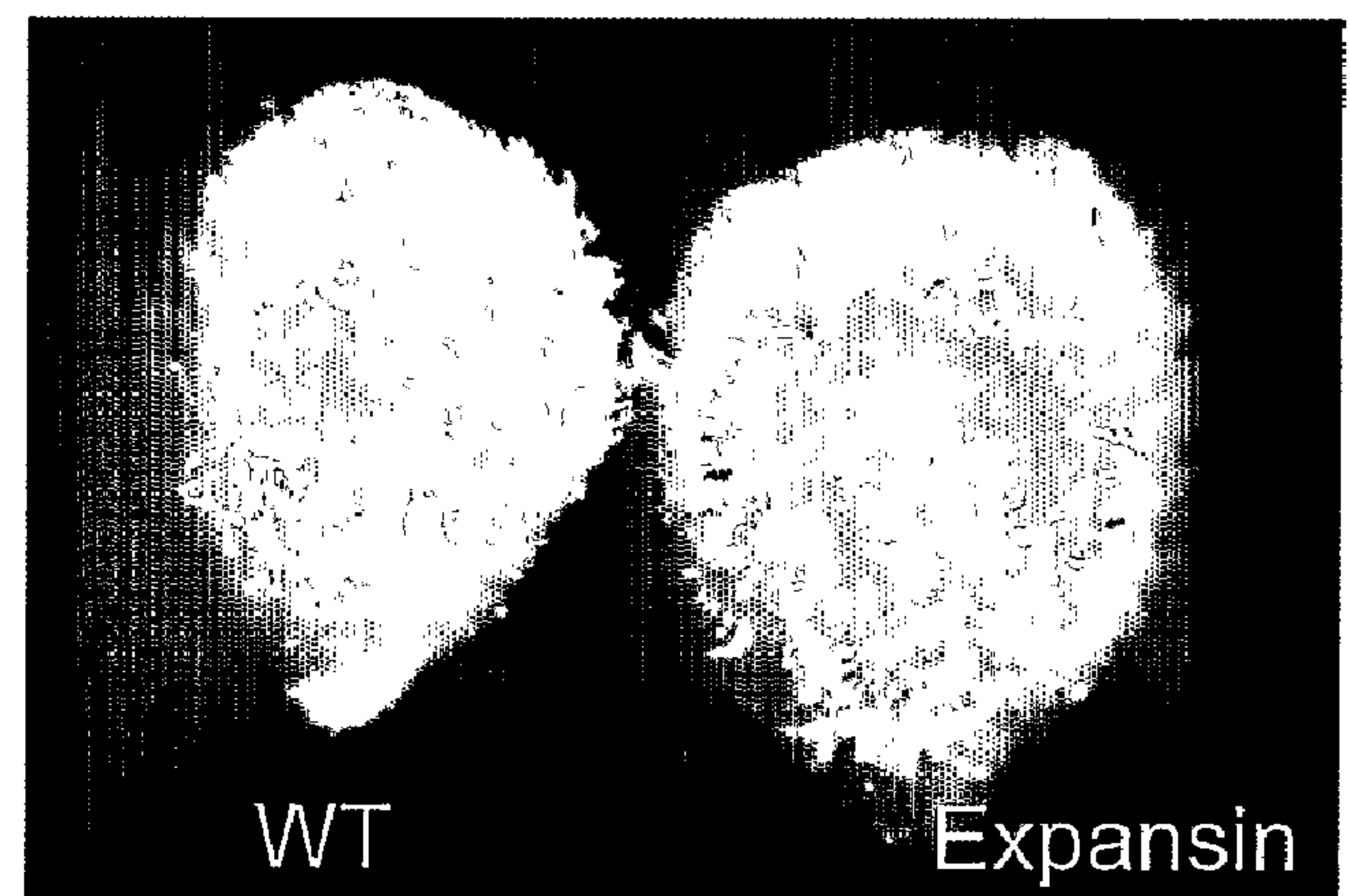
(74) Agent: OGILVY RENAULT LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : POLYNUCLEOTIDES ET POLYPEPTIDES IMPLIQUES DANS LE DEVELOPPEMENT DE LA FIBRE  
VEGETALE ET PROCEDES PERMETTANT DE LES UTILISER

(54) Title: POLYNUCLEOTIDES AND POLYPEPTIDES INVOLVED IN PLANT FIBER DEVELOPMENT AND METHODS  
OF USING SAME



**a**



**b**

(57) **Abrégé/Abstract:**

Isolated polynucleotides are provided. Each of the isolated polynucleotides comprise a nucleic acid sequence encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96, wherein the polypeptide is capable of regulating cotton fiber development. Also provided are methods of using such polynucleotides for improving fiber quality and/or yield of a fiber producing plant, as well as methods of using such polynucleotides for producing plants having increased biomass/vigor/yield.



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

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
22 December 2005 (22.12.2005)

PCT

(10) International Publication Number  
**WO 2005/121364 A2**

(51) International Patent Classification<sup>7</sup>: **C12Q 1/68**

(74) Agent: G.E. EHRLICH (1995) LTD.; 11 Menachem Begin Street, 52 521 Ramat Gan (IL).

(21) International Application Number:  
PCT/IL2005/000627

(22) International Filing Date: 14 June 2005 (14.06.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/578,833 14 June 2004 (14.06.2004) US

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

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **EVO-GENE LTD.** [IL/IL]; 13 Gad Finstein Street, 76121 Rehovot (IL).

(72) Inventors; and

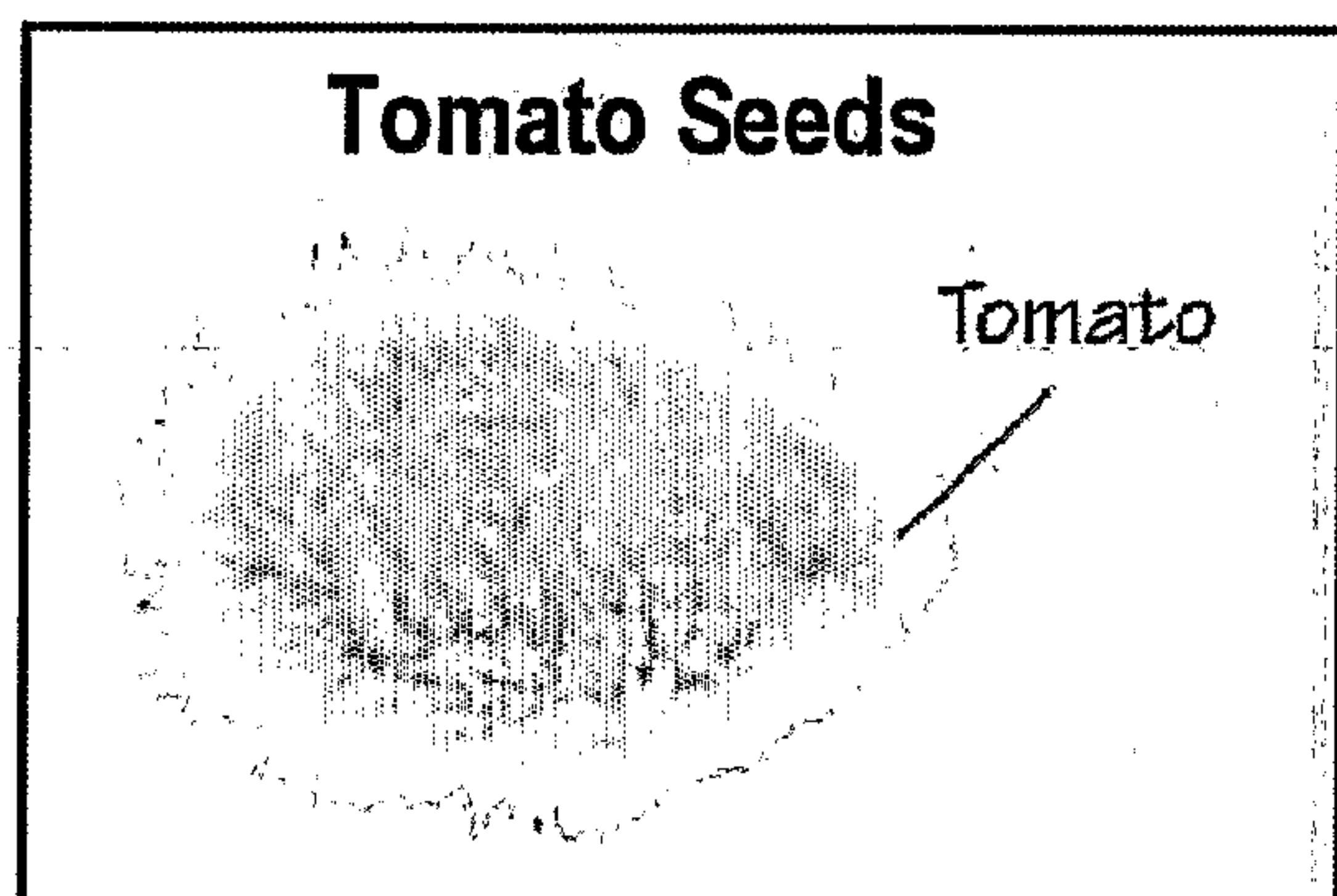
(75) Inventors/Applicants (*for US only*): **RONEN, Gil** [IL/IL]; 20 Moshav Ometz, 33870 Emeq-Hefer (IL). **GOLD, Evgenia** [IL/IL]; 11/3 Yavne Street, 76360 Rehovot (IL). **YELIN, Rodrigo** [IL/IL]; 12 Odem Street, 44862 Zur-Yigal (IL). **MEISSNER, Rafael** [IL/IL]; 11 HaLevona Street, 76350 Rehovot (IL). **KARCHI, Hagai** [IL/IL]; 249 Sapir Street, Moshav Sitriya, 76834 Doar-Na Emek Soreq (IL).

**Published:**

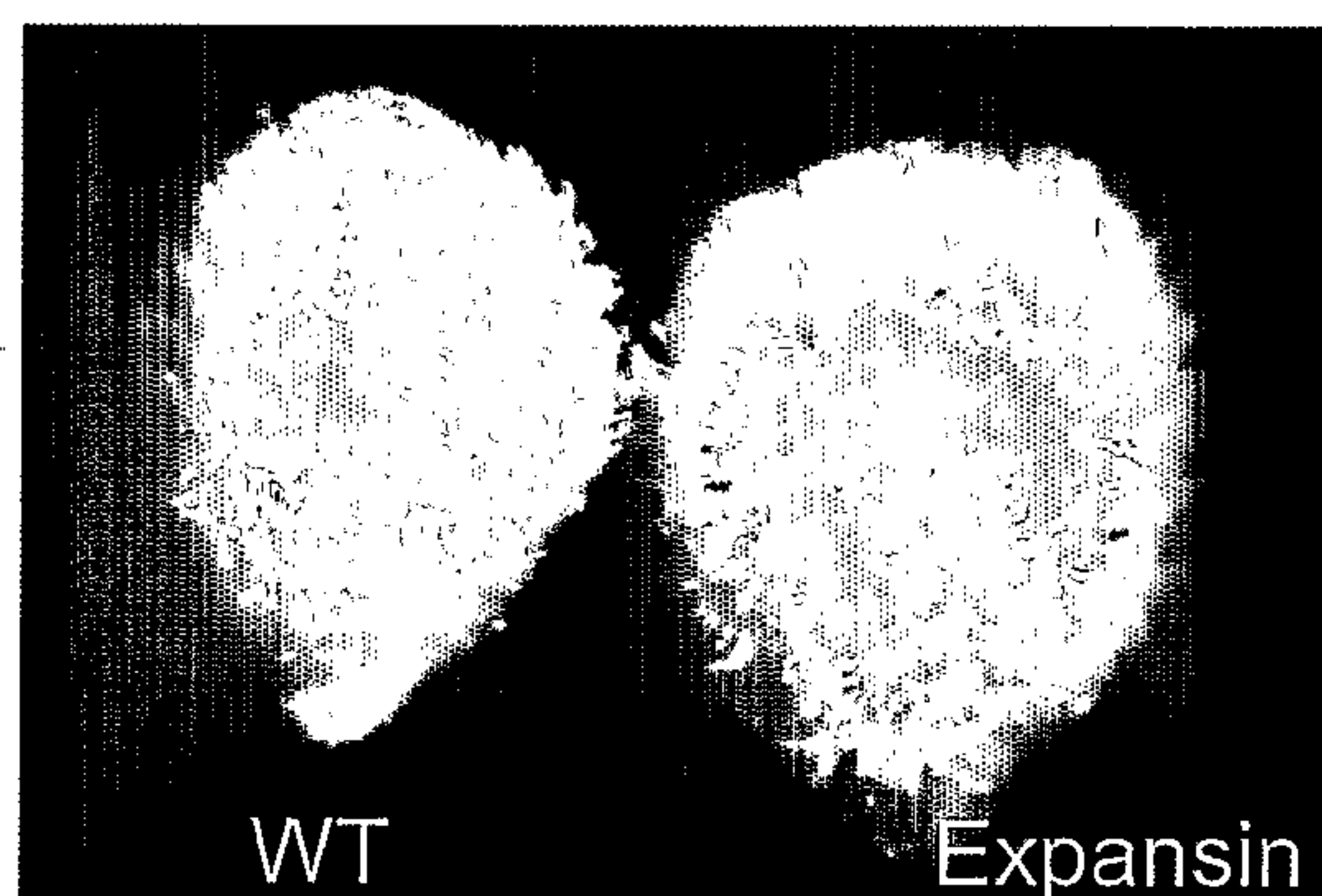
— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: POLYNUCLEOTIDES AND POLYPEPTIDES INVOLVED IN PLANT FIBER DEVELOPMENT AND METHODS OF USING SAME



a



b

(57) Abstract: Isolated polynucleotides are provided. Each of the isolated polynucleotides comprise a nucleic acid sequence encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96, wherein the polypeptide is capable of regulating cotton fiber development. Also provided are methods of using such polynucleotides for improving fiber quality and/or yield of a fiber producing plant, as well as methods of using such polynucleotides for producing plants having increased biomass/vigor/yield.



## DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME           1    DE    2  
CONTENANT LES PAGES    1    À    67

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

## JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME           1    OF    2  
CONTAINING PAGES    1    TO    67

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:



# POLYNUCLEOTIDES AND POLYPEPTIDES INVOLVED IN PLANT FIBER DEVELOPMENT AND METHODS OF USING SAME

5

## FIELD AND BACKGROUND OF THE INVENTION

The present invention relates to polynucleotides and polypeptides involved in plant-fiber development and methods of using same.

The present invention relates to a novel computational approach that utilizes comparative genomics to identify genes which play a role in fiber development.

Cotton and cotton by-products provide raw materials that are used to produce a wealth of consumer-based products in addition to textiles including cotton foodstuffs, livestock feed, fertilizer and paper. The production, marketing, consumption and trade of cotton-based products generate an excess of \$100 billion annually in the U.S. alone, making cotton the number one value-added crop.

It is estimated that the use of cotton as a fiber by humans dates back 7000 years in Central America and 5000 years in India. Even with the growth of synthetic fibers in the last 50 years, cotton still accounts for approximately 50 % of the world's textile fiber [Agrow Reports, Global Seed markets DS208, October 2000].

Even though 90 % of cotton's value as a crop resides in the fiber (lint), yield and fiber quality has declined, especially over the last decade [Meredith (2000), Proc. World Cotton Research Conference II, Athens, Greece pp.97-101]. This decline has been attributed to general erosion in genetic diversity of cotton varieties, and an increased vulnerability of the crop to environmental conditions [Bowman et al., Crop Sci. 36:577-581 (1996); Meredith, supra].

There are many varieties of cotton plant, from which cotton fibers with a range of characteristics can be obtained and used for various applications. Cotton fibers may be characterized according to a variety of properties, some of which are considered highly desirable within the textile industry for the production of increasingly high quality products and optimal exploitation of modern spinning technologies. Commercially desirable properties include length, length uniformity, fineness, maturity ratio, decreased fuzz fiber production, micronaire, bundle strength, and single fiber strength. Much effort has been put into the improvement of the characteristics of cotton fibers mainly focusing on fiber length and fiber fineness. In particular, there is a great demand for cotton fibers of specific lengths.



Methods for improving the characteristics or yield of cotton fibers can be classified into the following three categories:

1. Variety improvement by cross breeding

This method has been utilized most widely so far. At present, almost all the cultivated varieties of cotton plant are bred by this method. However, improvement of cotton fiber yield using traditional breeding is relatively slow and inefficient and the degree of variability which can be achieved is limited.

2. Treatment with plant hormones

Plant hormones such as auxin, gibberellin, cytokinin and ethylene have been widely used in field crops or horticultural products. The influence of plant hormones, particularly gibberellin, auxin and brassinolide, on the fiber characteristics of cotton plants is known [e.g. U.S. Pat. No. 5880110 produces cotton fibers with improved fiber characteristics by treatment with brassinosteroids]. However, no measurable effect has been documented, making practical use of these hormones on a large scale highly unlikely.

3. Variety improvement by genetic engineering:

The broad acceptance of genetically engineered cotton in the leading producing countries and the fact that it is a non-food crop make it an attractive candidate for genetic engineering for improvement of fiber yield and/or quality.

In recent years, remarkable progress has been made in plant genetic engineering, as a result several cases of successful variety improvement of commercially important crop plants have been reported (e.g., cotton, soybean, corn, canola, tomato). For example, methods of improving insect resistance by the introduction of a gene coding for BT toxin (i.e., insecticidal protein toxin produced by *Bacillus thuringiensis*) in a cotton plant, have been developed and put to practical use. In addition, cotton plants with improved herbicide (Glyphosate) resistance have been genetically engineered by the introduction of a gene coding for 5-enol-pyruvyl-shikimic acid 3-phosphate synthetase.

The availability and success of plant genetic engineering combined with the fact that cotton is an excellent candidate for genetic manipulation via recombinant techniques have led researchers to postulate that if a gene associated with an improved cotton fiber property could be identified, it could be up-regulated using recombinant techniques thus improving the characteristics or yield of cotton fibers.



Conversely, if a gene associated with a decline in a cotton fiber property could be identified, it could be down-regulated using gene silencing methods. For this purpose, the mechanisms of fiber elongation and formation must be elucidated on the genetic level and genes closely associated with these mechanisms must be identified.

5 A cotton fiber is composed of a single cell that has differentiated from an epidermal cell of the seed coat, developing through four stages, i.e., initiation, elongation, secondary cell wall thickening and maturation stages. More specifically, the elongation of a cotton fiber commences in the epidermal cell of the ovule immediately following flowering, after which the cotton fiber rapidly elongates for  
10 approximately 21 days. Fiber elongation is then terminated, and a secondary cell wall is formed and grown through maturation to become a mature cotton fiber.

Several candidate genes have been isolated which are associated with the elongation and formation of cotton fibers. For example, five genes from cotton plants have been identified that are specifically expressed at the cotton fiber elongation stage  
15 by differential screening method and differential display method, [U.S. Pat. No. 5,880,100 and U.S. patent applications Ser. Nos. 08/580,545, 08/867,484 and 09/262,653].

WO0245485 describes methods and means to modulate fiber quality in fiber-producing plants, such as cotton, by modulating sucrose synthase (a sugar important  
20 for cell wall synthesis) activity and/or expression in such plants.

U.S. Pat. No. 6,472,588 and WO0117333 provide methods for increasing the quality of cotton fiber produced from a cotton plant by transformation with a DNA encoding sucrose phosphate synthase. The fiber qualities include strength, length, fiber maturity ratio, immature fiber content, fiber uniformity and micronaire.

25 WO9508914 discloses a fiber producing plant comprising in its genome a heterologous genetic construct. The genetic construct comprises a fiber-specific promoter and a coding sequence encoding a plant peroxidase, such as a cotton peroxidase.

WO9626639 provides methods whereby an ovary specific promoter sequence  
30 is utilized to express plant growth modifying hormones in cotton ovule tissue. The methods permit the modification of the characteristics of boll set in cotton plants and provide a mechanism for altering fiber quality characteristics such as fiber dimension and strength.



U.S. Pat. No. 5,981,834, U.S. Pat. No. 5,597,718, U.S. Pat. No. 5,620,882, U.S. Pat. No. 5,521,708 and U.S. Pat. No. 5,495,070 all disclose a method for genetically engineering a fiber-producing plant and the identification of cDNA clones useful for identifying fiber genes in cotton. The cDNA clones are useful in developing  
5 corresponding genomic clones from fiber producing plants to enable genetic engineering of cotton and other plants using these genes. Coding sequences from these isolated genes are used in sense or antisense orientation to alter the fiber characteristics of transgenic fiber producing plants.

U.S. patent applications U.S. 2002049999 and U.S. 2003074697 both disclose  
10 cotton plants of the genus *Gossypium* with improved cotton fiber characteristics. The cotton plant has an expression cassette containing a gene coding for an enzyme selected from the group consisting of endoxyloglucan transferase, catalase and peroxidase so that the gene is expressed in cotton fiber cells to improve the cotton fiber characteristics.

15 WO 01/40250 provides methods for improving cotton fiber quality by modulating transcription factor gene expression.

WO 96/40924 provides novel DNA constructs which may be used as molecular probes or alternatively inserted into a plant host to provide for modification of transcription of a DNA sequence of interest during various stages of cotton fiber  
20 development. The DNA constructs comprise a cotton fiber transcriptional initiation regulatory region associated with a gene, which is expressed in cotton fiber. Also provided is a novel cotton having a cotton fiber which has a natural color. The color was achieved by the introduction and expression in cotton fiber cell of a pigment gene construct.

25 EP0834566 provides a gene which controls the fiber formation mechanism in cotton plant and which can be used for industrially useful improvement.

However, beside Sucrose Synthase, there is no evidence to date that the expression of any particular gene plays an essential role in cotton fiber formation or enhanced fiber characteristics.

30 Thus, there remains a need for identifying other genes associated with fiber characteristics of cotton plants and a more thorough search for quality-related genes is required.



While reducing the present invention to practice the present inventors devised and employed a novel computational approach that utilizes comparative genomics to identify genes which play a pivotal role in fiber development. As demonstrated herein, expression of such genes correlates with fiber length and their overexpression  
5 is sufficient to modify tomato seed hair, an ultimate model for cotton fibers. These results suggest that polynucleotides of the present invention can be used for generating transgenic cotton plants which are characterized by fibers of desired length.

## 10 SUMMARY OF THE INVENTION

According to one aspect of the present invention there is provided an isolated polynucleotide comprising a nucleic acid sequence encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96, wherein the polypeptide is capable of regulating cotton fiber development.

According to further features in preferred embodiments of the invention described below, the nucleic acid sequence is selected from the group consisting of SEQ ID NOs. 1, 2, 4, 5, 7, 9, 10, 16, 17, 20, 21, 22, 24, 25, 27 and 13.

According to still further features in the described preferred embodiments the polypeptide is as set forth in SEQ ID NO. 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96.

According to still further features in the described preferred embodiments the amino acid sequence is as set forth in SEQ ID NO. 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96.

According to still further features in the described preferred embodiments the cotton fiber development comprises fiber formation.

According to still further features in the described preferred embodiments the cotton fiber development comprises fiber elongation.

According to another aspect of the present invention there is provided an isolated polynucleotide comprising a nucleic acid sequence at least 80 % identical to SEQ ID NO: 85 or 91, wherein the nucleic acid sequence is capable of regulating expression of at least one polynucleotide sequence operably linked thereto in an ovule endothelial cell.



According to still further features in the described preferred embodiments the ovule endothelial cell is of a plant fiber or a trichome.

According to yet another aspect of the present invention there is provided an oligonucleotide capable of specifically hybridizing to the isolated polynucleotide.

According to another aspect of the present invention there is provided a nucleic acid construct comprising the isolated polynucleotide.

According to still further features in the described preferred embodiments the nucleic acid construct further comprising at least one cis-acting regulatory element operably linked to the isolated polynucleotide.

According to still further features in the described preferred embodiments the polynucleotide sequence is selected from the group consisting of SEQ ID NOs: 1, 2, 4, 5, 7, 9, 10, 16, 17, 20, 21, 22, 24, 25, 27 and 13.

According to still further features in the described preferred embodiments the cis-acting regulatory element is as set forth in SEQ ID NO: 74, 75, 85 or 91 or functional equivalents thereof.

According to an additional aspect of the present invention there is provided a transgenic cell comprising the nucleic acid construct.

According to yet an additional aspect of the present invention there is provided a transgenic plant comprising the nucleic acid construct.

According to yet another aspect of the present invention there is provided a method of improving fiber quality and/or yield of a fiber producing plant, the method comprising regulating an expression level or activity of at least one polynucleotide encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96 in the fiber producing plant, thereby improving the quality and/or yield of the fiber producing plant.

According to still further features in the described preferred embodiments the quality of the fiber producing plant comprises at least one parameter selected from the group consisting of fiber length, fiber strength, fiber weight per unit length, maturity ratio, uniformity and micronaire.

According to still further features in the described preferred embodiments the regulating expression or activity of the at least one polynucleotide is up-regulating.



According to still further features in the described preferred embodiments the up-regulating is effected by introducing into the cotton the nucleic acid construct.

According to still further features in the described preferred embodiments the regulating expression or activity of the at least one polynucleotide is down-regulating.

According to still further features in the described preferred embodiments the down-regulating is effected by gene silencing.

According to still further features in the described preferred embodiments the gene silencing is effected by introducing into the cotton the oligonucleotide.

According to still further features in the described preferred embodiments the fiber producing plant is selected from the group consisting of cotton, silk cotton tree (Kapok, *Ceiba pentandra*), desert willow, creosote bush, winterfat, balsa, ramie, kenaf, hemp, roselle, jute, sisal abaca and flax.

According to still an additional aspect of the present invention there is provided a method of increasing a biomass of a plant, the method comprising regulating an expression level or activity of at least one polynucleotide encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96 in the plant, thereby increasing the biomass of the plant.

According to still further features in the described preferred embodiments the plant is a monocot plant.

According to still further features in the described preferred embodiments the plant is a dicot plant.

According to a further aspect of the present invention there is provided a method of identifying genes which are involved in cotton fiber development, the method comprising:

- (a) providing expressed nucleic acid sequences derived from cotton fibers;
- (b) providing expressed nucleic acid sequences derived from an ovule tissue;
- (c) computationally assembling the expressed nucleic acid sequences of (a) and (b) to generate clusters; and
- (d) identifying clusters of the clusters which comprise expressed nucleic acid sequences of (a) and (b), thereby identifying genes which are involved in cotton fiber development.



According to still further features in the described preferred embodiments the method further comprising identifying genes which are differentially expressed in the cotton fiber following (d).

According to still further features in the described preferred embodiments the differentially expressed comprises:

- (a) specific expression; and/or
- (b) change in expression over fiber development.

According to yet an additional aspect of the present invention there is provided a method of producing an insect resistant plant, comprising regulating an expression level or activity of at least one polynucleotide encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96 in a trichome of the plant, thereby producing the insect resistant plant.

According to still an additional aspect of the present invention there is provided a method of producing cotton fibers, the method comprising:

- (a) generating a transgenic cotton plant expressing at least one polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96; and
- (b) harvesting the fibers of the transgenic cotton plant, thereby producing the cotton fibers.

5 The present invention successfully addresses the shortcomings of the presently known configurations by providing genes involved in cotton fiber development and methods of using same.

10 Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. In case of conflict, the patent specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.



BRIEF DESCRIPTION OF THE DRAWINGS

The invention is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for a fundamental understanding of the invention, the description taken with the drawings making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

In the drawings:

FIG. 1 is an illustration depicting the bioinformatic methodology of the present invention effected to identify genes which may be used to improve cotton fiber yield and quality.

FIGs. 2a-d are bar graphs showing expression patterns of fiber specific genes (CT\_11 Figure 2b), elongation associated genes (CT\_1, Figure 2c) and initiation associated genes (CT\_22, Figure 2d).

FIG. 3 is a graph depicting expression of CT\_76 in varieties of cotton (*G. hirsutum* var Tamcot, Coker and Acala, and *G. barbadense* var Pima S5) plants, as determined by RT-PCR.

FIG. 4 is a schematic illustration of the pPi binary plasmid.

FIGs. 5a-l are photographs of wild-type and transgenic arabidopsis plants over-expressing genes of the present invention. Figure 5a shows two week old rosette of wt plants; Figure 5b shows two week old rosette of CT11 over-expressing arabidopsis plants; Figure 5c shows two week old roots of CT11; Figure 5d shows three week old wild type arabidopsis; Figure 5e shows three week old CT\_20; Figure 5f shows three week old CT\_22; Figure 5g shows 30 days old rosettes of wt and CT\_9; Figure 5h shows 30 days inflorescence of wt and CT\_9; Figure 5i shows two week old roots of CT9; Figure 5j shows 30 days old rosettes of wt and CT\_40; Figure 5k shows rosette of 5 week old wt and CT81 over-expressing plants; Figure 5l shows a leaf of wt and CT81 over-expressing arabidopsis plants;



FIGs. 6a-e are photographs depicting wild-type and transgenic tomato plants over-expressing CT\_20. Figure 6a shows a leaf of wild-type plant; Figure 6b shows a leaf of CT\_20 transgenic tomato; Figure 6c shows seed hairs of WT and CT\_20 over-expressing tomato plants; Figure 6d shows section of a wt tomato seed; Figure 6e shows section of a CT\_20 over-expressing tomato seed; Figure 6f seed hairs of WT and CT\_82.

FIGs. 7a-b are photographs depicting transgenic tomato plants over-expressing GUS under the expression of the CT\_2 promoter. Figure 7a is a cut through transgenic tomato fruit, over-expressing GUS under CT2 promoter in the mature green stage (x 5 magnification). Figure 7b similar to Figure 7a showing x 25 magnification;

FIGs. 8a-b are photographs depicting various magnifications of wild-type and transgenic tomato fruits or tomato seeds. Figure 8a is a single wild type tomato seed covered with seed hairs x 10 magnification; Figure 8b shows tomato seed over expressing expansin under 35S (x 10 magnification).

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is of polypeptides and polynucleotides encoding same which are involved in plant fiber development and which can be used to improve fiber quality and/or yield/biomass of a fiber producing plant.

The principles and operation of the present invention may be better understood with reference to the drawings and accompanying descriptions.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details set forth in the following description or exemplified by the Examples. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

Cotton and cotton by-products provide raw materials that are used to produce a wealth of consumer-based products; in addition to textiles, cotton is used to produce foodstuffs, livestock feed, fertilizer and paper. The production, marketing, consumption and trade of cotton-based products generate an excess of \$100 billion annually in the U.S. alone, making cotton the number one value-added crop.



Over the past decade cotton fiber production has sharply declined prompting cotton growers and researchers to look for approaches, which can be used to improve fiber yield and quality.

Increasing fiber quality and/or yield under diverse environmental conditions will increase the profitability of cotton crop production and provide a new spectrum of material properties for exploitation by the processing industries.

While reducing the present invention to practice, the present inventors have configured a novel computational approach that utilizes comparative genomics to identify genes which play a role in fiber development. Genes identified using this approach may be successfully used for generating transgenic plants which are featured by fibers of desired properties.

Thus, according to one aspect of the present invention there is provided a method of identifying genes which are involved in cotton fiber development.

As used herein the term "cotton" refers to a wild-type, a cultivated variety (e.g., hybrid) or a transgenic cotton (*Gossypium*) plant.

As used herein the phrase "fiber development" refers to the development of the hair of the cotton seed.

As used herein the term "development" when used in context of cotton fibers refers to initiation of the fiber and/or elongation thereof, as well as to the fiber secondary cell wall thickening and maturation.

The method according to this aspect of the present invention is effected by:

- (a) providing expressed nucleic acid sequences derived from cotton fibers;
- (b) providing expressed nucleic acid sequences derived from an ovule tissue (i.e., a tissue developed from an ovary of a seed plant. Examples include, but are not limited to, carpels, seed coat, embryo, endosperm);
- (c) computationally assembling the expressed nucleic acid sequences of (a) and (b) to generate clusters; and
- (d) identifying clusters of said clusters which comprise expressed nucleic acid sequences of (a) and (b), thereby identifying genes which are involved in cotton fiber development.

Expressed nucleic acid sequences used as a potential source for identifying genes involved in cotton fiber development according to this aspect of the present invention are preferably libraries of expressed messenger RNA [i.e., expressed



sequence tags (EST), cDNA clones, contigs, pre-mRNA, etc.] obtained from tissue or cell-line preparations which can include genomic and/or cDNA sequence.

Expressed nucleic acid sequences, according to this aspect of the present invention can be retrieved from pre-existing publicly available databases (see  
5 Example 1 of the Examples section which follows or private databases).

Alternatively, the expressed nucleic acid sequences utilized by the present invention can be generated from sequence libraries (e.g., cDNA libraries, EST libraries, mRNA libraries and others).

cDNA libraries are suitable sources for expressed sequence information.

10 Generating a sequence database in such a case is typically effected by tissue or cell sample preparation, RNA isolation, cDNA library construction and sequencing.

It will be appreciated that such cDNA libraries can be constructed from RNA isolated from whole plant, specific tissues, or cell populations.

Once expressed sequence data is obtained from both cotton fibers and an ovule  
15 tissue, sequences may be clustered to form contigs. See Example 1 of the Examples section which follows

Such contigs are then assembled to identify homologous sequences (of cotton fibers and ovule tissue) present in the same cluster, such contigs are considered to be involved in cotton fiber development.

20 A number of commonly used computer software fragment read assemblers capable of forming clusters of expressed sequences are commercially available. These packages include but are not limited to, The TIGR Assembler [Sutton G. et al. (1995) Genome Science and Technology 1:9-19], GAP [Bonfield JK. et al. (1995) Nucleic Acids Res. 23:4992-4999], CAP2 [Huang X. et al. (1996) Genomics 33:21-  
25 31], The Genome Construction Manager [Laurence CB. Et al. (1994) Genomics 23:192-201], Bio Image Sequence Assembly Manager, SeqMan [Swindell SR. and Plasterer JN. (1997) Methods Mol. Biol. 70:75-89], LEADS and GenCarta (Compugen Ltd. Israel).

Once genes which are involved in cotton fiber development are identified their pattern of expression can be analyzed as described in Example 2 of the Examples section which follows, to thereby identify genes which are differentially expressed in the cotton fiber (i.e., specific expression) or during cotton fiber development (i.e., change in expression during cotton fiber development).



Methods of identifying differentially expressed genes are well known in the art.

Using the above methodology, the present inventors were able to successfully identify genes which are involved in cotton fiber development.

As is illustrated in the Examples section which follows genes identified using the teachings of the present invention can be classified into 6 functional categories according to their sequence homology to known proteins and enzymes (Table 3, below). The Two genes were classified into a cell fate commitment category: homologous to the MYB transcription factor and to GL3 which are known to be involved in trichome development in arabidopsis. The expression pattern of both genes and the phenotype of CT20 transgene both in arabidopsis and tomato T1 plants support their involvement mainly in the initiation phase. Two other genes (Table 3, above) are transcription factors from the MYB and MADS BOX families. Many studies demonstrated the function of these two transcription factor families as homeotic genes with key role in different developmental processes, among them are trichome and fiber morphogenesis (Suo. J. *et. al.* 2003, Ferrario S *et. al.* 2004). Their role in early stages of fiber development is supported also by their RNA expression pattern, which, is induced before, and during the day of anthesis. One gene belongs to the pathways of starch and sucrose metabolism. A recent work demonstrates that another gene (SUS), which, belongs to this pathway, is a limiting factor in both fiber initiation and development. Another gene (Table 3, below) is classified as lipid transport whose RNA expression is highly induced during early fiber elongation stage fit to the fact that lipids are key components in fiber formation. Several genes (Table 3, below) were classified either as genes involved in desiccation, salinity response stimulated by abscisic acid and genes involved in electron transfer. Out of them 3 genes were selected by RNA expression pattern to be induced in the elongation stage.

In view of the above and together with the experimental results which correlate gene expression with fiber length, it is suggested that genes of the present invention can be used to generate fiber producing plants with commercially desired fiber quality.

Thus, the present invention encompasses polynucleotides identified using the present methodology and their encoded polypeptide as well as functional equivalents of the polypeptides identified herein (i.e., , polypeptides which are capable of



regulating cotton fiber development, as can be determined according to the assays described in the Examples section which follows) and their coding sequences. Such functional equivalents can be at least about 70 %, at least about 75 %, at least about 80 %, at least about 81 %, at least about 82 %, at least about 83 %, at least about 84 %, at least about 85 %, at least about 86 %, at least about 87 %, at least about 88 %, at least about 89 %, at least about 90 %, at least about 91 %, at least about 92 %, at least about 93 %, at least about 94 %, at least about 95 %, at least about 75 %, at least about 75 %, at least about 75 %, say 100 % homologous to SEQ ID NO: 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 95 or 96.

Polynucleotides encoding functional equivalents can be at least about 70 %, at least about 75 %, at least about 80 %, at least about 81 %, at least about 82 %, at least about 83 %, at least about 84 %, at least about 85 %, at least about 86 %, at least about 87 %, at least about 88 %, at least about 89 %, at least about 90 %, at least about 91 %, at least about 92 %, at least about 93 %, at least about 94 %, at least about 95 %, at least about 75 %, at least about 75 %, at least about 75 %, say 100 % identical to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 27.

Homology (e.g., percent homology) can be determined using any homology comparison software, including for example, the BlastP software of the National Center of Biotechnology Information (NCBI) such as by using default parameters.

Identity (e.g., percent homology) can be determined using any homology comparison software, including for example, the BlastN software of the National Center of Biotechnology Information (NCBI) such as by using default parameters.

As used herein the phrase "an isolated polynucleotide" refers to a single or double stranded nucleic acid sequences which is isolated and provided in the form of an RNA sequence, a complementary polynucleotide sequence (cDNA), a genomic polynucleotide sequence and/or a composite polynucleotide sequences (e.g., a  
5 combination of the above).

As used herein the phrase "complementary polynucleotide sequence" refers to a sequence, which results from reverse transcription of messenger RNA using a reverse transcriptase or any other RNA dependent DNA polymerase. Such a sequence



can be subsequently amplified *in vivo* or *in vitro* using a DNA dependent DNA polymerase.

As used herein the phrase "genomic polynucleotide sequence" refers to a sequence derived (isolated) from a chromosome and thus it represents a contiguous  
5 portion of a chromosome.

As used herein the phrase "composite polynucleotide sequence" refers to a sequence, which is at least partially complementary and at least partially genomic. A composite sequence can include some exonal sequences required to encode the polypeptide of the present invention, as well as some intronic sequences interposing  
10 therebetween. The intronic sequences can be of any source, including of other genes, and typically will include conserved splicing signal sequences. Such intronic sequences may further include cis acting expression regulatory elements.

According to a preferred embodiment of this aspect of the present invention, the nucleic acid sequence is as set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 19, 21, 22, 23, 24, 25 or 26.

According to another preferred embodiment of this aspect of the present invention, the isolated polynucleotide is as set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 27.

According to yet another preferred embodiment of this aspect of the present invention, the polypeptide is as set forth in SEQ ID NO: 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 95 or 96.

According to still another preferred embodiment of this aspect of the present invention, the amino acid sequence is as set forth in SEQ ID NO: 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 95 or 96.

The isolated polynucleotides of this aspect of the present invention can also be qualified using a hybridization assay by incubating the isolated polynucleotides  
15 described above in the presence of oligonucleotide probe or primer under moderate to stringent hybridization conditions.

Moderate to stringent hybridization conditions are characterized by a hybridization solution such as containing 10 % dextrane sulfate, 1 M NaCl, 1 % SDS and  $5 \times 10^6$  cpm  $^{32}\text{P}$  labeled probe, at 65 °C, with a final wash solution of 0.2 x SSC  
20 and 0.1 % SDS and final wash at 65°C and whereas moderate hybridization is



16

effected using a hybridization solution containing 10 % dextrane sulfate, 1 M NaCl, 1 % SDS and  $5 \times 10^6$  cpm  $^{32}\text{P}$  labeled probe, at 65 °C, with a final wash solution of 1 x SSC and 0.1 % SDS and final wash at 50 °C.

Thus, the present invention encompasses nucleic acid sequences described  
5 hereinabove; fragments thereof, sequences hybridizable therewith, sequences homologous thereto, sequences encoding similar polypeptides with different codon usage, altered sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or man induced, either randomly or in a targeted fashion.

10 Since the polynucleotide sequences of the present invention encode previously unidentified polypeptides, the present invention also encompasses novel polypeptides or portions thereof, which are encoded by the isolated polynucleotides and respective nucleic acid fragments thereof described hereinabove.

Thus, the present invention also encompasses polypeptides encoded by the  
15 polynucleotide sequences of the present invention. The amino acid sequences of these novel polypeptides are set forth in SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96.

The present invention also encompasses homologues of these polypeptides, such homologues can be at least about 70 %, at least about 75 %, at least about 80 %, at least about 81 %, at least about 82 %, at least about 83 %, at least about 84 %, at least about 85 %, at least about 86 %, at least about 87 %, at least about 88 %, at least about 89 %, at least about 90 %, at least about 91 %, at least about 92 %, at least about 93 %, at least about 93 %, at least about 94 %, at least about 95 %, at least about 96 %, at least about 97 %, at least about 98 %, at least about 99 %, or more say  
25 100 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96.

The present invention also encompasses fragments of the above described polypeptides and polypeptides having mutations, such as deletions, insertions or substitutions of one or more amino acids, either naturally occurring or man induced,  
30 either randomly or in a targeted fashion.

The ability of polynucleotides of the present invention and their products to regulate cotton fiber development can be determined directly on at least one structural parameter of a cotton fiber such as fiber length or fiber finesse, or fiber growth rate



(further described hereinbelow). However cotton fiber development can also be determined indirectly such as by plant model systems for cotton fiber development. For example, it is well established that trichome cells and root hairs share common characteristics with cotton fiber cells, and as such can be used as model systems for cotton fiber development [Reviewed in Wagner, G.J. et. al. (2004)], as demonstrated in details in Example 12 of the Examples section which follows.

By analyzing expression profiles, the present inventors were able to determine the involvement of the biomolecular sequences (i.e., polynucleotides and polypeptides) of the present invention in fiber initiation and/or elongation. These results were further substantiated by establishing a correlation between gene expression and fiber length (see Example 7).

These results suggest that biomolecular sequences of the present invention (e.g., polynucleotides, polypeptides, promoters, oligonucleotides, antibodies, also referred to herein as agents) can be used to improve fiber quality and/or yield of a fiber producing plant.

Thus, according to yet another aspect of the present invention there is provided a method of improving fiber quality and/or yield of a fiber producing plant.

The method of this aspect of the present invention is effected by regulating an expression level or activity of at least one polynucleotide or polypeptide of the present invention (described hereinabove) in the fiber producing plant, thereby improving the quality and/or yield of the fiber producing plant.

As used herein the phrase "fiber producing plant" refers to plants that share the common feature of having an elongated shape and abundant cellulose in thick cell walls, typically termed as secondary walls. Such walls may or may not be lignified, and the protoplast of such cells may or may be viable at maturity. Such fibers have many industrial uses, for example in lumber and manufactured wood products, paper, textiles, sacking and boxing material, cordage, brushes and brooms, filling and stuffing, caulking, reinforcement of other materials, and manufacture of cellulose derivatives.

According to a preferred embodiment of this aspect of the present invention the fiber producing plant is cotton.

The term "fiber" is usually inclusive of thick-walled conducting cells such as vessels and tracheids and to fibrillar aggregates of many individual fiber cells. Hence,



the term "fiber" refers to (a) thick-walled conducting and non-conducting cells of the xylem; (b) fibers of extraxylary origin, including those from phloem, bark, ground tissue, and epidermis; and (c) fibers from stems, leaves, roots, seeds, and flowers or inflorescences (such as those of *Sorghum vulgare* used in the manufacture of brushes and brooms).

Example of fiber producing plants, include, but are not limited to, agricultural crops such as cotton, silk cotton tree (Kapok, *Ceiba pentandra*), desert willow, creosote bush, winterfat, balsa, kenaf, roselle, jute, sisal abaca, flax, corn, sugar cane, hemp, ramie, kapok, coir, bamboo, spanish moss and *Agave* spp. (e.g. sisal).

As used herein the phrase "fiber quality" refers to at least one fiber parameter which is agriculturally desired, or required in the fiber industry (further described hereinbelow). Examples of such parameters, include but are not limited to, fiber length, fiber strength, fiber fitness, fiber weight per unit length, maturity ratio and uniformity (further described hereinbelow).

Cotton fiber (lint) quality is typically measured according to fiber length, strength and fineness. Accordingly, the lint quality is considered higher when the fiber is longer, stronger and finer.

As used herein the phrase "fiber yield" refers to the amount or quantity of fibers produced from the fiber producing plant.

As used herein the term "improving" refers to at least about 5 %, at least about 10 %, at least about 15 %, at least about 20 %, at least about 30 %, at least about 40 %, at least about 50 %, change in fiber quality/yield as compared to a native plant (i.e., not modified with the biomolecular sequences of the present invention).

As used herein the term "regulating" refers to up regulating, down regulating or a combination thereof. For example, when an increase in fiber number is desired the present invention can be effected by upregulating at least one polynucleotide of the present invention, which is involved in fiber initiation (e.g., SEQ ID NOs: 4, 10, 9, 12, 16 and 25). Alternatively, when short fibers are desired such as for example, in corn, then the present invention is effected by down regulating at least one polynucleotide of the present invention which is involved in fiber elongation (e.g., SEQ ID NOs. 1, 2, 3, 5, 6, 7, 17, 18, 19, 20, 21, 22, 23, 24 and 27). Alternatively, the present invention can be effected by upregulating expression of at least one polynucleotide (such as involved in fiber elongation) and down regulating at least one



polynucleotide (such as involved in fiber initiation) of the polynucleotides of the present invention. In this manner it is feasible to obtain a fiber producing plant with improved fiber yield of each of short length.

Up regulating an expression level of at least one of the polynucleotides of the present invention can be effected at the genomic level (e.g., activation of transcription by means of promoters, enhancers, or other regulatory elements), at the transcript level, or at the protein level.

Following is a non-comprehensive list of agents capable of upregulating the expression level and/or activity of the biomolecular sequences (i.e., nucleic acid or protein sequences) of the present invention.

An agent capable of upregulating expression of a polynucleotide of interest may be an exogenous polynucleotide sequence designed and constructed to express at least a functional portion thereof (e.g., improving fiber yield/quality, increasing biomass etc.). Accordingly, the exogenous polynucleotide sequence may be a DNA or RNA sequence encoding a polypeptide molecule, capable of improving fiber yield or quantity. Alternatively, the exogenous polynucleotide may be a cis-acting regulatory region (e.g., SEQ ID NO: 74, 75, 85, 88 or 91) which may be introduced into the plant to increase expression of any polynucleotide which is involved in fiber development (e.g., sucrose phosphate synthase, as described in U.S. Pat. No. 6,472,588).

To express exogenous polynucleotides in plant cells, a polynucleotide sequence of the present invention is preferably ligated into a nucleic acid construct suitable for plant cell expression. Such a nucleic acid construct includes a cis-acting regulatory region such as a promoter sequence for directing transcription of the polynucleotide sequence in the cell in a constitutive or inducible manner. The promoter may be homologous or heterologous to the transformed plant/cell.

Preferred promoter sequences which can be used in accordance with this aspect of the present invention are endothelial cell promoters.

For example, promoter sequences of each of the polynucleotide sequences of the present invention may be preferably used in the nucleic acid constructs of the present invention.

According to a preferred embodiment of this aspect of the present invention the promoter is at least about 80 %, at least about 81 %, at least about 82 %, at least about 83 %, at least about 84 %, at least about 85 %, at least about 86 %, at least



about 87 %, at least about 88 %, at least about 89 %, at least about 90 %, at least about 91 %, at least about 92 %, at least about 93 %, at least about 94 %, at least about 95 %, at least about 96 %, at least about 97 %, at least about 98 %, at least about 99 %, or 100 % identical to SEQ ID NO. 85 or 91, which is capable of  
5 regulating expression of at least one polynucleotide sequence operably linked thereto in an ovule endothelial cell (i.e., capable of exerting a regulatory effect on the coding sequence linked thereto).

As is clearly illustrated in the Examples section which follows, such promoter sequences are capable of regulating expression of a coding nucleic acid sequence  
10 (e.g., GUS) operably linked thereto.

Other examples of cotton fiber-enhanced promoters include those of the cotton fiber-expressed genes E6 (John et al., Plant Mol. Biol., 30:297-306 (1996) and John et al., Proc. Natl. Acad. Sci., 93:12768-12773 (1996) e), H6 (John et al., Plant Physiol., 108:669-676, (1995)), FbL2A (Rinehart et al., Plant Physiol., 112:1331-1341 (1996)  
15 and John et al, Proc. Natl. Acad. Sci. USA, 93:12768-12773 (1996)), rac (Delmer et al., Mol. Gen. Genet., 248:43-51 (1995)); CelA (Pear et al., Proc. Natl. Acad. Sci USA, 93:12637-12642 (1996)); CAP (Kawai et al., Plant Cell Physiol. 39:1380-1383 (1998)); ACP (Song et al., Biochim. Biophys. Acta 1351:305-312 (1997); and LTP (Ma et al., Biochim. Biophys. Acta 1344:111-114 (1997)). Other cotton fiber  
20 specific promoters are disclosed in U.S. Pat. No. 5,495,070.

Other promoters which can be used in accordance with this aspect of the present invention are those that ensure expression only in specified organs, such as the leaf, root, tuber, seed, stem, flower or specified cell types such as parenchyma, epidermal, trichome or vascular cells.

25 Preferred promoters for enhancing expression in trichome cells are disclosed in WO 2004/111183, to Evogene Ltd.

Preferred promoters enhancing expression in vascular tissue include the CAD 2 promoter (Samaj et al., Planta, 204:437-443 (1998)), the Pt4C11 promoter (Hu et al., Proc. Natl. Acad. Sci. USA, 95:5407-5412 (1998)), the C4H promoter (Meyer et al., Proc. Natl. Acad. Sci. USA, 95:6619-6623 (1998)), the PtX3H6 and PtX14A9  
30 promoters (Loopstra et al., Plant Mol. Biol., 27:277-291 (1995)), the RolC promoter (Graham, Plant Mol. Biol., 33:729-735 (1997)), the Hvhspl7 promoter (Raho et al., J.



Expt. Bot., 47:1587-1594 (1996)), and the COMT promoter (Capellades et al., Plant Mol. Biol., 31:307-322 (1996)).

Preferred promoters enhancing expression in stem tissue include pith promoters (Datta, Theor. Appl. Genet., 97:20-30 (1998) and Ohta et al., Mol. Gen. Genet., 225:369-378 (1991)), and the anionic peroxidase promoter (Klotz et al., Plant Mol. Biol., 36:509-520 (1998)). Preferred promoters enhancing expression in phloem, cortex and cork, but not xylem or pith, include the Psam-1 promoter (Mijnsbrugge et al., Plant and Cell Physiol., 37:1108-1115 (1996)).

Preferred promoters enhancing expression in seeds include the phas promoter (Geest et al., Plant Mol. Biol. 32:579-588 (1996)); the GluB-1 promoter (Takaiwa et al., Plant Mol. Biol. 30:1207-1221 (1996)); the gamma-zein promoter (Torrent et al., Plant Mol. Biol. 34:139-149 (1997)), and the oleosin promoter (Sarmiento et al., The Plant Journal 11:783-796 (1997)).

Other promoter sequences which mediate constitutive, inducible, tissue-specific or developmental stage-specific expression are disclosed in WO 2004/081173 to Evogene Ltd.

Truncated or synthetic promoters including specific nucleotide regions conferring tissue-enhanced expression may also be used, as exemplified by identification of regulatory elements within larger promoters conferring xylem-enhanced expression (Seguin et al., Plant Mol. Biol., 35:281-291 (1997); Torres-Schumann et al., The Plant Journal, 9:283-296 (1996); and Leyva et al., The Plant Cell, 4:263-271 (1992)).

The nucleic acid construct can be, for example, a plasmid, a bacmid, a phagemid, a cosmid, a phage, a virus or an artificial chromosome. Preferably, the nucleic acid construct of the present invention is a plasmid vector, more preferably a binary vector.

The phrase "binary vector" refers to an expression vector which carries a modified T-region from Ti plasmid, enable to be multiplied both in *E. coli* and in *Agrobacterium* cells, and usually comprising reporter gene(s) for plant transformation between the two boarder regions. A binary vector suitable for the present invention includes pBI2113, pBI121, pGA482, pGAH, pBIG, pBI101 (Clontech), pPI (see Example 5 of the Examples section which follows) or modifications thereof.



The nucleic acid construct of the present invention can be utilized to transform a host cell (e.g., bacterial, plant) or plant.

As used herein, the terms "transgenic" or "transformed" are used interchangeably referring to a cell or a plant into which cloned genetic material has  
5 been transferred.

In stable transformation, the nucleic acid molecule of the present invention is integrated into the plant genome, and as such it represents a stable and inherited trait. In transient transformation, the nucleic acid molecule is expressed by the cell transformed but not integrated into the genome, and as such represents a transient  
10 trait.

There are various methods of introducing foreign genes into both monocotyledonous and dicotyledonous plants (Potrykus, I. (1991). *Annu Rev Plant Physiol Plant Mol Biol* 42, 205-225; Shimamoto, K. et al. (1989). Fertile transgenic rice plants regenerated from transformed protoplasts. *Nature* (1989) 338, 274-276).

15 The principal methods of the stable integration of exogenous DNA into plant genomic DNA includes two main approaches:

(i) *Agrobacterium-mediated gene transfer*. See: Klee, H. J. et al. (1987). *Annu Rev Plant Physiol* 38, 467-486; Klee, H. J. and Rogers, S. G. (1989). *Cell Culture and Somatic Cell Genetics of Plants*, Vol. 6, *Molecular Biology of Plant Nuclear Genes*, pp. 2-25, J. Schell and L. K. Vasil, eds., Academic Publishers, San Diego, Cal.; and  
20 Gatenby, A. A. (1989). *Regulation and Expression of Plant Genes in Microorganisms*, pp. 93-112, *Plant Biotechnology*, S. Kung and C. J. Arntzen, eds., Butterworth Publishers, Boston, Mass.

(ii) *Direct DNA uptake*. See, e.g.: Paszkowski, J. et al. (1989). *Cell Culture and Somatic Cell Genetics of Plants*, Vol. 6, *Molecular Biology of Plant Nuclear Genes*, pp. 52-68, J. Schell and L. K. Vasil, eds., Academic Publishers, San Diego, Cal.; and Toriyama, K. et al. (1988). *Bio/Technol* 6, 1072-1074 (methods for direct uptake of DNA into protoplasts). See also: Zhang et al. (1988). *Plant Cell Rep* 7, 379-384; and Fromm, M. E. et al. (1986). Stable transformation of maize after gene  
25 transfer by electroporation. *Nature* 319, 791-793 (DNA uptake induced by brief electric shock of plant cells). See also: Klein et al. (1988). *Bio/Technology* 6, 559-563; McCabe, D. E. et al. (1988). Stable transformation of soybean (*Glycine max*) by  
30 particle acceleration. *Bio/Technology* 6, 923-926; and Sanford, J. C. (1990). *Biolistic*



plant transformation. *Physiol Plant* 79, 206-209 (DNA injection into plant cells or tissues by particle bombardment). See also: Neuhaus, J. M. et al. (1987). *Theor Appl Genet* 75, 30-36; and Neuhaus, J. M. and Spangenberg, G. C. (1990). *Physiol Plant* 79, 213-217 (use of micropipette systems). See U.S. Pat. No. 5,464,765 (glass fibers or silicon carbide whisker transformation of cell cultures, embryos or callus tissue). See also: DeWet, J. M. J. et al. (1985). "Exogenous gene transfer in maize (*Zea mays*) using DNA-treated pollen," *Experimental Manipulation of Ovule Tissue*, G. P. Chapman et al., eds., Longman, New York-London, pp. 197-209; and Ohta, Y. (1986). High-Efficiency Genetic Transformation of Maize by a Mixture of Pollen and Exogenous DNA. *Proc Natl Acad Sci USA* 83, 715-719 (direct incubation of DNA with germinating pollen).

The *Agrobacterium*-mediated system includes the use of plasmid vectors that contain defined DNA segments which integrate into the plant genomic DNA. Methods of inoculation of the plant tissue vary depending upon the plant species and the *Agrobacterium* delivery system. A widely used approach is the leaf-disc procedure, which can be performed with any tissue explant that provides a good source for initiation of whole-plant differentiation (Horsch, R. B. et al. (1988). "Leaf disc transformation." *Plant Molecular Biology Manual* 45, 1-9, Kluwer Academic Publishers, Dordrecht). A supplementary approach employs the *Agrobacterium* delivery system in combination with vacuum infiltration. The *Agrobacterium* system is especially useful for in the creation of transgenic dicotyledenous plants.

There are various methods of direct DNA transfer into plant cells. In electroporation, the protoplasts are briefly exposed to a strong electric field, opening up mini-pores to allow DNA to enter. In microinjection, the DNA is mechanically injected directly into the cells using micropipettes. In microparticle bombardment, the DNA is adsorbed on microprojectiles such as magnesium sulfate crystals or tungsten particles, and the microprojectiles are physically accelerated into cells or plant tissues.

Following stable transformation, plant propagation occurs. The most common method of plant propagation is by seed. The disadvantage of regeneration by seed propagation, however, is the lack of uniformity in the crop due to heterozygosity, since seeds are produced by plants according to the genetic variances governed by Mendelian rules. In other words, each seed is genetically different and each will grow with its own specific traits. Therefore, it is preferred that the regeneration be effected



such that the regenerated plant has identical traits and characteristics to those of the parent transgenic plant. The preferred method of regenerating a transformed plant is by micropropagation, which provides a rapid, consistent reproduction of the transformed plants.

5           Micropropagation is a process of growing second-generation plants from a single tissue sample excised from a selected parent plant or cultivar. This process permits the mass reproduction of plants having the preferred tissue and expressing a fusion protein. The newly generated plants are genetically identical to, and have all of the characteristics of, the original plant. Micropropagation allows for mass production  
10 of quality plant material in a short period of time and offers a rapid multiplication of selected cultivars with preservation of the characteristics of the original transgenic or transformed plant. The advantages of this method of plant cloning include the speed of plant multiplication and the quality and uniformity of the plants produced.

          Micropropagation is a multi-stage procedure that requires alteration of culture  
15 medium or growth conditions between stages. The micropropagation process involves four basic stages: stage one, initial tissue culturing; stage two, tissue culture multiplication; stage three, differentiation and plant formation; and stage four, greenhouse culturing and hardening. During stage one, the tissue culture is established and certified contaminant-free. During stage two, the initial tissue culture is  
20 multiplied until a sufficient number of tissue samples are produced to meet production goals. During stage three, the newly grown tissue samples are divided and grown into individual plantlets. At stage four, the transformed plantlets are transferred to a greenhouse for hardening where the plants' tolerance to light is gradually increased so that they can continue to grow in the natural environment.

25           Although stable transformation is presently preferred, transient transformation of, for instance, leaf cells, meristematic cells, or the whole plant is also envisaged by the present invention.

          Transient transformation can be effected by any of the direct DNA transfer methods described above or by viral infection using modified plant viruses.

30           Viruses that have been shown to be useful for the transformation of plant hosts include cauliflower mosaic virus (CaMV), tobacco mosaic virus (TMV), and baculovirus (BV). Transformation of plants using plant viruses is described in, for example: U.S. Pat. No. 4,855,237 (bean golden mosaic virus, BGMV); EPA 67,553



(TMV); Japanese Published Application No. 63-14693 (TMV); EPA 194,809 (BV); EPA 278,667 (BV); and Gluzman, Y. et al. (1988). Communications in Molecular Biology: Viral Vectors, Cold Spring Harbor Laboratory, New York, pp. 172-189. The use of pseudovirus particles in expressing foreign DNA in many hosts, including  
5 plants, is described in WO 87/06261.

Construction of plant RNA viruses for the introduction and expression of non-viral exogenous nucleic acid sequences in plants is demonstrated by the above references as well as by: Dawson, W. O. et al. (1989). A tobacco mosaic virus-hybrid expresses and loses an added gene. *Virology* 172, 285-292; French, R. et al. (1986)  
10 *Science* 231, 1294-1297; and Takamatsu, N. et al. (1990). Production of enkephalin in tobacco protoplasts using tobacco mosaic virus RNA vector. *FEBS Lett* 269, 73-76.

If the transforming virus is a DNA virus, one skilled in the art may make suitable modifications to the virus itself. Alternatively, the virus can first be cloned into a bacterial plasmid for ease of constructing the desired viral vector with the  
15 foreign DNA. The virus can then be excised from the plasmid. If the virus is a DNA virus, a bacterial origin of replication can be attached to the viral DNA, which is then replicated by the bacteria. Transcription and translation of the DNA will produce the coat protein, which will encapsidate the viral DNA. If the virus is an RNA virus, the virus is generally cloned as a cDNA and inserted into a plasmid. The plasmid is then  
20 used to make all of the plant genetic constructs. The RNA virus is then transcribed from the viral sequence of the plasmid, followed by translation of the viral genes to produce the coat proteins which encapsidate the viral RNA.

Construction of plant RNA viruses for the introduction and expression in plants of non-viral exogenous nucleic acid sequences, such as those included in the  
25 construct of the present invention, is demonstrated in the above references as well as in U.S. Pat. No. 5,316,931.

In one embodiment, there is provided for insertion a plant viral nucleic acid, comprising a deletion of the native coat protein coding sequence from the viral nucleic acid, a non-native (foreign) plant viral coat protein coding sequence, and a  
30 non-native promoter, preferably the subgenomic promoter of the non-native coat protein coding sequence, and capable of expression in the plant host, packaging of the recombinant plant viral nucleic acid, and ensuring a systemic infection of the host by the recombinant plant viral nucleic acid. Alternatively, the native coat protein coding



sequence may be made non-transcribable by insertion of the non-native nucleic acid sequence within it, such that a non-native protein is produced. The recombinant plant viral nucleic acid construct may contain one or more additional non-native subgenomic promoters. Each non-native subgenomic promoter is capable of transcribing or expressing adjacent genes or nucleic acid sequences in the plant host and incapable of recombination with each other and with native subgenomic promoters. In addition, the recombinant plant viral nucleic acid construct may contain one or more *cis*-acting regulatory elements, such as enhancers, which bind a trans-acting regulator and regulate the transcription of a coding sequence located downstream thereto. Non-native nucleic acid sequences may be inserted adjacent to the native plant viral subgenomic promoter or the native and non-native plant viral subgenomic promoters if more than one nucleic acid sequence is included. The non-native nucleic acid sequences are transcribed or expressed in the host plant under control of the subgenomic promoter(s) to produce the desired products.

15 In a second embodiment, a recombinant plant viral nucleic acid construct is provided as in the first embodiment except that the native coat protein coding sequence is placed adjacent to one of the non-native coat protein subgenomic promoters instead of adjacent to a non-native coat protein coding sequence.

20 In a third embodiment, a recombinant plant viral nucleic acid construct is provided comprising a native coat protein gene placed adjacent to its subgenomic promoter and one or more non-native subgenomic promoters inserted into the viral nucleic acid construct. The inserted non-native subgenomic promoters are capable of transcribing or expressing adjacent genes in a plant host and are incapable of recombination with each other and with native subgenomic promoters. Non-native nucleic acid sequences may be inserted adjacent to the non-native subgenomic plant viral promoters such that said sequences are transcribed or expressed in the host plant under control of the subgenomic promoters to produce the desired product.

30 In a fourth embodiment, a recombinant plant viral nucleic acid construct is provided as in the third embodiment except that the native coat protein coding sequence is replaced by a non-native coat protein coding sequence.

Viral vectors are encapsidated by expressed coat proteins encoded by recombinant plant viral nucleic acid constructs as described hereinabove, to produce a recombinant plant virus. The recombinant plant viral nucleic acid construct or



recombinant plant virus is used to infect appropriate host plants. The recombinant plant viral nucleic acid construct is capable of replication in a host, systemic spread within the host, and transcription or expression of one or more foreign genes (isolated nucleic acid) in the host to produce the desired protein.

5 In addition to the above, the nucleic acid molecule of the present invention can also be introduced into a chloroplast genome thereby enabling chloroplast expression.

A technique for introducing exogenous nucleic acid sequences to the genome of the chloroplasts is known. This technique involves the following procedures. First, plant cells are chemically treated so as to reduce the number of chloroplasts per cell to  
10 about one. Then, the exogenous nucleic acid is introduced into the cells preferably via particle bombardment, with the aim of introducing at least one exogenous nucleic acid molecule into the chloroplasts. The exogenous nucleic acid is selected by one ordinarily skilled in the art to be capable of integration into the chloroplast's genome via homologous recombination, which is readily effected by enzymes inherent to the  
15 chloroplast. To this end, the exogenous nucleic acid comprises, in addition to a gene of interest, at least one nucleic acid sequence derived from the chloroplast's genome. In addition, the exogenous nucleic acid comprises a selectable marker, which by sequential selection procedures serves to allow an artisan to ascertain that all or substantially all copies of the chloroplast genome following such selection include the  
20 exogenous nucleic acid. Further details relating to this technique are found in U.S. Pat. Nos. 4,945,050 and 5,693,507, which are incorporated herein by reference. A polypeptide can thus be produced by the protein expression system of the chloroplast and become integrated into the chloroplast's inner membrane.

Downregulation of a gene of interest can be effected on the genomic and/or  
25 the transcript level using a variety of molecules that interfere with transcription and/or translation (e.g., antisense, siRNA), or on the protein level using, e.g., antibodies, immunization techniques and the like.

For example, an agent capable of downregulating an activity of a polypeptide of interest is an antibody or antibody fragment capable of specifically binding a  
30 polypeptide of the present invention. Preferably, the antibody specifically binds at least one epitope of the polypeptide of interest. As used herein, the term "epitope" refers to any antigenic determinant on an antigen to which the paratope of an antibody binds.



Down-regulation at the RNA level can be effected by RNA-based silencing strategies which are effective in plants. See for example, Kusaba (2004) RNA interference in crop plants. *Curr. Opin. Biotechnol.* 15(2):139-43; Matzke (2001) RNA based silencing strategies in plants. *Curr. Opin. Genet.* 11:221-7.

5 For example, an agent capable of downregulating a polynucleotide of interest is a small interfering RNA (siRNA) molecule in the process of RNA interference (RNAi).

dsRNAs can be delivered to plants in several ways (reviewed in Waterhouse P, Helliwell C. 2003. Exploring plant genomes by RNA-induced gene silencing. *Nature Genet* 4: 29-38): microprojectile bombardment with dsRNA or intron-containing hairpin RNA (ihpRNA)-expressing vectors; infiltration of plant tissue with an *Agrobacterium* strain carrying a T-DNA expressing an ihpRNA transgene; virus induced gene silencing (VIGS), in which the target sequence is integrated into viral sequences which are used to infect the plant, or are expressed from *Agrobacterium*-  
15 introduced transgenes, and by stable transformation with ihpRNA expressing transgenes. The various RNAi techniques each have advantages and disadvantages with respect to how persistent their effect is and the range of plants to which they can be applied, e.g. bombardment can be applied to any plant, but produces only transient effects. Alternatively, transformation with ihpRNA-expressing transgenes provides  
20 stable and heritable gene silencing, but requires efficient plant transformation techniques. ihpRNA transgenes have been shown to be very effective for a wide range of target genes in various plant species (reviewed in Waterhouse P, Helliwell C. 2003. Exploring plant genomes by RNA-induced gene silencing. *Nature Genet* 4: 29-38; Wesley S, Helliwell C, Smith N, *et al.* 2001. Construct design for efficient, effective  
25 and high-throughput gene silencing in plants. *Plant J* 27: 581-590), indicating that the RNAi mechanism is probably conserved in all plant species. This is supported by a recent report of RNAi in the non-vascular moss *Physcomitrella patens* (Bezanilla M, Pan A, Quatrano R. 2003. RNA interference in the moss *Physcomitrella patens*. *Plant Physiol* 133: 470-474).

30 Antisense genetic constructs for fiber specific promoters (e.g., for SEQ ID NO: 85, 91) can be used to inhibit or lessen the expression of one or more fiber genes in fiber cells. The use of antisense constructs is described in U.S. Pat. No. 5,495,070 and in Smith, *et al.* *Nature* 334 724-726, 1988; Bird, *et al.* *Bio/Technology* 9: 635-



639, 1991; Van der Krol, et al. Gene 72: 45-50, 1988.

It will be appreciated that the generation of fiber producing plant of desired traits according to the present invention can also be effected by crossing each of the above genetically modified plants with wild type, hybrid or transgenic plants, using  
5 methods which are well known in the art.

Once the transgenic plants of the present invention are generated, fibers are harvested (for example by mechanical picking and/or hand-stripping) and fiber yield and quality is determined.

The following describes methods of qualifying cotton fibers.

10 Fiber length - Instruments such as a fibrograph and HVI (high volume instrumentation) systems are used to measure the length of the fiber. HVI instruments compute length in terms of "mean" and "upper half mean" (UHM) length. The mean is the average length of all the fibers while UHM is the average length of the longer half of the fiber distribution.

15 Fiber strength - As mentioned, fiber strength is usually defined as the force required to break a bundle of fibers or a single fiber. In HVI testing the breaking force is converted to "grams force per tex unit." This is the force required to break a bundle of fibers that is one tex unit in size. In HVI testing the strength is given in grams per tex units (grams/tex). Fibers can be classified as low strength (e.g., 19-22 gms/tex),  
20 average strength (e.g., 23-25 gms/tex), high strength (e.g., 26-28 gms/tex), and very high strength (e.g., 29-36 gms/tex).

Micronaire - The micronaire reading of a fiber is obtained from a porous air flow test. The test is conducted as follows. A weighed sample of cotton is compressed to a given volume and controlled air flow is passed through the sample. The resistance  
25 to the air flow is read as micronaire units. The micronaire readings reflects a combination of maturity and fineness. Since the fiber diameter of fibers within a given variety of cotton is fairly consistent, the micronaire index will more likely indicate maturity variation rather than variations in fineness. A micronaire reading of 2.6-2.9 is low while 3.0-3.4 is below average, 3.5-4.9 is average and 5.0 and up are high. For  
30 most textile applications a micronaire of 3.5-4.9 is used. Anything higher than this is usually not desirable. It will be appreciated though, that different applications require different fiber properties. Thus, it is understood that a fiber property that is disadvantageous in one application might be advantageous in another.



As is illustrated in the Examples section, which follows, biomolecular sequences of the present invention are capable of increasing trichome/leaf hair number and length, as well as seed hair. As such biomolecular sequences of the present invention can be used to generate transgenic plants with increased trichome number/length which better deter herbivores, guide the path of pollinators, or affect photosynthesis, leaf temperature, or water loss through increased light reflectance. Additionally such transgenic plants may be used for the compartmentalized production of recombinant proteins and chemicals in trichomes, as described in details in WO 2004/111183 to Evogene Ltd.

Interestingly and unexpectedly, the present inventors found that polynucleotide sequences of the present invention are capable of increasing a biomass of a plant. It will be appreciated that the ability of the polypeptides of the present invention to increase plant yield/biomass/vigor is inherent to their ability to promote the increase in plant cell-size or volume (as described herein).

Thus, the present invention also envisages a method of increasing a biomass/vigor/yield of a plant (coniferous plants, moss, algae, monocot or dicot, as well as other plants listed in [www.nationmaster.com/encyclopedia/Plantae](http://www.nationmaster.com/encyclopedia/Plantae)). This is effected by regulating expression and/or activity of at least one of the polynucleotides of the present invention, as described above.

As used herein the phrase "plant biomass" refers to the amount or quantity of tissue produced from the plant in a growing season, which could also determine or affect the plant yield or the yield per growing area.

As used herein the phrase "plant vigor" refers to the amount or quantity of tissue produced from the plant in a given time. Hence increase vigor could determine or affect the plant yield or the yield per growing time or growing area.

As used herein the phrase "plant yield" refers to the amount or quantity of tissue produced and harvested as the plant produced product. Hence increase yield could affect the economic benefit one can obtain from the plant in a certain growing area and/or growing time.

Thus, the present invention is of high agricultural value for promoting the yield of commercially desired crops (e.g., biomass of vegetative organ such as poplar wood, or reproductive organ such as number of seeds or seed biomass).



As used herein the term "about" refers to  $\pm 10\%$ .

Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

10

### EXAMPLES

Reference is now made to the following examples, which together with the above descriptions, illustrate the invention in a non limiting fashion.

Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook et al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson et al., "Recombinant DNA", Scientific American Books, New York; Birren et al. (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I-III Cellis, J. E., ed. (1994); "Current Protocols in Immunology" Volumes I-III Coligan J. E., ed. (1994); Stites et al. (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific literature, see, for example, U.S. Pat. Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M.



J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., Eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and  
5 "Methods in Enzymology" Vol. 1-317, Academic Press; "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, CA (1990); Marshak et al., "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The  
10 procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

### EXAMPLE 1

#### 15 *In silico identification of cotton genes involved in fiber formation*

##### *Experimental Procedures*

*Interspecies comparison of expressed sequences*- Two main tools were used during the data mining stage. Large numbers of gene profiles were queried from an ORACLE database housing Compugen's GeneCarta platform (Compugen Ltd. Israel).  
20 This data was loaded into MicroSoft Excel spreadsheets for further manual refinement. Using this data a cross species genomic comparison was effected, aiming at defining organs from other plant species for which publically available EST libraries can be used both as models and as new sources of information to define new genes with key role in fiber formation (Figure 1). This comparison analysis used  
25 mainly the cotton, arabidopsis and tomato databases.

*Clustering and inter-species clustering of EST sequences* - The cotton genomic database included less than 50,000 ESTs (Genbank release #135) originating primarily from two species *Gossypium arboreum* (~ 35,000 ESTs) and *Gossypium hirsutum* L. (~ 9,000 ESTs, Table 1, below). These ESTs were clustered and  
30 assembled using the LEADS<sup>TM</sup> software platform (Compugen Ltd, Israel) in two alternative approaches.

In the first approach, the ESTs from two species were clustered and assembled together (thereby mimicking the evolutionary process since *G. arboreum* is an



ancestor of *G. hirsutum*). This process revealed 6478 clusters among them 3243 new clusters (without mRNA in the public database) that were defined as high quality clusters (Table 1, below).

In the second approach, ESTs from each species were clustered and assembled separately. Comparison between the two approaches showed that using the first approach adds valuable information to the cotton clusters without a significant bias in the analysis. The tomato genomic database contains 126,156 ESTs originating from about 30 well defined libraries that through the clustering and assembling process revealed 14034 clusters of which a large group of 12787 new high quality clusters (Table 1). The genomic data of arabidopsis includes 99417 ESTs (ftp://ftp.ncbi.nih.gov/genbank/), 8573 full length cDNA (Rikken and genbank mRNAs ftp://ftp.ncbi.nih.gov/genbank/) and the entire DNA sequence. Using the LEADS software 23,148 clusters and 6777 singeltones (Single ESTs which no other EST was clustered therewith) were revealed, all of which were supported by ESTs sequences, contrary to the public consortium (TAIR, www.arabidopsis.org/).

EST libraries from other plants and organs that share similar biological processes as cotton fiber were sought. Such ESTs are expected to serve as models and as new information sources for the identification of genes which are involved in the fiber development. To this end, a list of known genes that are suspected to be involved in fiber formation was generated. These genes originated from arabidopsis and were shown in various studies to have a key role in trichome formation (i.e., GL2, CPC, bHLH, TTG1, GL1, reviewed in Larkin J.C. et.al. 2003, Schellmann S. et al. 2002). Extensive comparative genomic analysis revealed that tomato genes, with high homology to cotton fiber genes and to arabidopsis trichome genes have a significant EST content in either leaf trichome and specific flower development libraries. Further analysis compared the genomic data of these three species – cotton, Arabidopsis and tomato (focusing on the tomato libraries mentioned above) as key parameters in the present database search (Figure 1).

**Table 1**  
**Genomic databases of Cotton, Tomato and Arabidopsis**

| Species            | EST Lib description | EST count | mRNA | After LEADS (clusters)   |
|--------------------|---------------------|-----------|------|--------------------------|
| <i>G. arboreum</i> | Fiber 6DPA          | 37,276    | 12   | 16,294 clusters on mixed |
| <i>G. hirsutum</i> | Fiber 7-10 DPA      | 7,944     | 236  |                          |



34

|                      |                       |         |        |                                           |
|----------------------|-----------------------|---------|--------|-------------------------------------------|
| <i>G. hirsutum</i>   | Flower ovule<br>1DPA  | 1,272   | 870    | production*                               |
| <i>L. esculentum</i> | All libraries         | 115,859 | 7      | 25,678 clusters<br>on mixed<br>production |
| <i>L. hirsutum</i>   | Trichome<br>libraries | 2,409   | 7      |                                           |
| <i>L. pennellii</i>  | Trichome<br>libraries | 2,723   | 24,450 |                                           |
| <i>A. thaliana</i>   | All libraries         | 160,698 | mRNA   | 25,678 clusters                           |

\*clusters derived from different species, cotton *G. arboreum* and *G. hirsutum*, tomato *L. esculentum*, *L. hirsutum* and *L. pennellii*

*In silico identification of cotton genes with a role in fiber development* To  
5 find whether tomato genomic data can be used as a relevant source of genomic data to  
study cotton fiber development an extensive genomic comparison was effected to  
identify both tomato and cotton genes that have high homology to key genes  
determining arabidopsis trichome development (e.g., GL2, CPC, bHLH, TTG1, GL1).

Homologous genes were identified in cotton and tomato. Because almost all  
10 cotton ESTs were produced from cotton fibers, it was impossible to do in-silico  
prediction of the expression profile of those genes. However, wide tissue sources used  
for the production of the tomato EST database enabled identification of tissues in  
which trichome specific genes are expressed.

In tomato it was revealed that both trichome and ovule ESTs are enriched in  
15 clusters representing trichome specific genes. Interestingly, it was found that cotton  
fibers are produced from ovule coat cells. As tomato seeds are covered with hairy like  
tissue, similarly to cotton seeds, it was postulated that those hairs are developmentally  
linked to trichome and cotton fiber formation.

In tomato ~1100 clusters were found to include at least one EST from  
20 trichome libraries. Among them about 1000 sequences included sequences also  
originating from tomato flower libraries (in which the ovule tissue is present).  
Comparing this group of genes to cotton data revealed ~2300 cotton genes with high  
homology to the tomato trichome genes. Mining the database using these two groups  
of genes together with other bioinformatic information [cross species homology, Gene  
25 Ontology (GO)] revealed 80 cotton clusters predicted to have a key role in fiber  
formation. Those genes were selected based on the following criteria:

Cotton clusters with at least 2 ESTs;

Homology to tomato cluster with e-score higher than 1e-5;



Homology to tomato cluster with at least one EST coming from trichome libraries or one EST coming from ovule containing tissues;

The following criteria were considered as advantageous although not necessary:

- 5        Large number of ESTs in a cluster;
- Transcription factor/ signal transduction proteins;
- Gene annotation related to cell expansion, turgor pressure, cell-wall synthesis.

The new genes together with the control cotton genes known to be involved in fiber formation were further analysed for their RNA expression profile in cotton  
10    plants.

## **EXAMPLE 2**

### ***mRNA expression analysis of genes identified according to the teachings of the present invention***

To study the RNA expression profile of candidate genes identified as described in Example 1 above, a reverse transcription was effected followed by real time PCR (RT-qPCR).

#### 15        ***Experimental Procedures***

***Quantitative Real time PCR analysis (qRT PCR)*** - To verify the levels of expression specificity and trait-association, Reverse Transcription following quantitative (Real-Time) PCR (RTqPCR) was effected. Total RNA was extracted at different stages of fiber development (from the day of anthesis till day 20 - post  
20    anthesis). To study the specificity of expression, RNA from other tissues of the cotton plants were collected and analysed for control expression (i.e., young leaves, young stems, mature stems, young roots, sepals, petals, and stamen). For this purpose, RNA was extracted from Cotton tissue using Hot Borate RNA Extraction protocol according to [www.eeob.iastate.edu/faculty/WendelJ/ultramicroRNA.html](http://www.eeob.iastate.edu/faculty/WendelJ/ultramicroRNA.html). Reverse  
25    transcription was effected using 1.5 µg total RNA, using 300 U Super Script II Reverse Transcriptase enzyme (Invitrogen), 225ng random deoxynucleotide hexamers (Invitrogen), 500 µM dNTPs mix (Takara, Japan), 0.2 volume of x 5 RT buffer (Invitrogen), 0.01M DTT, 60U RNAsin (Promega), DEPC treated double distilled water was added up to 37.5 µl. RT reactions were incubated for 50 min at 42 °C,



followed by 70 °C for 15 min. cDNA was diluted 1:20 in Tris EDTA, pH=8. 5mL of the diluted cDNA was used for qRT-PCR.

Quantitative RT-PCR was performed on cDNA (5 µL), using x1 SYBR GREEN PCR master mix (Applied Biosystems), forward and reverse primers 0.3 µM each. The ABI7000real-time PCR machine was used with the following conditions 50 °C for 2 min, 95 °C for 10 min, 40 times of 95 °C for 15 sec and 1 min at 60 °C, followed by 95 °C for 15 sec, 60 °C for 60 sec, and 70 times of 60 °C for 10 sec +0.5 °C increase in each cycle. For each gene, a standard curve was prepared from a pool of RTs from all samples, in 5 dilutions (dilutions – 1:60, 1:200, 1:600, 1:2000, 1:10000). The standard curve plot [ct (cycle threshold) vs. log (concentration)] should have  $R \geq 0.98$  with an efficiency in the range of 100% +/- 5%. The levels of expression (Qty) measured in the qPCR were calculated using the efficiency (E) of the amplification reaction and the corresponding C.T. (the cycle at which the samples crossed the threshold)  $Qty = E - C.T.$ . The dissociation curves obtained were inspected for the absence of unwanted additional PCR products or primer-dimers. Reactions were repeated at least twice. The calculation method is based in the fact that the efficiencies of the reactions of the GOI (gene of interest) and of the housekeeping genes are similar.

To normalize the expression level between the different tissues, specific primers were designed for specifically hybridizing with the following housekeeping genes: Actin (GenBank Accession No. D88414 SEQ ID NO: 28, Forward and reverse primers are set forth in SEQ ID NO: 68 and 69, respectively), GAPDH (GenBank Accession No. COTCWPPR, partial sequence, SEQ ID NO: 29, Forward and reverse primers are set forth in SEQ ID NO: 97 and 98, respectively), and RPL19 (GenBank Accession No. AI729179, SEQ ID NO: 30, Forward and reverse primers are set forth in SEQ ID NO: 99 and 100, respectively).

Using this methodology it was possible to identify genes that show elevated expression during fiber elongation, as well as genes that show unique cotton fiber specificity. Genes that showed elevated expression during anthesis that decreases during fiber elongation were considered good candidates to be involved in fiber differentiation and initiation. Notably, the above-described quantification methodology did not provide absolute expression levels, but provided good parameters for scoring the relative gene expression along fiber development as



differences as high as over 1000 fold in the maximal levels of expression reached by different genes were detected (Table 2, below).

### ***Results***

88 cotton genes were evaluated for expression profile in different tissues of  
5 cotton (*Gossypium hirsutum*, var Acala). According to the gene expression results, 23 genes were predicted to improve fiber yield and quality. Expression profile of all the candidate genes are presented in Table 2.



Table 2

| Gene ID/SEQ ID NO. | -DPA*<br>0.053* | 0-1 dpa | 12-14 dpa | 15-17 dpa | 18-20 dpa | 2-3 dpa | 4-5 dpa | 6-8 dpa | 9-11 dpa | mature leaves | mature stems | petals | sepals | stamen | young leaves | young roots | young stems |
|--------------------|-----------------|---------|-----------|-----------|-----------|---------|---------|---------|----------|---------------|--------------|--------|--------|--------|--------------|-------------|-------------|
| CT1/1              | 0.053*          | 0.049   | 2.034     | 2.138     | 2.477     | 0.295   | 0.976   | 1.347   | 1.118    | 0.53          | 0.029        | 9.368  | 0.336  | 0.277  | 0.347        | 0.002       | 0.202       |
| CT2/2              | 0.025           | 0.040   | 0.870     | 0.735     | 0.819     | 0.060   | 0.183   | 0.238   | 0.267    | 0.014         | 0.000        | 0.001  | 0.008  | 0.01   | 0.021        | 0.068       | 0.025       |
| CT3/3              | 0.082           | 0.070   | 0.511     | 0.632     | 0.819     | 0.057   | 0.084   | 0.116   | 0.092    | 0.109         | 0.032        | 0.038  | 0.086  | 0.020  | 0.142        | 0.037       | 0.063       |
| CT4/4              | 1.313           | 0.719   | 0.389     | 0.561     | 0.419     | 0.622   | 0.666   | 0.757   | 0.774    | 0.001         | 0.001        | 0.004  | 0.000  | 0.044  | 0.001        | 0.003       | 0.003       |
| CT6/5              | 0.093           | 0.075   | 0.580     | 0.732     | 0.916     | 0.066   | 0.095   | 0.104   | 0.110    | 0.113         | 0.028        | 0.037  | 0.085  | 0.026  | 0.148        | 0.037       | 0.044       |
| CT7/6              | 0.074           | 0.055   | 0.362     | 0.297     | 0.197     | 0.112   | 0.219   | 0.228   | 0.263    | 0.066         | 0.001        | 0.125  | 0.007  | 0.001  | 0.055        | 0.000       | 0.049       |
| CT9/7              | 0.276           | 0.980   | 1.166     | 0.715     | 0.960     | 0.980   | 1.265   | 1.103   | 2.095    | 0.012         | 0.000        | 0.019  | 0.032  | 0.004  | 0.008        | 0.000       | 0.012       |
| CT11/8             | 0.148           | 0.163   | 0.132     | 0.163     | 0.121     | 0.142   | 0.131   | 0.163   | 0.097    | 0.000         | 0.000        | 0.000  | 0.000  | 0.068  | 0.000        | 0.000       | 0.000       |
| CT20/9             | 0.074           | 0.035   | 0.021     | 0.013     | 0.016     | 0.045   | 0.042   | 0.032   | 0.033    | 0.051         | 0.051        | 0.459  | 0.076  | 0.572  | 0.037        | 0.069       | 0.067       |
| CT22/10            | 2.989           | 1.631   | 0.870     | 0.838     | 0.749     | 1.693   | 1.268   | 1.017   | 1.589    | 0.541         | 0.636        | 0.168  | 0.408  | 0.521  | 0.463        | 1.308       | 0.762       |
| CT26/11            | 0.022           | 0.001   | 0.017     | 0.001     | 0.018     | 0.017   | 0.028   | 0.039   | 0.017    | .             | .            | 0.006  | .      | 0.001  | .            | .           | 0.000       |
| CT27/12            | 0.010           | 0.009   | 0.009     | 0.009     | 0.010     | 0.008   | 0.005   | 0.005   | 0.003    | .             | 0.007        | 0.008  | 0.005  | 0.001  | 0.001        | 0.001       | 0.007       |
| CT40/16            | 0.016           | 0.016   | 0.014     | 0.023     | 0.024     | 0.012   | 0.013   | 0.016   | 0.017    | 0.007         | 0.000        | 0.002  | 0.022  | 0.005  | 0.005        | 0.001       | 0.004       |
| CT49/17            | 0.056           | 0.114   | 0.156     | 0.131     | 0.111     | 0.161   | 0.283   | 0.315   | 0.332    | 0.031         | 0.002        | 0.011  | 0.007  | 0.007  | 0.060        | 0.005       | 0.047       |
| CT70/18            | 1.406           | 2.247   | 8.460     | 7.782     | 10.709    | 2.152   | 5.313   | 7.361   | 4.796    | 1.065         | 0.492        | 9.976  | 0.671  | 1.207  | 1.904        | 1.177       | 1.294       |
| CT71/19            | 0.095           | 0.403   | 1.736     | 2.079     | 2.670     | 0.338   | 0.685   | 1.139   | 0.809    | 0.627         | 1.708        | 1.258  | 1.268  | 6.599  | 1.301        | 0.004       | 0.480       |
| CT74/20            | 2.971           | 2.555   | 3.474     | 4.398     | 5.859     | 3.135   | 4.301   | 4.272   | 6.983    | 0.017         | 0.002        | 0.203  | 0.015  | 0.136  | 0.030        | 0.003       | 0.464       |
| CT75/21            | 1.727           | 0.282   | 16.012    | 15.856    | 20.171    | 3.812   | 8.935   | .       | 20.295   | 4.473         | 3.644        | 83.72  | 6.317  | 28.659 | 8.534        | 0.872       | 2.759       |
| CT76/22            | 0.000           | 0.002   | 0.041     | 0.039     | 0.080     | 0.007   | 0.020   | 0.015   | 0.036    | 0.000         | 0.000        | 0.000  | 0.000  | 0.000  | .            | 0.000       | 0.000       |
| CT77/23            | 0.005           | 0.011   | 0.555     | 0.892     | 1.434     | 0.057   | 0.161   | 0.166   | 0.123    | 0.016         | 0.026        | 0.020  | 0.009  | .      | 0.023        | 0.001       | 0.003       |
| CT81/24            | 0.161           | 0.196   | 3.455     | 4.880     | 14.028    | 0.210   | 0.354   | 0.515   | 1.153    | 9.477         | 26.444       | 1.165  | 0.913  | 0.021  | 6.614        | 0.004       | 1.089       |
| CT82/25            | 0.024           | 0.022   | 0.005     | 0.004     | 0.006     | 0.018   | 0.016   | 0.014   | 0.011    | 0.053         | 0.034        | 0.017  | 0.045  | 0.036  | 0.004        | .           | 0.000       |
| CT84/27            | 0.007           | 0.005   | 0.136     | 0.167     | 0.371     | 0.004   | 0.014   | 0.027   | 0.031    | 0.036         | 0.346        | 0.034  | 0.196  | 0.101  | 0.061        | 0.071       | 0.035       |
| CT88/13            | 0.002           | 0.371   | 0.841     | 2.978     | 3.045     | 4.947   | 14.725  | 17.514  | 28.290   | 0.001         | 0.034        | 0.005  | 0.000  | .      | 0.005        | 0.004       | 0.007       |

Reverse-transcription following quantitative PCR was performed using real-time PCR, on tissues of either young or mature cotton (G. hirsutum var Acala) plants. Relative amounts of mRNA of each gene are presented in all examined tissues. dpa- days post anthesis, of ovule and fibers tissues (until 10 dpa) or only fiber tissue (after 10 dpa).



Two main criteria were used to select cotton genes as candidates that may be involved in fiber development according to their RNA profiling. Genes showing a high degree of fiber expression specificity and genes displaying expression level, which changes concomitantly with fiber development (Table 3, below).

5 Twenty three genes met these selection criteria:

CT-1 (SEQ ID NOs. 1 and 106), CT\_2 (SEQ ID NOs.2 and 107), CT\_3 (SEQ ID NOs. 3 and 108), CT\_4 (SEQ ID NOs. 4 and 109) CT\_6 (SEQ ID NOs. 5 and 110), CT\_7 (SEQ ID NOs. 6 and 111), CT\_9 (SEQ ID NOs. 7 and 112), CT\_11 (SEQ ID NOs. 8 and 113), CT\_20 (SEQ ID NOs. 9 and 114), CT\_22 (10 and 115), CT\_26  
10 (SEQ ID NOs. 11 and 116), CT\_27 (SEQ ID NOs. 12 and 117), CT\_40 (SEQ ID NOs. 16 and 118), CT\_49 (SEQ ID NOs. 17 and 119), CT\_70 (SEQ ID NOs. 18 and 120), CT\_71 (SEQ ID NOs. 19 and 121), CT\_74 (SEQ ID NOs.20 and 122), CT\_75 (SEQ ID NOs. 21 and 123), CT\_76 (SEQ ID NOs. 22 and 124), CT\_77 (SEQ ID NOs. 23 and 125), CT\_81 (SEQ ID NOs. 24 and 126), CT\_82 (SEQ ID NOs. 25 and  
15 95), CT\_84 (SEQ ID NOs. 27 and 96) and CT\_88 (SEQ ID NOs. 13 and 26).

CT-4, 22, 20, 27, 40, 82 (SEQ ID NOs: 4, 10, 9, 12, 16 and 25, respectively) were chosen mainly as candidate genes that may have a role in fiber initiation (Table 3) while CT 27 (SEQ ID NO: 12), which is a homologue gene to GL3, was also used as a control (in Figure 2d CT 22, SEQ ID NO: 10 is shown).

20 CT-1, 2, 3, 6, 7, 9, 49, 70, 71, 74, 75, 76, 77, 81, 84 (SEQ ID NOs. 1, 2, 3, 5, 6, 7, 17, 18, 19, 20, 21, 22, 23, 24 and 27, respectively, see Figures 2a, c) were predicted to be involved in the fiber elongation and quality (strength and finesse) according to their expression pattern (Table 3, Figure 2C CT 1 is shown).

CT11, 40, 74 and CT 26 (SEQ ID NOs. 8, 16, 20 and 11, respectively, see  
25 Figures 2a, b) which are homologous to Glabrous1 from Arabidopsis (GenBank Accession No. AB006078) are fiber specific genes that showed uniform and fiber-specific expression during all stages of fiber development (Table 3, in Figure 2B CT 11 is shown as an example). Expression profile of all the chosen genes are shown in Table 2, above.



Table 3

| CT #  | Gene annotation                            | Initiation | Fiber Quality<br>& Elongation | Stable and Specific Fiber<br>Expression | Fiber Specific | Biological Process                      |
|-------|--------------------------------------------|------------|-------------------------------|-----------------------------------------|----------------|-----------------------------------------|
| CT_2  | Acid sucrose-6-phosphate hydrolase         |            | v                             |                                         | Yes            | carbohydrate metabolism                 |
| CT_7  | Putative acyltransferase                   |            | v                             |                                         |                | unknown                                 |
| CT_9  | Hypothetical protein                       |            | v                             |                                         | Yes            | tRNA processing                         |
| CT_49 | Hypothetical protein                       |            | v                             |                                         |                | unknown                                 |
| CT_1  | GDSL-motif lipase/hydrolase-like protein   |            | v                             |                                         |                | unknown                                 |
| CT_3  | Putative mitochondrial protein             |            | v                             |                                         |                | unknown                                 |
| CT_6  | Aspartyl protease                          |            | v                             |                                         |                | proteolysis and peptidolysis            |
| CT_70 | Cysteine protease                          |            | v                             |                                         |                | water deprivation                       |
| CT_71 | Dehydration-responsive protein             |            | v                             |                                         |                | dessication                             |
| CT_75 | Lipid transfer protein, putative           |            | v                             |                                         |                |                                         |
| CT_76 | Putative receptor kinase                   |            | v                             |                                         | Yes            | protein amino acid phosphorylation      |
| CT_77 | Hypothetical protein                       |            | v                             |                                         | Yes            |                                         |
| CT_81 | APETAL2-like protein                       |            | v                             |                                         |                | cell wall organization and biogenesis   |
| CT_84 | Hypothetical protein                       |            | v                             |                                         |                | aromatic amino acid family biosynthesis |
| CT_4  | Cytochrome P450-like protein               | v          |                               |                                         | Yes            | electron transport                      |
| CT_20 | MYB-related protein homologue              | v          |                               |                                         |                | regulation of transcription             |
| CT_22 | Hypothetical protein                       | v          |                               |                                         |                | unknown                                 |
| CT_27 | bHLH transcription factor-like protein     | v          |                               |                                         |                | regulation of transcription             |
| CT_82 | MADS box protein-like                      | v          |                               |                                         |                | regulation of transcription             |
| CT_11 | Agamous-like MADS-box transcription factor |            |                               | v                                       | Yes            | regulation of transcription             |
| CT_26 | MYB-related protein homologue              |            |                               | v                                       | Yes            | cell fate commitment                    |
| CT_40 | Lipid-transfer protein 3 precursor (LTP 3) |            |                               | v                                       | Yes            | lipid transport                         |
| CT_74 | EN/SPM-like transposon protein             |            |                               | v                                       | Yes            | cell wall organization and biogenesis   |



The selected genes were over-expressed in transgenic arabidopsis and tomato, using the constitutive CaMV promoter of 35S (SEQ ID NO. 31). Transgenic plants were further evaluated for epidermal modifications, trichome density and length and seed hair yield (as further described hereinbelow).

5

### EXAMPLE 3

#### *Analysis of gene expression using publically available microarrays*

Further information about the expression of the selected genes (Example 2, above) was retrieved by statistical analysis of microarray data from arabidopsis. Essentially, the best homologs of the new candidate genes in arabidopsis were compared to a set of 77 microarrays experiment of different tissues of Arabidopsis (AtGenExpress databases, the Principal investigator for AFGN: Prof. Dr. Lutz Nover, Botanisches Institut, Molekulare Zellbiologie, FB Biologie und Informatik der J. W. Goethe Universität Frankfurt; Biozentrum N200 3OG, Marie-Curie-Strasse 9, 60439 Frankfurt am Main, [www.arabidopsis.org/info/expression/ATGenExpress.jsp](http://www.arabidopsis.org/info/expression/ATGenExpress.jsp)).

15

Polynucleotide sequences that were highly expressed in elongated cells or inflorescence meristems were selected for further analysis.

Table 4 below lists tissues which exhibit the highest levels of gene expression.

20

**Table 4**

|       | <i>Tissues with high expression</i>                      | <i>&lt; Fold change/<br/>specificity</i> | <i>Related to fiber</i>          |
|-------|----------------------------------------------------------|------------------------------------------|----------------------------------|
| CT_1  | Seed, siliques                                           | 10-20                                    | Elongated cells                  |
| CT_11 | carpels, flower, seed, siliques                          | Tissue specific                          | Flower specific                  |
| CT_2  | root, seedlin and sepals                                 | Tissue specific                          | Elongated cells,                 |
| CT_22 | carpels, flower, inflorescence, shoot                    | 4-10                                     | inflorescence                    |
| CT_4  | Petals, stamen                                           | >10                                      | Elongated cells,                 |
| CT49  | siliques                                                 | >2                                       | Elongated cells,                 |
| CT_7  | carpels, flower, inflorescence, petals, shoot, siliques, | 10-30                                    | inflorescence                    |
| CT_70 | flower, root, stamen                                     | Almost tissue specific                   |                                  |
| CT_76 | carpels, flower, inflorescence, shoot, siliques          | >2                                       | Elongated cells, & inflorescence |
| CT_77 | seeds, pollen, stemen, petals, sepals, siliques          | 10-50                                    | Elongated cells                  |
| CT_82 | inflorescence, shoot stem                                | 3-6                                      | inflorescence                    |
| CT_88 | petals, stamen                                           |                                          | Elongated cells                  |



**EXAMPLE 4*****Establishing a correlation between expression of candidate genes and fiber length***

In order to define correlations between the levels of RNA expression of the selected genes and fiber length, fibers from 4 different cotton lines were analyzed. These fibers were selected showing very good fiber quality and high lint index (Pima types, originating from other cotton species, namely *G. barbadense*) and different levels of quality and lint indexes from various *G. hirsutum* lines: good quality and high lint index (Acala type), medium lint index (Coker type) and poor quality and short lint index (Tamcot type).

***Experimental procedures***

**RNA extraction** - Fiber development stages, representing different fiber characteristic, at 5, 10 15 and 20 DPA were sampled and RNA was extracted as describe in Example 2.

**Fiber assessment** - Fiber length of the above lines was measured using fibrograph. The fibrograph system was used to compute length in terms of "Upper Half Mean" length. The upper half mean (UHM) is the average length of longer half of the fiber distribution. The fibrograph measures length in span lengths at a given percentage point ([www.cottoninc.com/ClassificationofCotton/?Pg=4#Length](http://www.cottoninc.com/ClassificationofCotton/?Pg=4#Length).)

**Results**

Four different cotton lines were grown in Rehovot, Israel during summer 2004, and their fiber length was measured. The fibers UHM values are summarized in Table 5, below:

25

**Table 5**

|           | <b>Length (UHM)</b> |
|-----------|---------------------|
| Pima S5   | $1.40 \pm 0$ a      |
| Acala     | $1.23 \pm 0.01$ b   |
| Coker 310 | $1.18 \pm 0.01$ c   |
| Tamcot    | $1.15 \pm 0.02$ c   |

Five genes were tested for correlation between gene expression and fiber length (presented for CT\_76 in Figure 3). The results are summarized in the Table 6 below:



Table 6

|       |           | Tissue Sampling Day (DPA) |                          |                                   |                          |                                   |                          |                                   |
|-------|-----------|---------------------------|--------------------------|-----------------------------------|--------------------------|-----------------------------------|--------------------------|-----------------------------------|
|       |           | 0                         | 5                        |                                   | 10                       |                                   | 15                       |                                   |
|       |           | Relative amounts of mRNA  | Relative amounts of mRNA | Relative expression Related to T0 | Relative amounts of mRNA | Relative expression Related to T0 | Relative amounts of mRNA | Relative expression Related to T0 |
| CT_1  | Tamcot    | 0.75                      | 2.99                     | 4.0                               | 4.71                     |                                   |                          |                                   |
|       | Coker 310 | 0.51                      | 4.80                     | 9.3                               | 7.56                     |                                   |                          |                                   |
|       | Acala     | 0.64                      | 5.08                     | 7.9                               | 8.01                     |                                   |                          |                                   |
| CT_2  | Tamcot    | 0.03                      | 0.19                     | 7.6                               | 8.17                     |                                   |                          |                                   |
|       | Coker 310 | 0.03                      | 0.35                     | 11.4                              | 15.04                    |                                   |                          |                                   |
|       | Acala     | 0.02                      | 0.36                     | 17.7                              | 15.28                    |                                   |                          |                                   |
|       | Pima S5   | 0.02                      | 0.41                     | 23.6                              | 17.58                    |                                   |                          |                                   |
| CT_40 | Tamcot    | 0.28                      |                          |                                   |                          |                                   | 0.47                     | 1.67                              |
|       | Coker 310 | 0.37                      |                          |                                   |                          |                                   | 0.46                     | 1.24                              |
|       | Acala     | 0.30                      |                          |                                   |                          |                                   | 0.67                     | 2.25                              |
|       | Pima S5   | 0.37                      |                          |                                   |                          |                                   | 1.03                     | 2.75                              |
| CT_76 | Tamcot    | 0.01                      | 0.03                     | 5.4                               | 0.01                     | 2.3                               | 0.00                     | 0.10                              |
|       | Coker 310 | 0.01                      | 0.08                     | 8.9                               | 0.04                     | 5.1                               | 0.00                     | 0.10                              |
|       | Acala     | 0.01                      | 0.12                     | 16.6                              | 0.06                     | 9.1                               | 0.00                     | 0.12                              |
|       | Pima S5   | 0.01                      | 0.13                     | 122.4                             | 0.18                     | 177.9                             | 0.12                     | 99.51                             |
| CT_81 | Tamcot    | 0.50                      | 1.33                     | 2.68                              | 5.03                     | 10.15                             | 1.11                     | 2.24                              |
|       | Coker 310 | 0.31                      | 2.64                     | 8.65                              | 4.51                     | 14.76                             | 0.84                     | 2.75                              |
|       | Acala     | 0.49                      | 4.38                     | 8.98                              | 6.36                     | 13.05                             | 3.65                     | 7.49                              |

Reverse-transcription following quantitative PCR was performed using real-time PCR, on tissues of 0, 5 10 and 15 DPA of cotton (*G. hirsutum* var Tamcot, Coker and Acala, and *G. barbadense* var Pima S5) plants. Relative amounts of mRNA and Relative expression related to T0 of each gene are presented in all examined tissues.

### EXAMPLE 5

#### Cloning of the selected genes in a binary vector under constitutive regulation and recombinant expression of the same

**ORF analysis** - Gene sequences of the present invention were analyzed for ORFs using Gene Runner software version 3.05 (Hasting Software, Inc: [www.generunner.com/](http://www.generunner.com/)). ORFs of each gene were compared to Genbank database, using Blast ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). By comparing to highest homologous ORFs, the position of the ATG initiation codon was determined. All the sequences described herein were shown to have a predicted full length ORF and to include the predicted ATG starting codon.

**Cloning into the pPI expression vector** - For cloning genes of the present invention, total RNAs from the various developmental stages of fiber producing cells was extracted, using Hot Borate RNA Extraction from Cotton Tissue according to [www.eeob.iastate.edu/faculty/WendelJ/rnaextraction.html](http://www.eeob.iastate.edu/faculty/WendelJ/rnaextraction.html). Complementary DNA (cDNA) molecules were produced from mRNA using M-MuLV reverse-transcriptase (RT) enzyme (Roche) and T<sub>16</sub>NN DNA primer, following protocol provided by the



manufacturer. cDNA amplification was done for 19 genes, out of the sequences above, namely CT clones number 1, 2, 3, 6, 7, 9, 11, 20, 22, 27, 40, 71, 74, 75, 76, 81, 82, 84 and 88, by PCR using PFU proof reading DNA polymerase enzyme (Promega [www.promega.com/pnotes/68/7381\\_07/7381\\_07.html](http://www.promega.com/pnotes/68/7381_07/7381_07.html)) following the protocol  
5 provided by the manufacturer. Primers for each gene were designed to span the full ORF. Additional restriction endonuclease sites were added to the 5' end of each primer to facilitate further cloning of the CTs to the binary vector (pPI). Table 7 below, lists the primers used for cloning each of the genes:



Table 7

| CT No | Forward<br>Primer/SEQ ID NO:       | Reverse<br>Primer/SEQ ID NO:        | upstream restriction<br>site | downstream restriction site |
|-------|------------------------------------|-------------------------------------|------------------------------|-----------------------------|
| CT_1  | ACCCGGGATGGATGGTTATTGTAGCAGAAGG/32 | GCCGAGCTCGAATCAAAATGAGGGCAATGCC/33  | SmaI                         | SacI                        |
| CT_2  | AATCTAGACAAAGTACAGAAGCTCAATTCCC/34 | TGATAATCATGTGGAAGCAACC/35           | XbaI                         |                             |
| CT_3  | CAGCCCGGTGATGGAACTGAGCATTTCAG/36   | CGTGAGCTCTGATTAGAGTTTCAAGTGCATG/37  | SmaI                         | SacI                        |
| CT_6  | TTTCCCGGTTGTTGTCATGGCTTCTCTGC/38   | ATGGAGCTCATATTTCATGGCCAAAACAC/39    | SmaI                         | SacI                        |
| CT_7  | G CACCCGGGAAAGGAATGGCAGGCGTC/40    | TTTCGATATCCACAGTACCCCTACTTCCATGC/41 | SmaI                         | EcoRV                       |
| CT_9  | TACCCGGGTACCATTAATCTACTACAGCTGC/42 | GAGAGCTCAACAGACAAAGACCAGACTGG/43    | SmaI                         | SacI                        |
| CT_11 | ACCCCGGGCAAGTGATCAAAAGAGAAATGG/44  | CATGAGCTCTTTCTCCAACTCCTCTACCC/45    | SmaI                         | SacI                        |
| CT_20 | CCCCCGGGTCCCCTATTGTCATGCCCTTTC/46  | TTGAGCTCACTCGATCTTACTCATCC/47       | SmaI                         | SacI                        |
| CT_22 | AGCCCGGGAGATAGAGAGATGGAGGTCC/48    | TCGAGCTCTGGGGCAACAATCATTTACC/49     | SmaI                         | SacI                        |
| CT_27 | TCCCGGGCATCTGATCTAATTTGTTGGTGG/50  | TTGGATATCGCACCTTATGACATGGGATC/51    | SmaI                         | EcoRV                       |
| CT_40 | TTCCCGGGTACAAAACATGGCTAGTTCCG/52   | TCGAGCTCATCAACCTCAGTGCACCTTG/53     | SmaI                         | SacI                        |
| CT_71 | TAGTCACTCCTGTTCTAGATGAAG/54        | CTGAGCTCCAGGATTTTACTTAGGGACCC/55    | XbaI                         | SacI                        |
| CT_74 | TACCCGGGCATACAGAGATGGAGAGGC/56     | ACGAGCTCAAAGGTGTTGCTTAGGTCC/57      | SmaI                         | SacI                        |
| CT_75 | AGCCCGGGAGAAAAGATGATGAAAAGGGG/58   | AAGATATCAAAATCCCATGCCAAAACCCC/59    | SmaI                         | EcoRV                       |
| CT_76 | AACCCGGGCGCAACTTAAAGAAAACC/60      | AAGAGCTCCTTTGTTGGCTTCTCAAG/61       | SmaI                         | SacI                        |
| CT_81 | GACCCGGGACTGTAAAAAAGCATAGG/62      | GCGAGCTCAGCTTAAGGATGATGGGGAG/63     | SmaI                         | SacI                        |
| CT_82 | ATCCCGGGGATGGTGAGAGGCAAAATTC/64    | ACGAGCTCTAGCAATGGCGATAACGTAC/65     | SmaI                         | SacI                        |
| CT_84 | ATCCCGGGTTCATGAAAAGGGTCTCG/66      | GTGAGCTCTATCGTCGTTGTCTTCAGC/67      | SmaI                         | SacI                        |



The resultant PCR blunt ended products, were purified using PCR Purification Kit (Qiagen, Germany), digested with the appropriate restriction endonucleases (Roche) and cloned into the pPI binary vector (Figure 4), while replacing the existing GUS reporter gene. pPI is a modified version of pBI101.3 (Clontech, Accession No. U12640). pPI was constructed by inserting a synthetic poly-(A) signal sequence, which originated from pGL3 Basic plasmid vector (Promega, Acc No U47295, where the synthetic poly-(A) signal sequence is located between base-pairs 4658-4811), into the HindIII restriction site of pBI101.3 (while reconstituting the HindIII site, downstream to the poly-(A) insert), to avoid the possibility of read-through effect of the upstream Nos-promoter. To replace the GUS gene with each one of the CT genes in the pPI binary vector, pPI was digested with the appropriate restriction enzymes [5' prime restriction enzyme is either SmaI or XbaI and 3' prime restriction enzyme is either SacI or EcoRV (Roche- using the protocol provided by the manufacturer)]. Open binary vector was purified using PCR Purification Kit (Qiagen, Germany). 5-75 ng of PCR product of each of the CT genes and 100 ng of open pPI plasmid vector were ligated in 10 µL ligation reaction volume using T4 DNA ligase enzyme (Roche), following the protocol provided by the manufacturer. Ligation products were introduced into *E. coli* cells.

**Recombinant expression in bacteria** - 60 µL of *E. coli*, strain DH5-α competent cells (about  $10^9$  cells/mL) were transformed using 1µl of ligation reaction mixture by electroporation, using a MicroPulser electroporator (Biorad), 0.2 cm cuvettes (Biorad) and EC-2 electroporation program (Biorad). *E. coli* cells were grown on 0.8 mL LB liquid medium at 37 °C for 1 hrs and 0.2 mL of the cell suspension were plated on LB-agar plates supplemented with the antibiotics kanamycin 50 mg/L (Sigma). Plates were then incubated at 37 °C for 16 hrs. Bacteria colonies were grown and expression was confirmed by PCR amplification using primers which were designed to span the inserted sequence in the binary vector. Primers used for DNA amplification of the inserts in the pPI binary vector were: 5'-GGTGGCTCCTACAAATGCCATC-3' (forward, SEQ ID NO. 70) and 5'-AAGTTGGGTAACGCCAGGGT-3' (reverse, SEQ ID NO. 71).

PCR products were separated on 1.5 % agarose gels and product sizes were estimated by comparing to DNA ladder (MBI Fermentas). PCR products with the



47

predicted size were sequenced using the same primers previously used for PCR amplification (See Table 7, above).

Additional primers, which were designed based on the sequence of each gene insert, were used to complete the sequencing of the full length ORF insert.

- 5 Sequencing of the inserted sequence was performed to verify that the clones were introduced in the right orientation, and to eliminate the possibility that sequence errors were included during PCR amplification. DNA sequences were determined using ABI 377 sequencer (Amersham Biosciences Inc).

10 Into each one of the 19 pPI binary constructs harboring the CT genes, the constitutive, Cauliflower Mosaic Virus 35S promoter was cloned.

Cauliflower Mosaic Virus 35S promoter sequence, originated from the pBI121 vector (Clontech, Accession No AF485783) was cloned by digesting the pBI121 vector with the restriction endonucleases HindIII and BamHI (Roche) and ligated into the binary constructs, digested with the same enzymes (SEQ ID NO. 31).

15

### EXAMPLE 6

#### *Agrobacterium transformation of binary plasmids harboring the genes of interest and expression in Arabidopsis and tomato plants*

Each of the nineteen binary constructs, comprising the 35S promoter upstream  
20 of each of the CTs genes was transformed into Arabidopsis or tomato plants via *Agrobacterium tumefaciens* transformation.

60  $\mu$ L of *Agrobacterium tumefaciens* GV301 or LB4404 competent cells (about  $10^9$  cells/mL) were transformed with 20 ng of binary plasmid via electroporation, using a MicroPulser electroporator (Biorad), 0.2 cm cuvettes (Biorad)  
25 and EC-2 electroporation program (Biorad).

*Agrobacterium* cells were grown on 0.8 mL LB liquid medium at 28 °C for 3 hrs and 0.2 mL of the cell suspension were plated on LB-agar plates supplemented with the antibiotics gentamycin 50 mg/L (for *Agrobacterium* strains GV301) or streptomycin 300 mg/L (for *Agrobacterium* strain LB4404) and kanamycin 50 mg/L  
30 (Sigma). Plates were then incubated at 28 °C for 48 hrs. *Agrobacterium* colonies were grown and PCR amplification was performed on *Agrobacterium* cells, using primers which were designed to span the inserted sequence in the binary vector.

Primers used for PCR amplification were: 5'-



GGTGGCTCCTACAAATGCCATC-3' (forward, SEQ ID NO. 70) and 5'-AAGTTGGGTAACGCCAGGGT-3' (reverse, SEQ ID NO. 71).

5 PCR products were separated on 1.5 % agarose gels and product sizes were determined by comparing to DNA ladder (MBI Fermentas). PCR products with the predicted size were sequenced using the primers which were used for the PCR amplification. Sequencing of the inserted sequence was performed to verify that the right clones were introduced into the *Agrobacterium* cells.

DNA sequencing was effected using ABI 377 sequencer (Amersham Biosciences Inc.).

10 ***Plant transformation and cultivation:***

***Transformation of Arabidopsis thaliana plants with putative cotton genes -***  
Arabidopsis thaliana Columbia plants (T0 plants) were transformed using the Floral Dip procedure described by Clough and Bent and by Desfeux et al., with minor modifications. Briefly, T0 Plants were sown in 250 ml pots filled with wet peat-based growth mix. The pots were covered with aluminum foil and a plastic dome, kept at 4 °C for 3–4 days, then uncovered and incubated in a growth chamber at 18–24 °C under 16/8 hr light/dark cycles. The T0 plants were ready for transformation six days prior to anthesis. Single colonies of *Agrobacterium* carrying the binary constructs, were cultured in LB medium supplemented with kanamycin (50 mg/L) and gentamycin (50 mg/L). The cultures were incubated at 28 °C for 48 hrs under vigorous shaking and then centrifuged at 4,000 rpm for 5 minutes. The pellets comprising *Agrobacterium* cells were re-suspended in a transformation medium containing half-strength (2.15 g/L) Murashig-Skoog (Duchefa); 0.044 µM benzylamino purine (Sigma); 112 µg/L B5 Gambourg vitamins (Sigma); 5 % sucrose; and 0.2 ml/L Silwet L-77 (OSI Specialists, CT) in double-distilled water, at pH of 5.7. Transformation of T0 plants was effected by inverting each plant into an *Agrobacterium* suspension, such that the above ground plant tissue was submerged for 3-5 seconds. Each inoculated T0 plant was immediately placed in a plastic tray, then covered with clear plastic dome to maintain humidity and was kept in the dark at room temperature for 18 hrs, to facilitate infection and transformation. Transformed (i.e., transgenic) plants were then uncovered and transferred to a greenhouse for recovery and maturation.

15  
20  
25  
30



The transgenic T0 plants were grown in the greenhouse for 3-5 weeks until siliques were brown and dry. Seeds were harvested from plants and kept at room temperature until sowing. For generating T1 transgenic plants harboring the genes, seeds collected from transgenic T0 plants were surface-sterilized by soaking in 70% ethanol for 1 minute, followed by soaking in 5% sodium hypochloride and 0.05% triton for 5 minutes. The surface-sterilized seeds were thoroughly washed in sterile distilled water then placed on culture plates containing half-strength Murashig-Skoog (Duchefa); 2 % sucrose; 0.8 % plant agar; 50 mM kanamycin; and 200 mM carbenicylin (Duchefa). The culture plates were incubated at 4 °C for 48 hours then transferred to a growth room at 25 °C for an additional week of incubation. Vital T1 Arabidopsis plants were transferred to a fresh culture plates for another week of incubation. Following incubation the T1 plants were removed from culture plates and planted in growth mix contained in 250 ml pots. The transgenic plants were allowed to grow in a greenhouse to maturity.

**Transformation of Micro-Tom tomato plants with putative cotton genes -** Tomato (*Lycopersicon esculentum*, var MicroTom) transformation and cultivation of transgenic plants was effected according to Curtis *et al.* 1995, and Meissner *et. al.* 2000.

20

### EXAMPLE 7

#### ***Growth of Arabidopsis transformed plants and phenotype characterizations***

T1 arabidopsis plants were grown as described above and phenotypes were characterized.

**PCR analysis of transgenic plants -** Arabidopsis T2 seeds were sown directly in growth mix contained in 250 ml pots. Positive transgenic plants were screen for kanamycin resistance in two weeks old leaves by PCR. Primers used for PCR amplification of the kanamycin were: 5'- CTATTCGGCTATGACTGGGC -3' (forward, SEQ ID NO. 72) and 5'- ATGTCCTGATAGCGGTCCGC -3' (reverse, SEQ ID NO. 73).

**Root performance -** In order to visualized root performance, T2 seeds were surface-sterilized by soaking in 70 % ethanol for 1 minute, followed by soaking in 5 % sodium hypochloride and 0.05 % triton for 5 minutes. The surface-sterilized seeds were thoroughly washed in sterile distilled water and then placed in culture plates



containing half-strength Murashig-Skoog (Duchefa); 2% sucrose; 0.8% plant agar; 50 mM kanamycin; and 200 mM carbenicylin (Duchefa). The culture plates were incubated at 4 °C for 48 hours then transferred to a growth room at 25 °C till reaching  
5 the right size for phenotypic characterization.



## Results

Table 8 – Analysis of *Arabidopsis* T2 plants carrying the putative cotton genes

| CT    | Putative Gene function                     | T generation | No of Independent plants | T2 Phenotype                                                                                                                                                                                                                                                                                                               |
|-------|--------------------------------------------|--------------|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CT_11 | Agamous-like MADS-box transcription factor | 2            | 5                        | Curled and narrow leaves, with long petioles, roots are longer and denser (Figures 5a-c)                                                                                                                                                                                                                                   |
| CT_9  | Hypothetical protein                       | 2            | 5                        | The rosette leaves and the inflorescent are longer and bigger compared to control. The roots are longer and denser.<br>The phenotype resembles the phenotype of <i>Arabidopsis</i> plants over expressing expansin as was characterized by Hyung-Taeg Cho and Daniel J. Cosgrove in PNAS u August 15, 2000. (Figures 5g-i) |
| CT_20 | MYB-related protein                        | 1            | 1                        | Small rankled and hairy leaves (Figures 5d and e)                                                                                                                                                                                                                                                                          |
| CT_40 | Lipid-transfer protein 3                   | 2            | 5                        | Longer and curlier leaves (Figure 5j)                                                                                                                                                                                                                                                                                      |
| CT_22 | Hypothetical protein                       |              |                          | Narrow leaves, with long petioles (Figures 5d and f)                                                                                                                                                                                                                                                                       |
| CT_81 | APETAL2-like protein                       | 1            | 1                        | The rosette leaves are almost double then wild type (Figures 5k and l)                                                                                                                                                                                                                                                     |
| CT_1  | hydrolase-like protein                     | 1            | 6                        | Narrow leaves, with long petioles (same as CT_22, not shown)                                                                                                                                                                                                                                                               |



***EXAMPLE 8******Growth of MicroTom transformed plants and phenotype characterizations******Experimental Procedures***

- 5        ***Transgenic tomato plants*** - Plant were transformed as described in Example 6, above. Following transformation, T1 MicroTom tomato plants were grown in mix contained in 1000 ml pots.



**Results**  
**Table 9 - Analyzing Micro-Tom tomato T1 and T2 plants and seeds caring the putative cotton genes**

| <i>CT</i>    | <i>Putative Gene function</i>    | <i>T generation</i> | <i>No of Independent plants</i> | <i>T1 seed hair length (wt 0.3mm)</i> | <i>T2 Phenotype</i>                                                                        |
|--------------|----------------------------------|---------------------|---------------------------------|---------------------------------------|--------------------------------------------------------------------------------------------|
| <i>CT20</i>  | MYB-related protein homologue    | I                   | 10                              | 0.366±0.006mm (Figures 6c-e)          | Small and wrinkled leaves, the trichome on the leaves are longer and denser. (Figure 6a-b) |
| <i>CT75</i>  | Lipid transfer protein, putative | I                   | 2                               | 0.347±0.019mm                         | Big inflorescent                                                                           |
| <i>CT_6</i>  | Aspartyl protease                | 1                   | 1                               | 0.343±0.019                           |                                                                                            |
| <i>CT_82</i> | MADS box protein-like            | 1                   | 3                               | 0.423±0.013mm (Figure 5f)             | Normal plants                                                                              |



### ***Discussion***

(Examples 1-8)

*In-silico identification of genes involved in cotton fiber development* - Little is known about the genetic control of cotton fiber initiation and elongation. Since both cotton fiber and Arabidopsis trichomes are developed from single epidermal cells they are assumed to share similar genetic regulation (Reviewed at Wagner G.J. *et. al.* 2004). In Arabidopsis, a large number of studies have revealed extensive information on the genetic mechanisms regulating trichome initiation and elongation. Several studies demonstrated the similarities between trichome and fiber by showing that cotton fiber specific promoters in arabidopsis and tobacco plants confer trichome specific expression (Kim and Triplett, 2001; Hsu *et. al.* 1999; Liu *et. al.* 2000, Wang *et al.* 2004). Most of the research that studies fiber development uses arabidopsis trichome as a model system to identify cotton genes in a small scale manner (Kim and Triplett, 2001; Wang *et al.* 2004).

In this study the present inventors have used tomato trichome and flower EST libraries as model systems to study cotton fiber development. Analysis of the EST libraries profile of the tomato homologous clusters to known arabidopsis trichome genes showed that tomato trichome and flower EST libraries significantly contributed to this set of clusters.

This result was confirmed while analyzing the EST libraries profile of the new cotton clusters that were selected by their RNA expression pattern as cotton fiber genes. 9 and 10 clusters contained ESTs which originated from the flower and trichome libraries respectively. Furthermore the group of tomato trichome clusters (trichome ESTs/total ESTs > 0.1) comprise large portion from the tomato genes that show high degree of homology to cotton (~ 50 %) even though their percentage in the total population is only ~5 %. It may indicate that both organ share common developmental processes. Even though there is a large group of studies about the genetic control of tomato fruit and trichome development no publications could be found to use these organs as a source of genomic data to study cotton fiber development. All of the 23 cotton genes were compared to unique EST data produced separately from embryo and suspensor of Scarlet Runner bean developing seeds ([www.mcdb.ucla.edu/Research/Goldberg/ests/intro-index.htm](http://www.mcdb.ucla.edu/Research/Goldberg/ests/intro-index.htm)). All sequences, except one, share high homologies with sequences originated from the suspensor, which is a



maternal tissue. This result supports the *in silico* results and identifies the role of these cotton clusters in fiber development, which originated from maternal cells as well.

***Identifying cotton genes with a role in fiber development through analysis of RNA expression profile*** - The differentiation/initiation phase is represented by gene expression at or before anthesis. The elongation phase mainly in hirsutum cultivars is represented by very fast growth rate mainly during 5 to 20 DPA. One pattern is represented by genes such as CT 1, 2, 3 expressed at their highest levels, slightly before and during the period of peak fiber expansion about 20 DPA. Another pattern of gene expression is displayed by the CT40, 11 or 70 which have the same expression level throughout all fiber development. Likewise, known genes encoding actin, endoxyloglucan transferase or Suc synthase also display unvarying RNA levels throughout fiber development (Shimizu et al., 1997).

Since the initiation occurs mainly before anthesis till 1 DPA it suggests that genes with a peak in expression during this time may have a role in fiber initiation. CT 4, 20, 22 and 11 have expression patterns that indicate their involvement at this stage.

One limitation of the current cotton EST database is the absence of ESTs that were extracted from flower at initiation stage (there is one library that was taken from ovary 1 DPA but of poor quality) most ESTs were taken only later on, between 6 to 10 DPA. This EST composition could explain why most of the chosen genes have expression pattern that indicate their association with the elongation stage.

***Role of the selected genes in fiber development, possible mechanisms*** - The 23 fiber-associated clusters could be classified into 6 functional categories according to their sequence homology to known proteins and enzymes (Table 3, above). The classification was made according to the GO consortium ([www.geneontology.org/](http://www.geneontology.org/)). The largest group comprises unique sequences without homology to any known protein. The rest of the clusters were classified according to categories known to be associated with fiber development. Two genes (Table 3, above) were classified into a cell fate commitment category: a new gene that belongs to the MYB transcription factor and a cotton homologous gene to GL3 that are known to be involved in trichome development in arabidopsis. The expression pattern of both genes and the phenotype of CT20 transgene both in arabidopsis and tomato T1 plants support their involvement mainly in the initiation phase.



Accumulative evidence link cotton MYB genes with fiber development (Suo. J. *et. al.* 2003, Cerdoni. M.L. *et. al.* 2003, Loguerico L.L. *et al* 1999). Over expression of a number of genes that work in the same pathway related to the initiation phase, could further induce initiation. Kirik *et al.* (2004) showed that by over-expressing two or three genes from the initiation phase they enhance the number of trichome and root hairs. Genes that relate to the initiation phase could be used for uniformity of fiber initiation on the cotton seed, initiate of more of the seeds epidermis cells into fibers. Over expression of those genes in vegetative meristems such as stems and leaves could be used as protect against insects (as has been shown in canola, [www.westerngrains.com/news/nr\\_050413.html](http://www.westerngrains.com/news/nr_050413.html)) and a-biotic stresses. However, there is no substantial evidence that proves direct involvement of any MYB gene to fiber development.

Two other genes (Table 3, above) are transcription factors from the MYB and MADS BOX families. Many studies demonstrated the function of these two transcription factor families as homeotic genes with key role in different developmental processes, among them are trichome and fiber morphogenesis (Suo. J. *et. al.* 2003, Ferrario S *et. al.* 2004). Their role in early stages of fiber development is supported also by their RNA expression pattern, which, is induced before, and during the day of anthesis. One gene (CT\_2, Table 3, above) was classified to the pathways of starch and sucrose metabolism. A recent work demonstrates that another gene (SUS), which, belongs to this pathway, is a limiting factor in both fiber initiation and development. CT\_40, 75 were classified as lipid transport whose RNA expression is highly induced during early fiber elongation stage fit to the fact that lipids are key components in fiber formation. Several genes (Table 3, above, CT\_4, 70, 71) were classified either as genes involved in desiccation, salinity response stimulated by abscisic acid and genes involved in electron transfer. Out of them 3 genes (CT 7, 9 and 49) were selected by RNA expression pattern to be induced in the elongation stage. Several studies consider changing proton and potassium pump mechanisms as key factor in the rapid growth rate of the fiber (Smart L.B, *et. al.* 1998). Combine the over-expression of several genes relate to fiber elongation such as genes relate to starch and sucrose metabolism that will enhance cell wall formation with lipid transport genes or genes relate to desiccation that my influence on the pressure in the cell, might result in longer fibers then over expressed of single gene.



### EXAMPLE 9

#### *Cloning and analyses of promoter sequences upstream of the genes of the present invention*

5 Differential gene expression in fiber tissues vs. other tissues in cotton is the result of complicated gene regulation. The genomic regions upstream of the 23 selected genes are predicted to possess promoter activities that direct gene expression to fiber cells in unique quantitative and qualitative manner. A precise gene expression, directed to fiber cells, is crucial for the development of cotton plants with enhanced  
10 fiber performance, without negatively affecting other plant tissues.

#### *Experimental Procedures*

*Cloning of promoter sequences* - The genomic sequence upstream of CT2 and CT6 were cloned from genomic DNA of cotton (*Gossypium hirsutum* L. var Acala), as follows. Total genomic DNA was extracted from plant leaf tissues of 4 week old  
15 cultivated cotton plants (*Gossypium hirsutum* L., var Acala), using DNA extraction kit (Dneasy plant mini kit, Qiagen, Germany). Inverse PCR (IPCR), DNA digestion, self-ligation, and PCR reaction were performed on genomic DNA, following common protocol ([www.pmci.unimelb.edu.au/core\\_facilities/manual/mb390.asp](http://www.pmci.unimelb.edu.au/core_facilities/manual/mb390.asp)) with the following modifications. To avoid mistakes in the IPCR, the genomic sequence of the  
20 5' sequence of a relevant cDNA (i.e. including introns) was first identified to produce Genomic Island (GI). The desired region from the genomic DNA was PCR-amplified using direct oligonucleotide primers designed based on the cDNA cluster sequence (for CT\_2 and CT\_6, respectively GI sequences are as set forth in SEQ ID NOs. 74 and 75 for CT\_2 and CT\_6. Primers are set forth in SEQ ID NOs. 14-15 (CT\_2) and  
25 101-102 CT\_6). PCR reaction was performed in a DNA thermal cycler, using common PCR protocols. For example:  
92 °C/3 min → 31 × [94 °C/30 sec → 56 °C/30 sec → 72 °C/3 min] → 72 °C/10 min).

PCR products were purified using PCR purification kit (Qiagen) and  
30 sequencing of the amplified PCR products was performed, using ABI 377 sequencer (Amersham Biosciences Inc).

In some cases, a different technique [UP-PCR (Dominguez and Lopez-Larrea. 1994)] was used when IPCR resulted in poor amplification. UP-PCR technique was



used in order to amplify unknown upstream region of known cluster sequences. Generally, the procedure involved four oligonucleotide primers: two sequence specific primers (SPs, external and internal) (listed below), both with same orientation of 3' end towards the unknown, yet desired, 5' region of the gene, and two universal walking primers (WP28 5'- TTTTTTTTTTTTGTGTTGTGTTGGGGGTGT (SEQ ID NO. 76 and sWP 5'- TTTTGTGTTGTGTTGTGGG, SEQ ID NO. 77). Reactions were carried out using the following reaction mixtures: sample mixture (SM) - genomic DNA of cotton species (30-40ng), WP28 primers (20 pmol), and double distilled water was added to a final volume of 10 µl. Polymerase mixture (PM) - dNTPs (Roche, Switzerland, 10nmol each), Expand Long Template Enzyme mix (Roche, Switzerland, 1U), 10 x buffer supplied with the enzyme and double distilled water was added to a final volume of 8 µl.

SMs were placed in a thermocycler (Biometra, USA), where it was subjected to an amplification program of 1 minute at 90 °C, held (pause) at 80 °C until PM was added, 30 seconds at 15 °C, 10 minutes at 25 °C, 3 minutes at 68 °C, held at 90 °C until the external SP (2 µl of 10 µM concentration) was added. The process was followed by external PCR reaction of 30 seconds at 92 °C, 10 seconds at 94 °C, 30 seconds at 65.5° C, 3 minutes at 68 °C, for 30 cycles followed by final extension of 10 minutes at 68 °C.

External PCR product diluted 5000 – 25000 fold was used as a template, and PCR amplification was effected using specific internal sWP and SP (30 pmol each) primers, 1U Ex Taq (Takara), in 50µl reaction volume. Internal PCR reaction was subjected to an amplification program of 2 minutes at 92 °C, followed by 30 seconds at 94 °C, 30 seconds at 58 °C, and 3 minutes at 72 °C for 30 cycles and a final extension of 10 minutes at 72 °C. IPCR / Up-PCR products were purified (PCR Purification Kit, Qiagen, Germany) and sequenced (ABI 377 sequencer, Amersham Biosciences Inc).

Primers for CT<sub>2</sub> were as follows (UP-PCR):

External primers:

sWP28- 5'- TTTTTTTTTTTTGTGTTGTGTTGGGGGTGT-3' (SEQ ID NO. 78)

SP (External)- 5'- CTGGGGTTACTTGCTAATGG -3' (SEQ ID NO: 79)

Internal (Nested) primers:

sWP- 5'- TTTTGTGTTGTGTTGTGGG -3' (SEQ ID NO: 80)



SP (Internal)- 5'- GCTCCGGGCTTTGGTTAACG -3' (SEQ ID NO: 81)

Internal genomic sequence of CT\_2 resulting from the above procedure is provided in SEQ ID NO: 14.

Primers for CT\_6 were as follows (UP-PCR):

5 External primers:

sWP28- 5'- TTTTTTTTTTTGTTTGTGTGGGGGTGT-3' (SEQ ID NO. 78)

SP (External)- 5'- GGCTTTGGGATGTTTGAGGTGG -3' (SEQ ID NO. 82)

Internal (Nested) primers:

sWP- 5'- TTTTGTGTTGTTGTGGG -3' (SEQ ID NO: 83)

10 SP (Internal)- 5'- GGTGGTGGGCTCTTGCAACAG -3' (SEQ ID NO: 84)

Internal genomic sequence of CT\_2 resulting from the above procedure is provided in SEQ ID NO: 85.

For cloning the putative promoters and 5' UTRs, PCR amplification was carried out using a new set of primers (below) to which 8-12 bp extension that  
15 included one restriction site (*Hind*III, *Sal*I, *Xba*I, *Bam*HI, or *Sma*I) on the 5' prime end. For each promoter, restriction sites that do not exist in the promoter sequence were selected. Moreover, the restriction sites in the primer sequences were design so the resultant PCR products will be cloned into the binary vector pPI in the right orientation, upstream of the GUS reporter gene.

20 The plasmid pPI was constructed by inserting a synthetic poly-(A) signal sequence, originating from pGL3 basic plasmid vector (Promega, Acc No U47295; bp 4658-4811) into the *Hind*III restriction site of the binary vector pBI101.3 (Clontech, Accession No. U12640).

Below are the primers used for promoter and 5' UTR (P+U) amplification and  
25 cloning into pPI, and the amplified and cloned sequence. Restriction sites within each primer are shown in bold letters:

CT\_2:

P+U forward (*Hind*III): 5'- ATTCAAGCTTTTGTGTTGTGGGGG -  
3' (SEQ ID NO: 86)

30 P+U reverse (*Bam*HI): 5'- TTGGATCCTTGGGCATTGAGCTTCTGTAC -  
3' (SEQ ID NO: 87)

P+U sequence of CT\_2 is as set forth in SEQ ID NO: 88.

CT6:



60

P+U forward (HindIII): 5'- TTAAAGCTTTGGGCTCTTGCAACAGAGGC -  
3' (SEQ ID NO: 89)

P+U reverse (BamHI): 5'- AAGGATCCGACGACGACAACAACAAC  
-3' (SEQ ID NO: 90)

5 P+U sequence of CT\_6 is as set forth in SEQ ID NO: 91.

Genomic DNA or the IPCR/UP-PCR product was used as DNA template for PCR-amplification, using the newly designed oligonucleotide primers. PCR products were purified (PCR Purification Kit, Qiagen, Germany) and digested with the restriction sites exist in the primers (Roche, Switzerland). The digested PCR products  
10 were re-purified and cloned into the binary vector pPI, which was digested with the same restriction enzymes. PCR product and the open plasmid vector were ligated using T4 DNA ligase enzyme (Roche, Switzerland).

### EXAMPLE 10

15 ***Transforming Agrobacterium tumefaciens cells with binary vectors harboring  
cotton fiber promoters***

pPi Binary vector, including either CT2 or CT6 promoter, upstream to the GUS reporter gene were used to transform *Agrobacterium* cells.

The binary vectors were introduced to *Agrobacterium tumefaciens* GV301, or  
20 LB4404 competent cells (about  $10^9$  cells/mL) by electroporation. Electroporation was performed using a MicroPulser electroporator (Biorad), 0.2 cm cuvettes (Biorad) and EC-2 electroporation program (Biorad). The treated cells were cultured in LB liquid medium at 28°C for 3 hr, then plated over LB agar supplemented with gentamycin (50 mg/L; for *Agrobacterium* strains GV301) or streptomycin (300 mg/L; for  
25 *Agrobacterium* strain LB4404) and kanamycin (50 mg/L) at 28°C for 48 hrs. *Agrobacterium* colonies which developed on the selective media were analyzed by PCR using the primers set forth in SEQ ID NOs: 70-71, which were designed to span the inserted sequence in the pPI plasmid. The resulting PCR products were isolated and sequenced as described in Example 4 above, to verify that the correct sequences  
30 were properly introduced to the *Agrobacterium* cells.

### EXAMPLE 11



***Cotton fiber specific promoters are expressed in tomato leaves and tomato fruits***

GUS staining was effected to illustrate specific expression in trichomes and tomato fruits.

***Experimental Procedures***

5 ***Transformation of Micro-Tom tomato plants with putative cotton promoter -***

As describe above.

***Transformation of Arabidopsis thaliana plants with putative cotton promoter -*** As describe above.

10 ***GUS staining of Arabidopsis-*** Gus staining of arabidopsis plants was effected as previously described (Jefferson RA. et. al. 1987, Meissner et. al. 2000).

***GUS staining of tomato leaves -*** Gus staining of tomato plants was effected as previously described (Jefferson RA. et. al. 1987, Meissner et. al. 2000).

Tissue fixation was effected as follows. Tomato leaves were immersed in 90 % ice cold acetone, then incubated on ice for 15 - 20 minutes following by removal of  
15 the acetone. Thereafter tissue was rinsed twice with the Working Solution [100 mM Sodium Phosphate (Sigma, USA) buffer pH=7, Ferricyanide (Sigma, USA) 5 mM, Ferrocyanide (Sigma, USA) 5 mM, EDTA (BioLab) pH=8 1 mM , Triton X-100 (Sigma, USA) 1 %] for 15-20 minutes in dark. Rinsing solution was then removed and replaced with X-gluc staining solution [Working Solution + 5-bromo-4-chloro-3-  
20 indolyl- $\beta$ -D-glucuronic acid (X-GlcA, Duchefa) solubilized in N,N-Dimethylformamide (BioLab) 0.75mg/ml , Dithiothreitol (BioLab) 100mM] and incubated for over night at 37 °C in the dark (tubes wrapped with aluminum foil). Distaining was effected by sinking the plant tissue in 70 % ethanol and heating at 50 °C for ~120 minutes. Distaining step was repeated until the plant tissue became  
25 transparent excluding the blue stained regions. Distained plants were stored in 70 % ethanol (BioLab) at room temperature.

***GAS staining of Tomato Fruits -*** Gus staining of tomato fruits was effected as previously described (Jefferson RA. et. al. 1987, Meissner et. al. 2000). Briefly: thin tomato fruit slice were sunk in staining solution [100 mM Sodium Phosphate (Sigma,  
30 USA) buffer pH=8, Ferricyanide (Sigma, USA) 5 mM, Ferrocyanide (Sigma, USA) 5 mM, EDTA (BioLab) pH=8 15mM , Methanol (BioLab) 20 % , 5-bromo-4-chloro-3-indolyl- $\beta$ -D-glucuronic acid (X-GlcA, Duchefa) solubilized in N,N-Dimethylformamide (BioLab) 0.75mg/ml] in the dark (tubes wrapped with aluminum



foil) and incubated for over night at 37 °C. Distaining was effected by sinking the plant tissue in 70 % ethanol and heating to 50 °C for ~20 minutes. Distaining step was repeated until the fruit slice became transparent except for the blue stained regions. Distained fruits were stored in 70 % ethanol (BioLab) at room temperature.

## 5 **Results**

### ***GUS staining was performed on Seeds of T1 tomato plants.***

GUS was expressed under the regulation of CT2 and CT6, promoters in the genetically transformed tomato plants (Figures 7a-b).

Results for tomato T1 generation are summarized in the Table 10, below.

10

**Table 10**

| Promoter | No of Independent T1 plants | Leaf | Leaf trichome | Seed cover of Young fruit | Seed cover of Mature green | Seed cover of Ripen fruit |
|----------|-----------------------------|------|---------------|---------------------------|----------------------------|---------------------------|
| CT2      | four                        | 0    | 2             | 3                         | 5                          | 3                         |
| CT6      | one                         | 0    | 1             | 1                         | 2.5                        | 1                         |

The numbers represent average grade, 0-not expressed, 5-high expression

## **EXAMPLE 12**

### ***Tomato seed hairs as a model system for cotton fibers***

15 The genetic modification of cotton is long and time consuming. Hence to find genes which are capable of improving cotton fiber yield and quality, a need exists for a model system for cotton fiber development in other plants.

Trichome cells and root hairs share common characteristics with cotton fiber cells, and are widely accepted as model systems for cotton fiber development  
20 [Reviewed in Wagner. G.J. et. al. 2004) and Wang et al. 2004].

However measuring changes in growth rate, length and thickness as well as other structural parameters is not an easy task because of the small size, remote accessibility and lack of uniformity in sizes of trichome cells.

To overcome these limitations, tomato seed hairs were analyzed for their  
25 possible use as a model tissue for cotton fiber development. To this end, the GUS reporter gene was over-expressed under the regulation of cotton fiber specific promoter element derived from CT2, as describe above.

Tomato transformation of the binary construct, plant regeneration and GUS staining was effected as described above.



Tomato seed hairs (Figure 8a) are maternal epidermal cells, covering the ovule surface of the seeds. In anatomical aspects, tomato seed hairs are much closer to cotton fibers than either trichome cells or root hairs.

4 independent transgenic tomato fruits over-expressing GUS gene under cotton specific promoter CT\_2 were produced. GUS staining of fruits at the mature-green stage (fruit is in full size just before the ripening process) was observed uniquely on the seed envelope, where seed hairs are being developed (Figures 7a and b).

Five independent transgenic tomato fruits over-expressing 35S-expansin (AF043284) were produced, and the seed hair length was measured and compare to wt. The seed hair of transgenic plants was significantly longer than of wt (Figures 8a-b).

*Table 11*

| Plant        | Number of Independent plant | Seed hair length (mm)   |
|--------------|-----------------------------|-------------------------|
| WT           | 3                           | 0.300±0.019             |
| 35S:expansin | 5                           | 0.357±0.017 (Figure 8b) |

15

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination.

20

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications, patents and patent applications and GenBank Accession numbers mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent or patent application or GenBank Accession number was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be

30



construed as an admission that such reference is available as prior art to the present invention.



**REFERENCES CITED BY AUTHOR NAME IN THE APPLICATION***(other references are cited in the document)*

5 Cedroni M.L, Cronn R.C, Adams K.L, Wilkins T.A, and Wendel J.F. 2003. Evolution and expression of MYB genes in diploid and polyploid cotton. *Plant Mol. Biol.* 51, 313-25.

10 Clough S.J, and Bent A.F (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735-43.

Curtis I.S, Davey M.R, and Power J.B. 1995. Leaf disk transformation. *Methods Mol. Biol.* 44, 59-70.

15 Desfeux C, Clough S.J, and Bent A.F (2000). Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. *Plant Physiol.* 123, 895-904.

Dominguez O, and Lopez-Larrea. C. 1994. Gene walking by unpredictably primed PCR. *Nucleic Acids Research.* 22:3247-3248.

20 Hsu C.Y, Creech R.G, Jenkins J.N, and Ma D.P. 1999. Analysis of promoter activity of cotton lipid transfer protein gene LPT6 in transgenic tobacco plants. *Plant Sci* 143, 63-70.

25 Kim H.J, and Triplett B.A. 2001. Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 2001 Dec;127(4):1361-6.

30 Larkin J.C, Brown M.L, and Schiefelbein J. 2003. How do cells know what they want to be when they grow up? Lessons from epidermal patterning in *Arabidopsis*. *Ann. Rev. Plant Mol. Biol.* 54, 403-430.



- Liu H.C, Creech R.G, Jenkins J.N, Ma D.P. 2000. Cloning and promoter analysis of the cotton lipid transfer protein gene Ltp3(1). *Biochim Biophys Acta.* 24):106-11
- Loguerico L.L, Zhang J.Q, and Wilkins T.A. 1999. Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.). *Mol. Gen. Genet.* 261, 660-71.
- Meissner R, Chague V, Zhu Q, Emmanuel E, Elkind Y, Levy A.A. 2000. Technical advance: a high throughput system for transposon tagging and promoter trapping in tomato. *Plant J.* 22, 265-74.
- Ruan Y.L, Llewellyn D.J, and Furbank R.T. 2003. Suppression of Sucrose Synthase gene expression represses cotton fiber cell initiation, elongation and seed development. *Plant Cell* 15, 952-964.
- Schellmann. S, Schnittger. A, Kirik. V, Wada. T, Okada. K, Beermann. A, Thumfahrt. J, Jurgens. G, and Hulskamp.M. 2002. *TRIPTYCHON* and *CAPRICE* mediate lateral inhibition during trichome and root hair patterning in Arabidopsis. *EMBO J.* 21, 5036-5046.
- Smart L.B, Vojdani F, Maeshima M, Wilkins T.A. 1998. Genes involved in osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated. *Plant Physiol.* 116, 1539-49.
- Suo. J, Liang. X, Pu. Li, Zhang. Y, and Xue. Y. 2003. Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). *Biochem. Biophys. Acta.* 1630, 25-34.
- Wagner. G.J, Wang. E and Shepherd. R.W. 2004. New approaches for studying and exploiting an old protuberance, the plant trichome. *Ann. Bot.* 93, 3-11.



67

Wang E, Gan S, and Wagner G.J. 2002. Isolation and characterization of the CYP71D16 trichome-specific promoter from *Nicotiana tabacum* L. *J Exp Bot.* 53(376):1891-7.



## DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME           1    DE    2  
CONTENANT LES PAGES    1    À    67

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

## JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME           1    OF    2  
CONTAINING PAGES    1    TO    67

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:



## WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleic acid sequence encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96, wherein said polypeptide is capable of regulating cotton fiber development.

2. The isolated polynucleotide of claim 1, wherein said nucleic acid sequence is selected from the group consisting of SEQ ID NOs. 1, 2, 4, 5, 7, 9, 10, 16, 17, 20, 21, 22, 24, 25, 27 and 13.

3. The isolated polynucleotide of claim 1, wherein said polypeptide is as set forth in SEQ ID NO. 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96.

4. The isolated polynucleotide of claim 1, wherein said amino acid sequence is as set forth in SEQ ID NO. 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96.

5. The isolated polynucleotide of claim 1, wherein said cotton fiber development comprises fiber formation.

6. The isolated polynucleotide of claim 1, wherein said cotton fiber development comprises fiber elongation.

7. An isolated polynucleotide comprising a nucleic acid sequence at least 80 % identical to SEQ ID NO: 85 or 91, wherein said nucleic acid sequence is capable of regulating expression of at least one polynucleotide sequence operably linked thereto in an ovule endothelial cell.

8. The isolated polynucleotide of claim 7, wherein said ovule endothelial cell is of a plant fiber or a trichome.



9. An oligonucleotide capable of specifically hybridizing to the isolated polynucleotide of claim 1 or 7.

10. A nucleic acid construct comprising the isolated polynucleotide of claim 1.

11. A nucleic acid construct comprising the isolated polynucleotide of claim 7.

12. The nucleic acid construct of claim 10, wherein the nucleic acid construct further comprising at least one cis-acting regulatory element operably linked to the isolated polynucleotide.

13. The nucleic acid construct of claim 7, wherein said polynucleotide sequence is selected from the group consisting of SEQ ID NOs: 1, 2, 4, 5, 7, 9, 10, 16, 17, 20, 21, 22, 24, 25, 27 and 13.

14. The nucleic acid construct of claim 12, wherein said cis-acting regulatory element is as set forth in SEQ ID NO: 74, 75, 85 or 91 or functional equivalents thereof.

15. A transgenic cell comprising the nucleic acid construct of claim 10 and/or 11.

16. A transgenic plant comprising the nucleic acid construct of claim 10 and/or 11.

17. A method of improving fiber quality and/or yield of a fiber producing plant, the method comprising regulating an expression level or activity of at least one polynucleotide encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96 in the fiber producing plant, thereby improving the quality and/or yield of the fiber producing plant.



18. The method of claim 17, wherein the quality of the fiber producing plant comprises at least one parameter selected from the group consisting of fiber length, fiber strength, fiber weight per unit length, maturity ratio, uniformity and micronaire.

19. The method of claim 17, wherein said regulating expression or activity of said at least one polynucleotide is up-regulating.

20. The method of claim 19, wherein said up-regulating is effected by introducing into the cotton the nucleic acid construct of claim 10 and/or 11.

21. The method of claim 17, wherein said regulating expression or activity of said at least one polynucleotide is down-regulating.

22. The method of claim 21, wherein said down-regulating is effected by gene silencing.

23. The method of claim 22, wherein said gene silencing is effected by introducing into the cotton the oligonucleotide of claim 9.

24. The method of claim 17, wherein said fiber producing plant is selected from the group consisting of cotton, silk cotton tree (Kapok, *Ceiba pentandra*), desert willow, creosote bush, winterfat, balsa, ramie, kenaf, hemp, roselle, jute, sisal abaca and flax.

25. A method of increasing a biomass of a plant, the method comprising regulating an expression level or activity of at least one polynucleotide encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96 in the plant, thereby increasing the biomass of the plant.

26. The method of claim 25, wherein the plant is a monocot plant.



27. The method of claim 25, wherein the plant is a dicot plant.

28. A method of identifying genes which are involved in cotton fiber development, the method comprising:

- (a) providing expressed nucleic acid sequences derived from cotton fibers;
- (b) providing expressed nucleic acid sequences derived from an ovule tissue;
- (c) computationally assembling said expressed nucleic acid sequences of (a) and (b) to generate clusters; and
- (d) identifying clusters of said clusters which comprise expressed nucleic acid sequences of (a) and (b), thereby identifying genes which are involved in cotton fiber development.

29. The method of claim 28, further comprising identifying genes which are differentially expressed in said cotton fiber following (d).

30. The method of claim 29, wherein said differentially expressed comprises:

- (a) specific expression; and/or
- (b) change in expression over fiber development.

31. A method of producing an insect resistant plant, comprising regulating an expression level or activity of at least one polynucleotide encoding a polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96 in a trichome of the plant, thereby producing the insect resistant plant.

32. A method of producing cotton fibers, the method comprising:

- (a) generating a transgenic cotton plant expressing at least one polypeptide having an amino acid sequence at least 80 % homologous to SEQ ID NO: 26, 106, 107, 109, 110, 112, 114, 115, 118, 119, 122, 123, 124, 126, 95 or 96; and



72

- (b) harvesting the fibers of said transgenic cotton plant, thereby producing the cotton fibers.



1/7

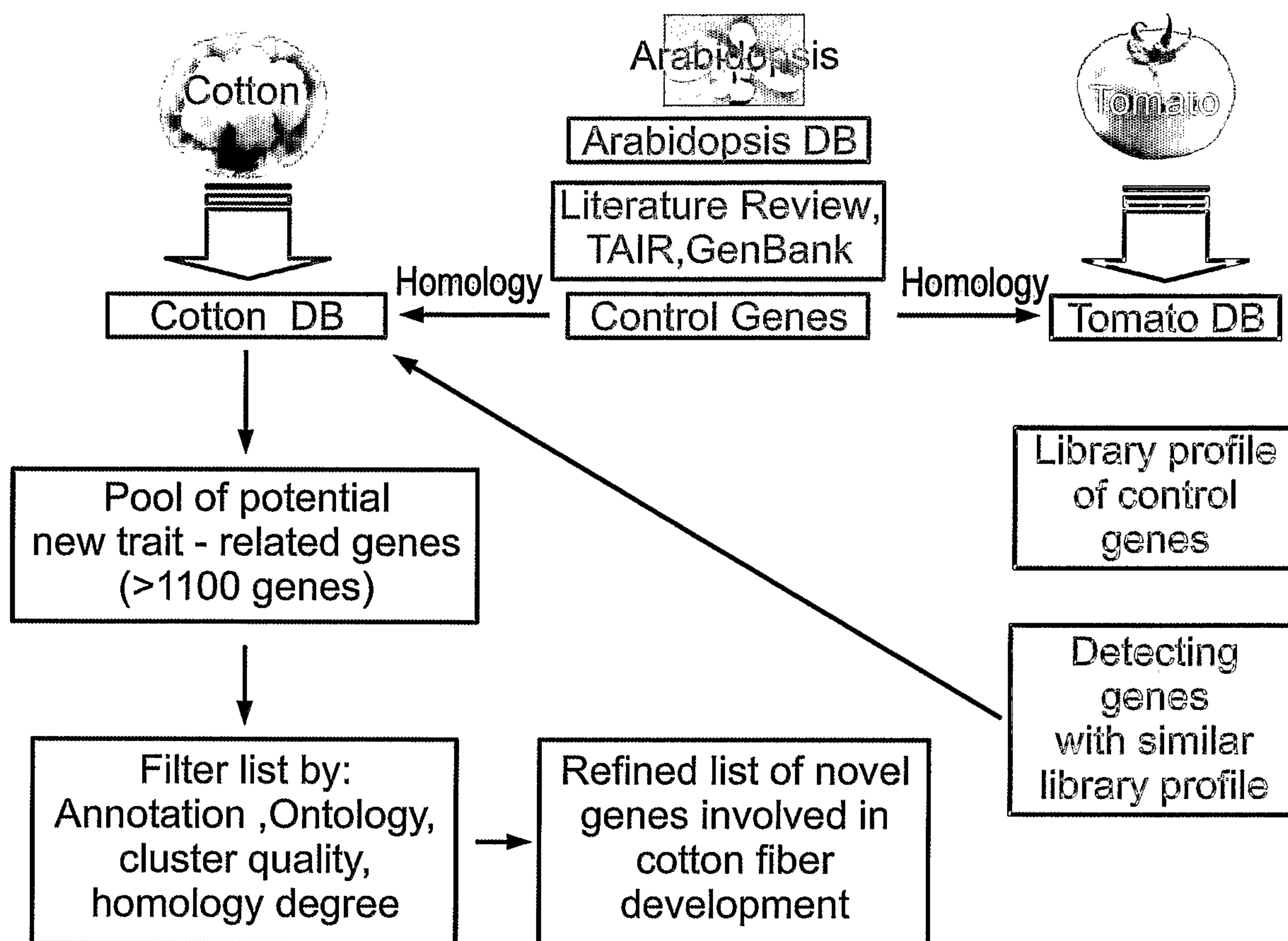
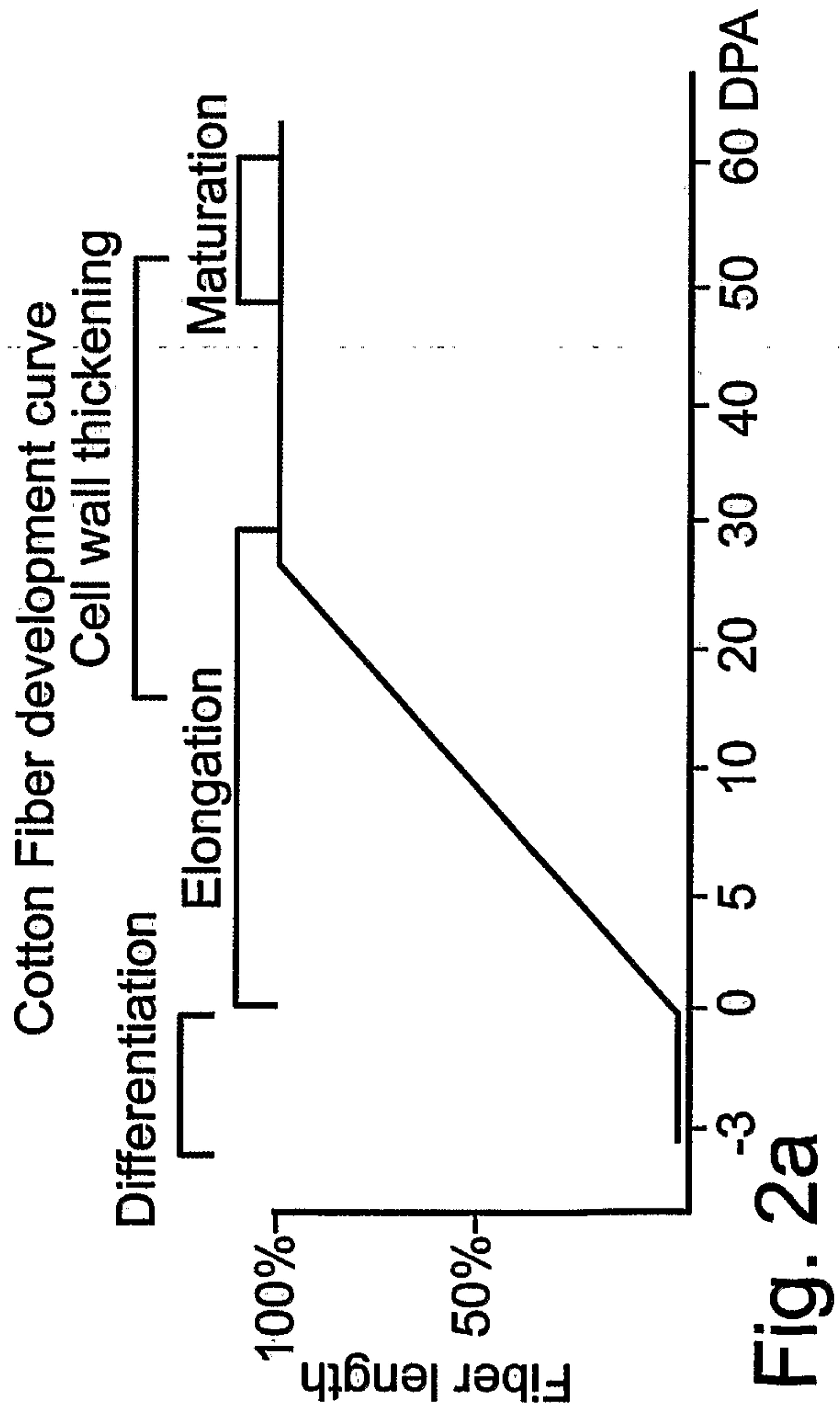
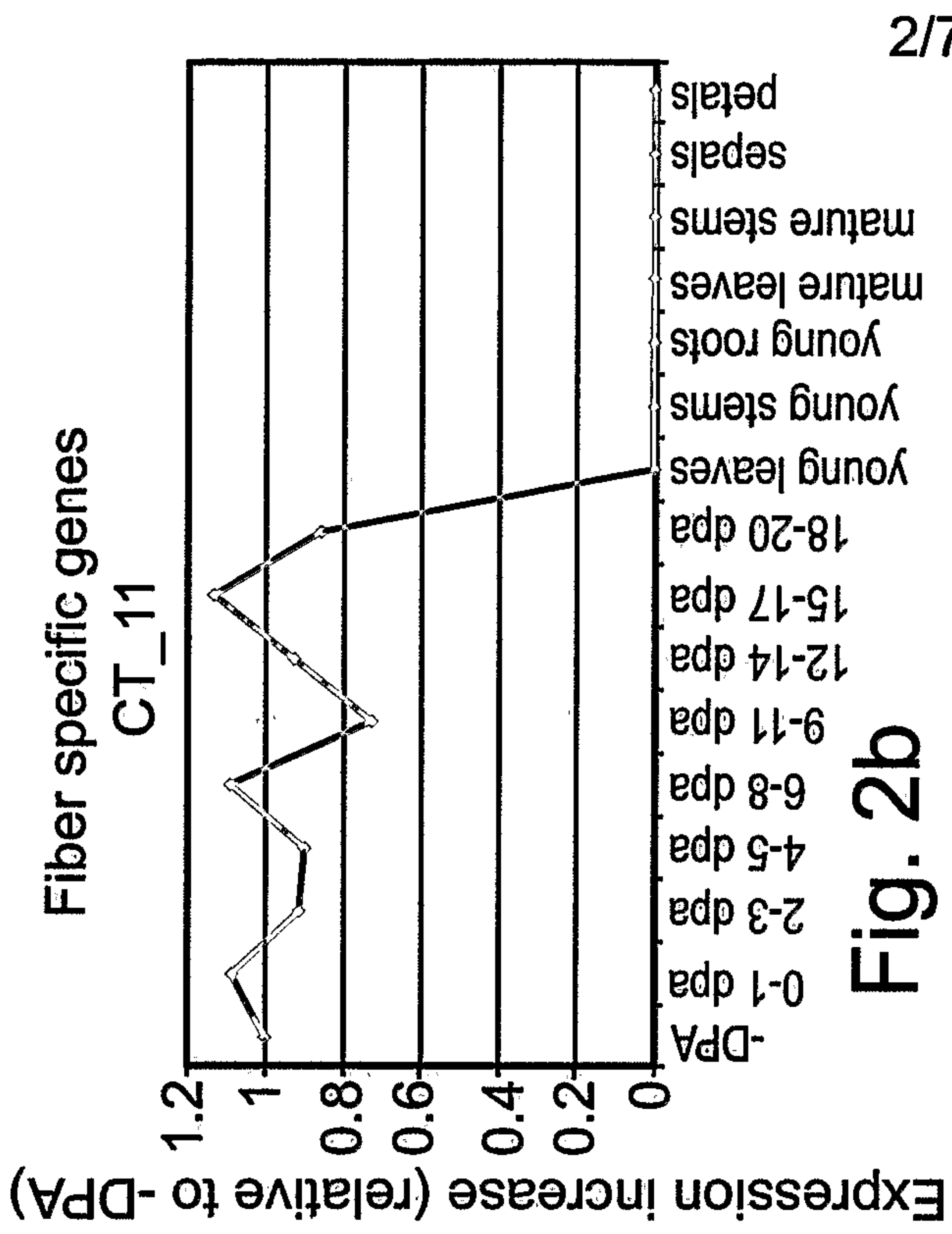


Fig. 1

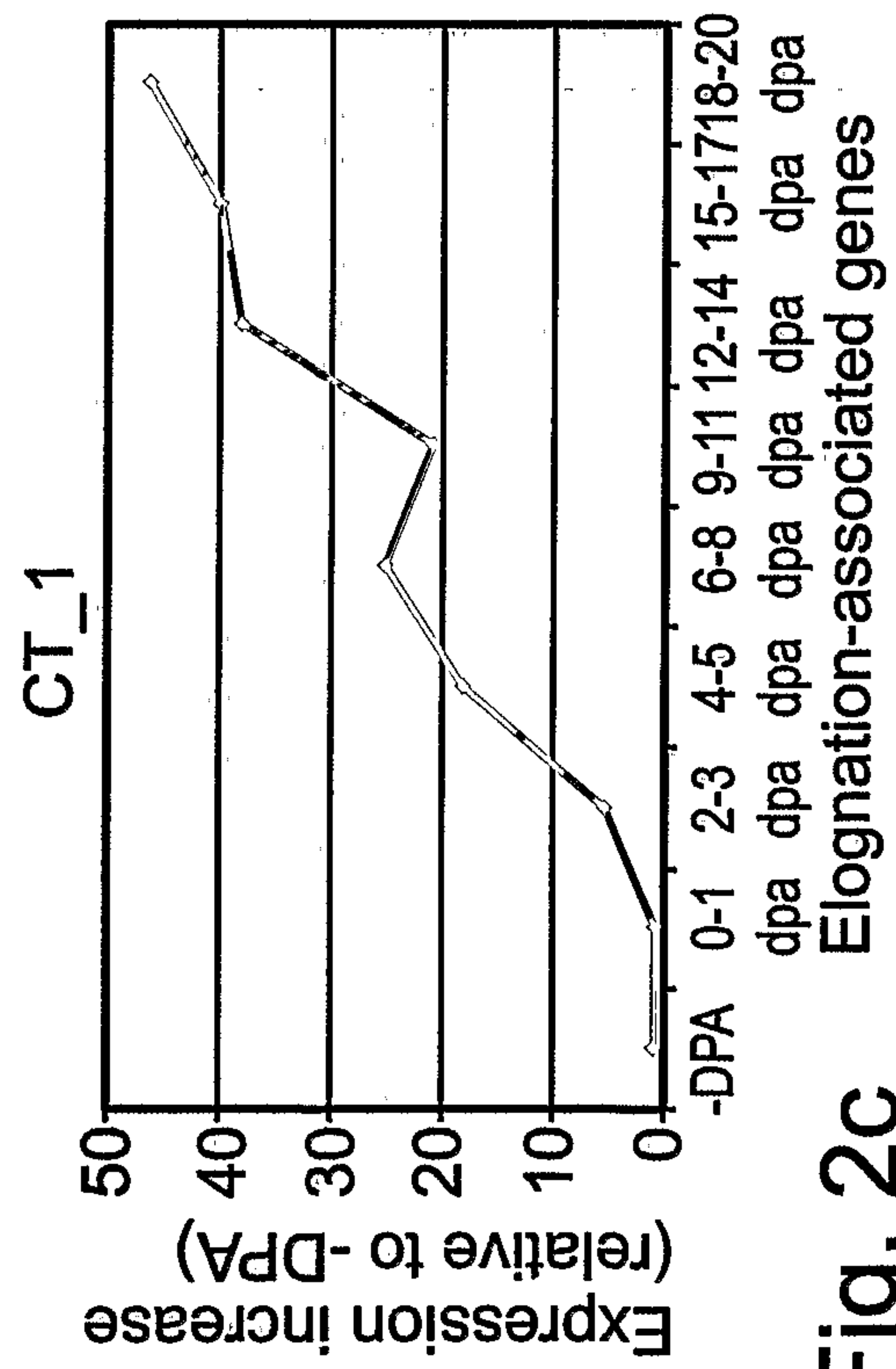




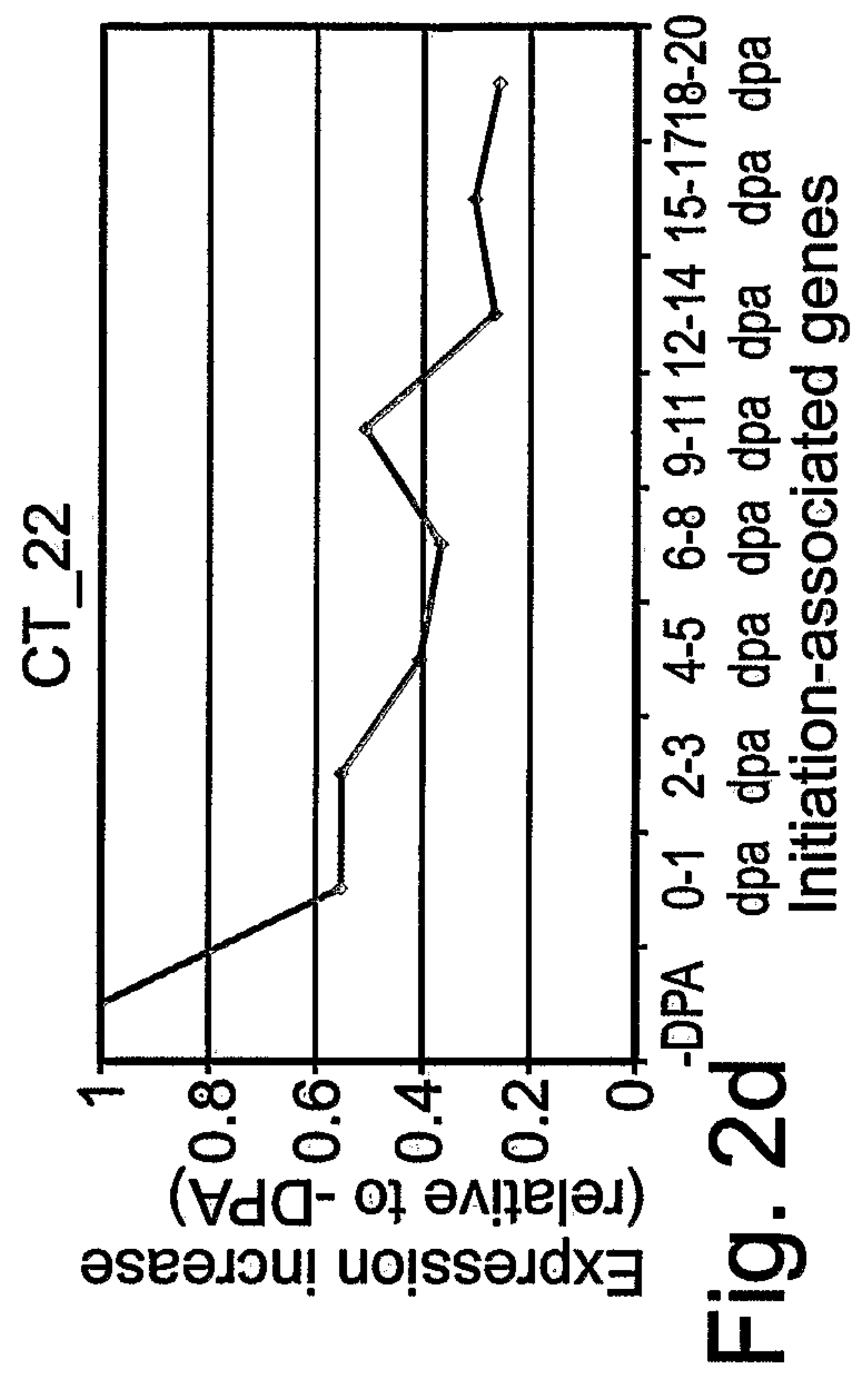
**Fig. 2a**



2/7



**Fig. 2c**





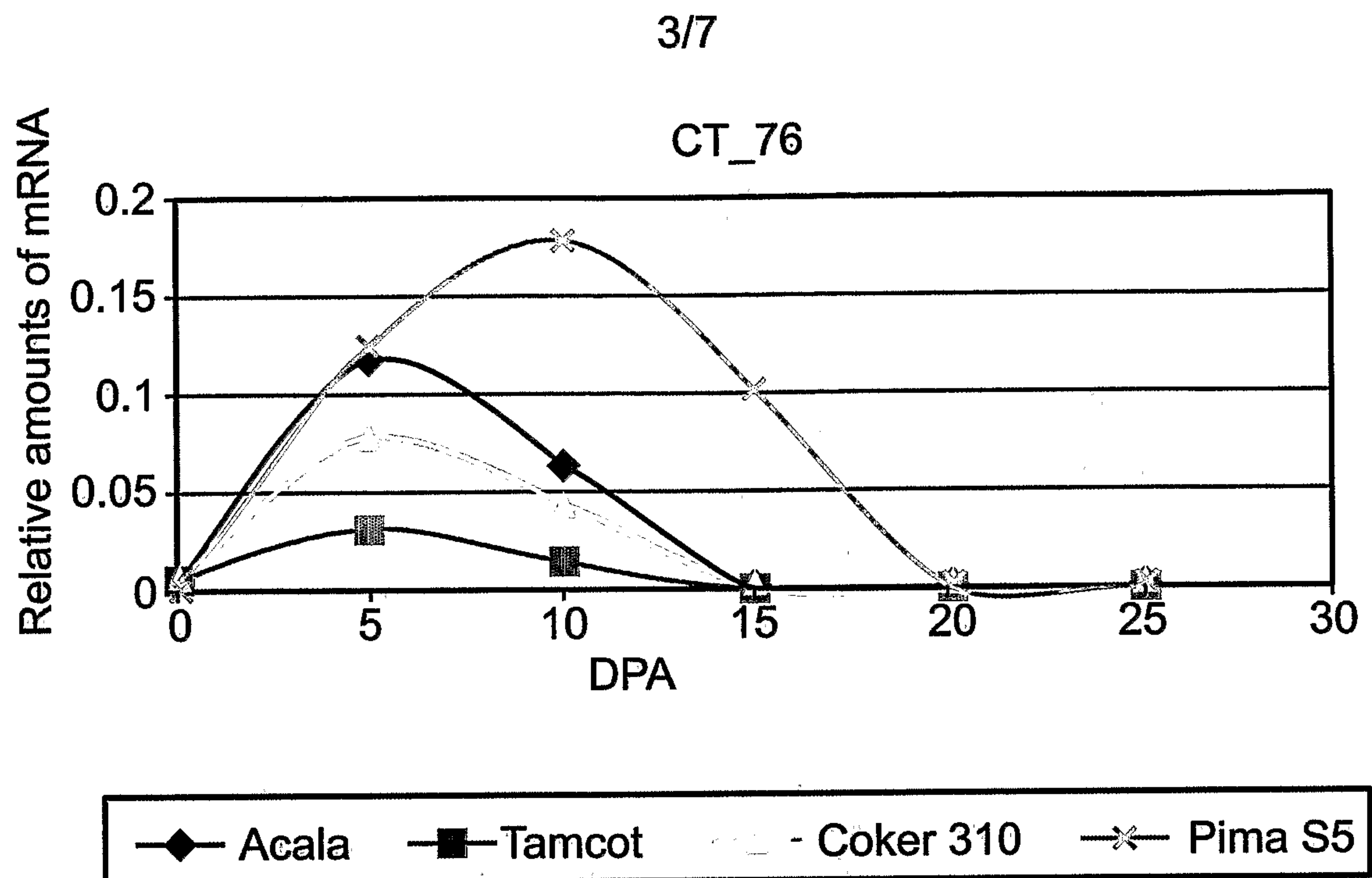
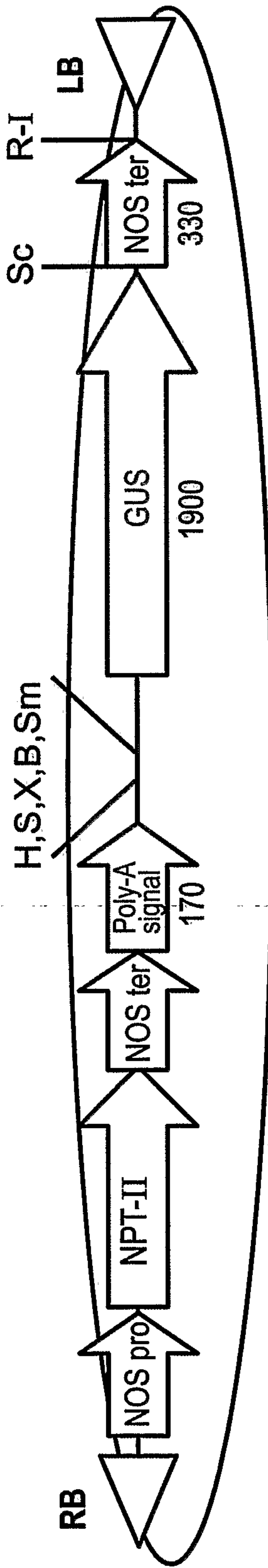


Fig. 3





- Legends:
- RB-T-DNA right border
  - LB-T-DNA left border
  - H-restriction enzyme-Hind-III
  - X-restriction enzyme-Xba-I
  - B-restriction enzyme-BamHI
  - S-restriction enzyme-Sal-I
  - Sma-restriction enzyme-Sma-I
  - R-I-restriction enzyme-Hind-III
  - H-restriction enzyme-Eco-RI
  - Sc-SacI/SstI/Ecl136II
  - (numbers)-Length in base-pairs

Fig. 4



5/7

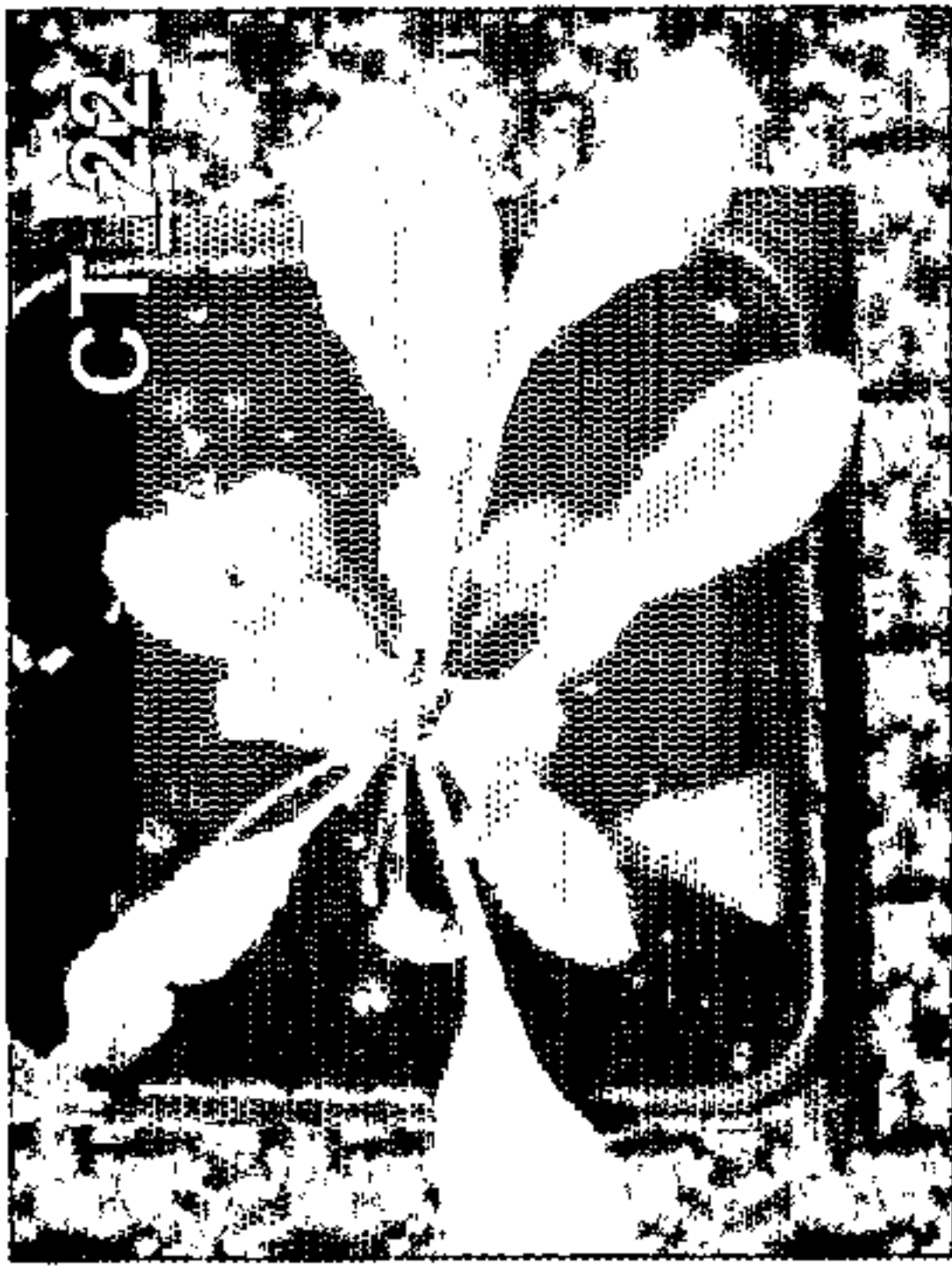


Fig. 5f



Fig. 5e

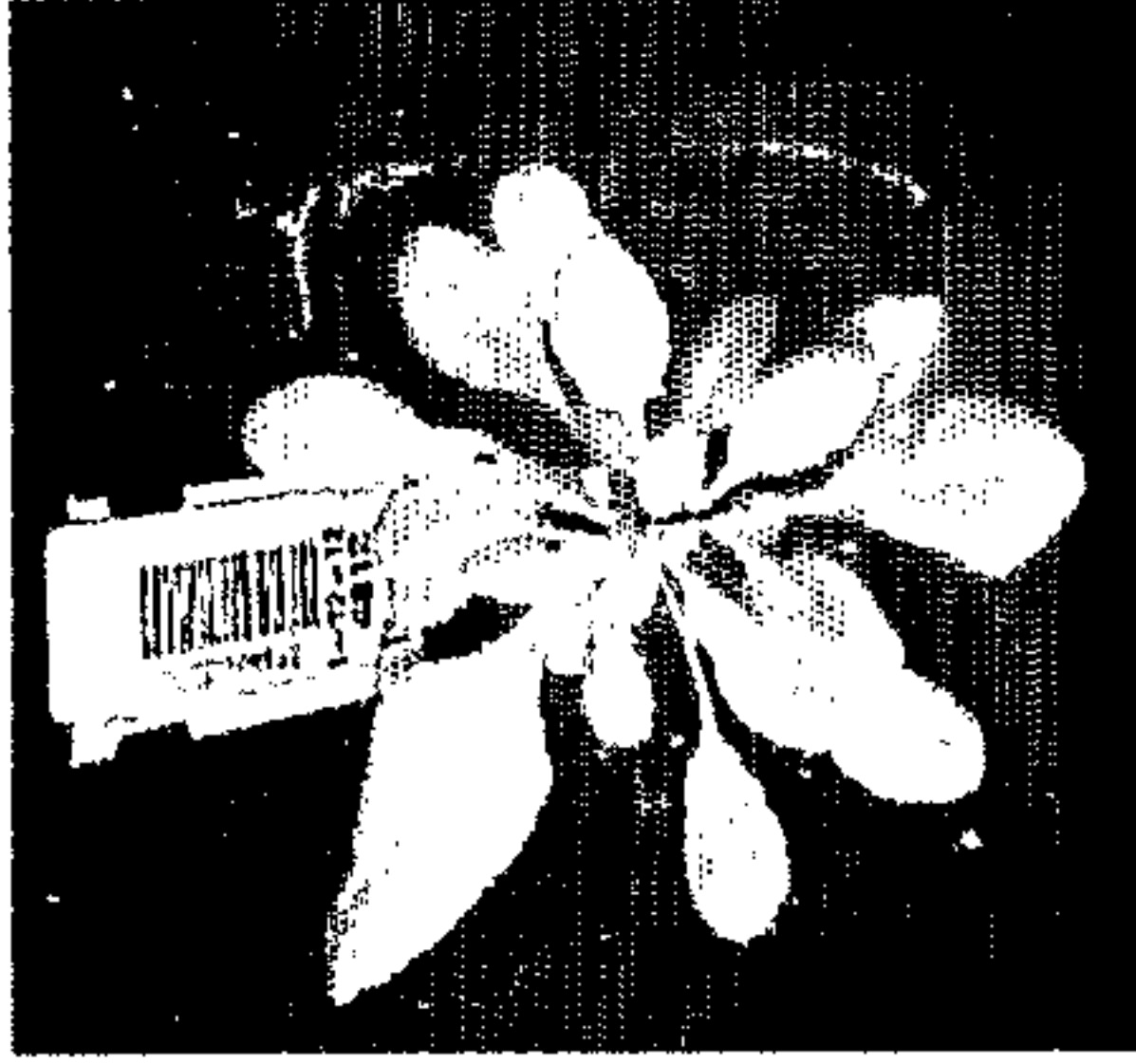


Fig. 5d

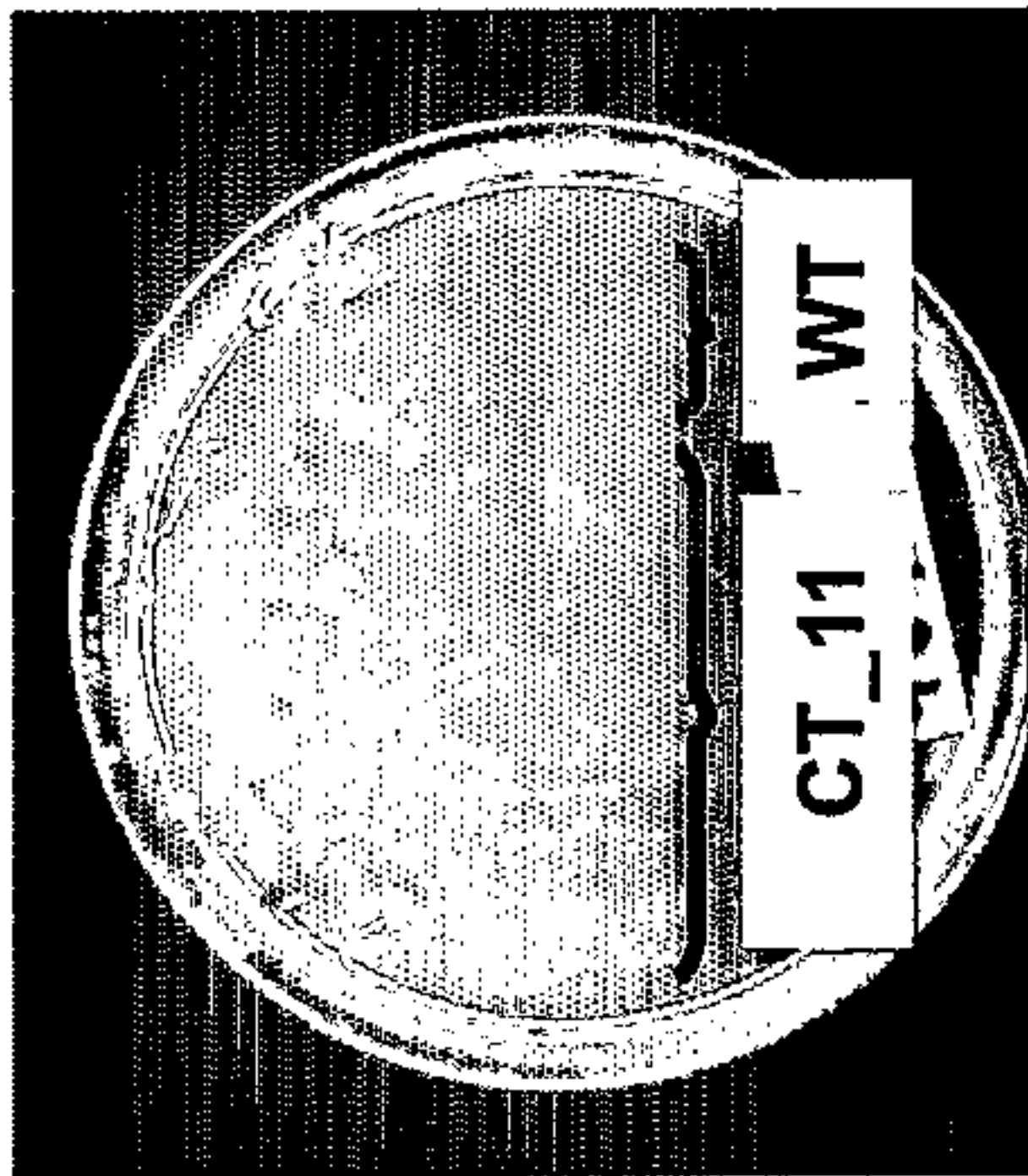


Fig. 5c

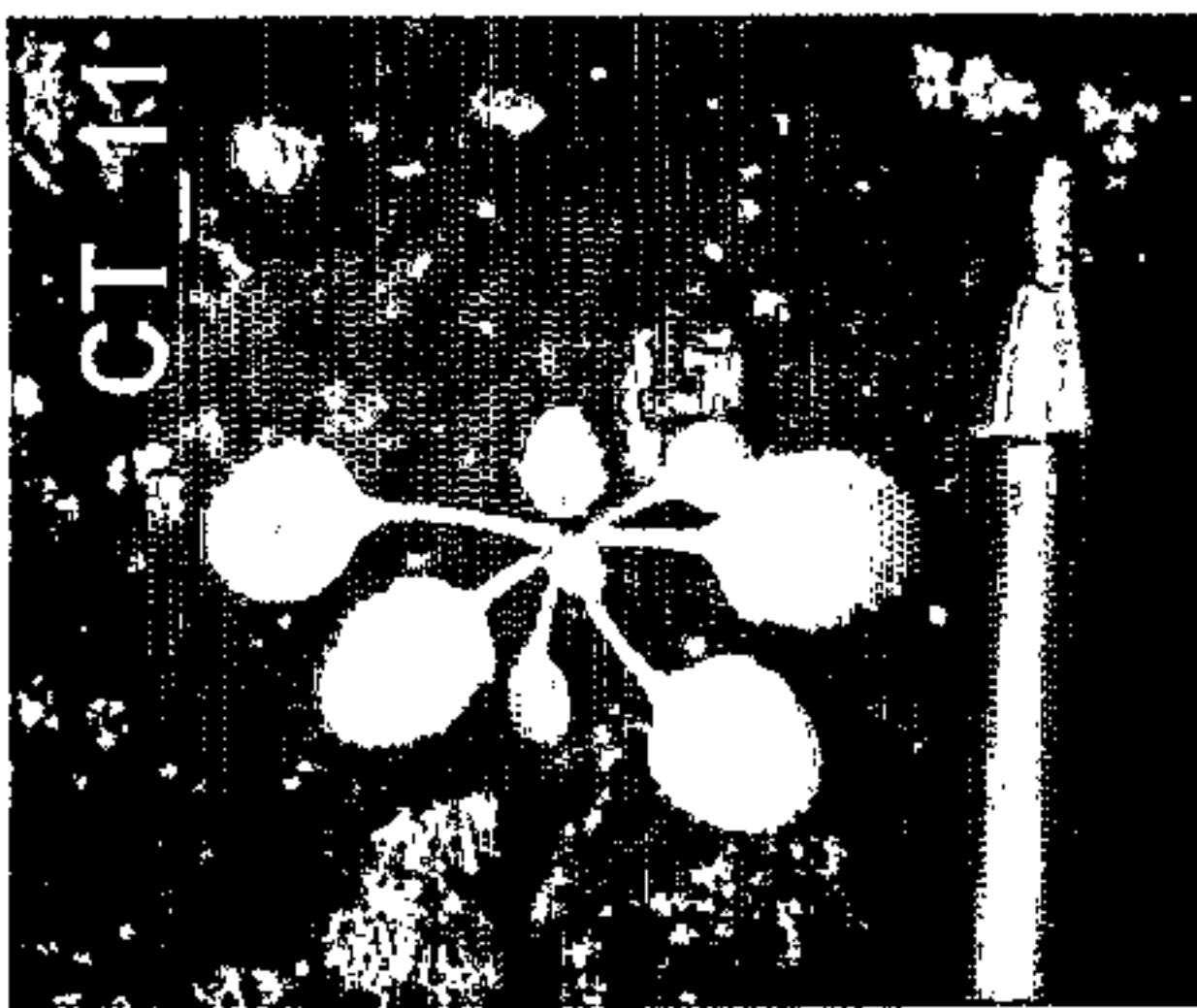


Fig. 5b

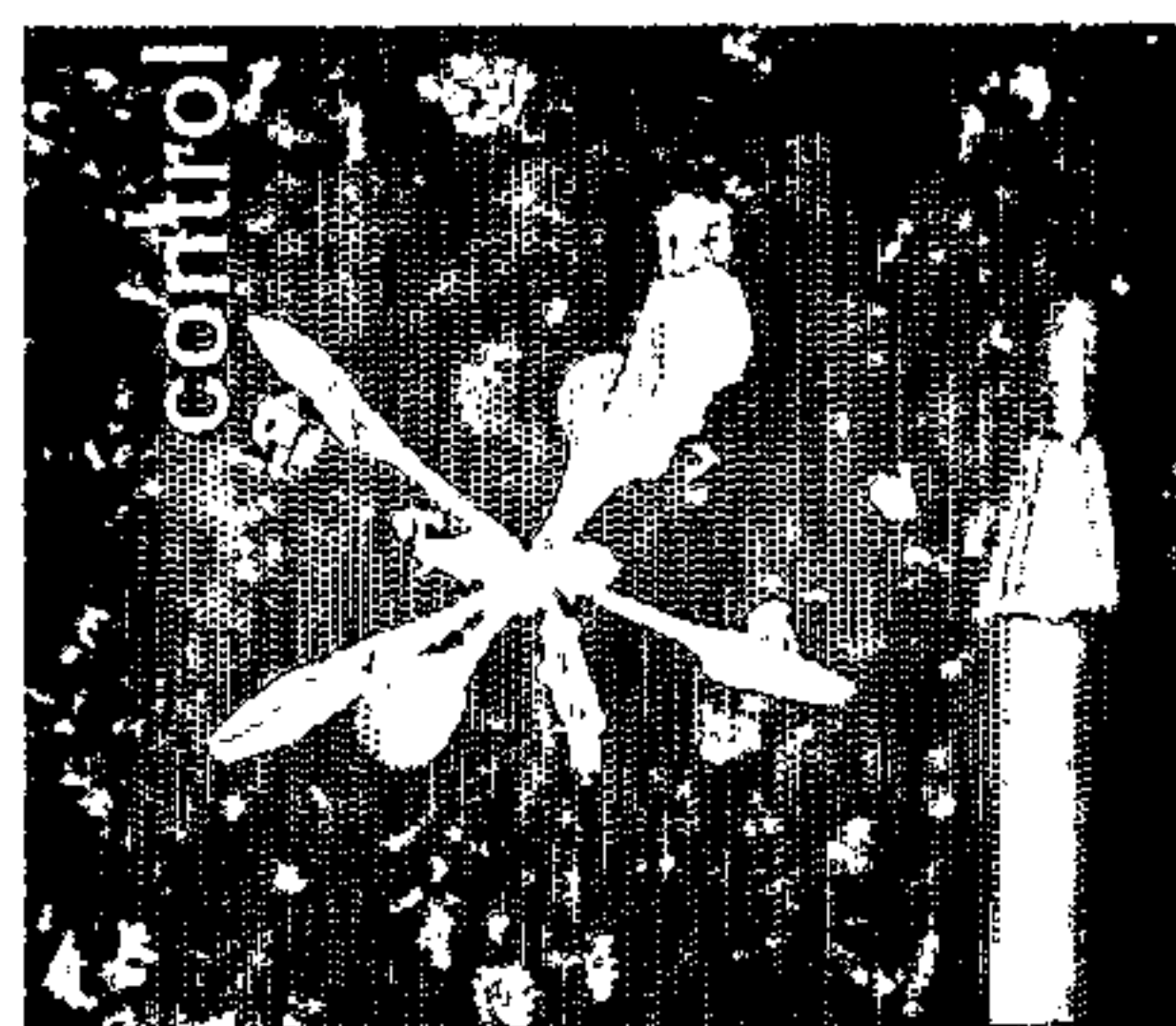


Fig. 5a

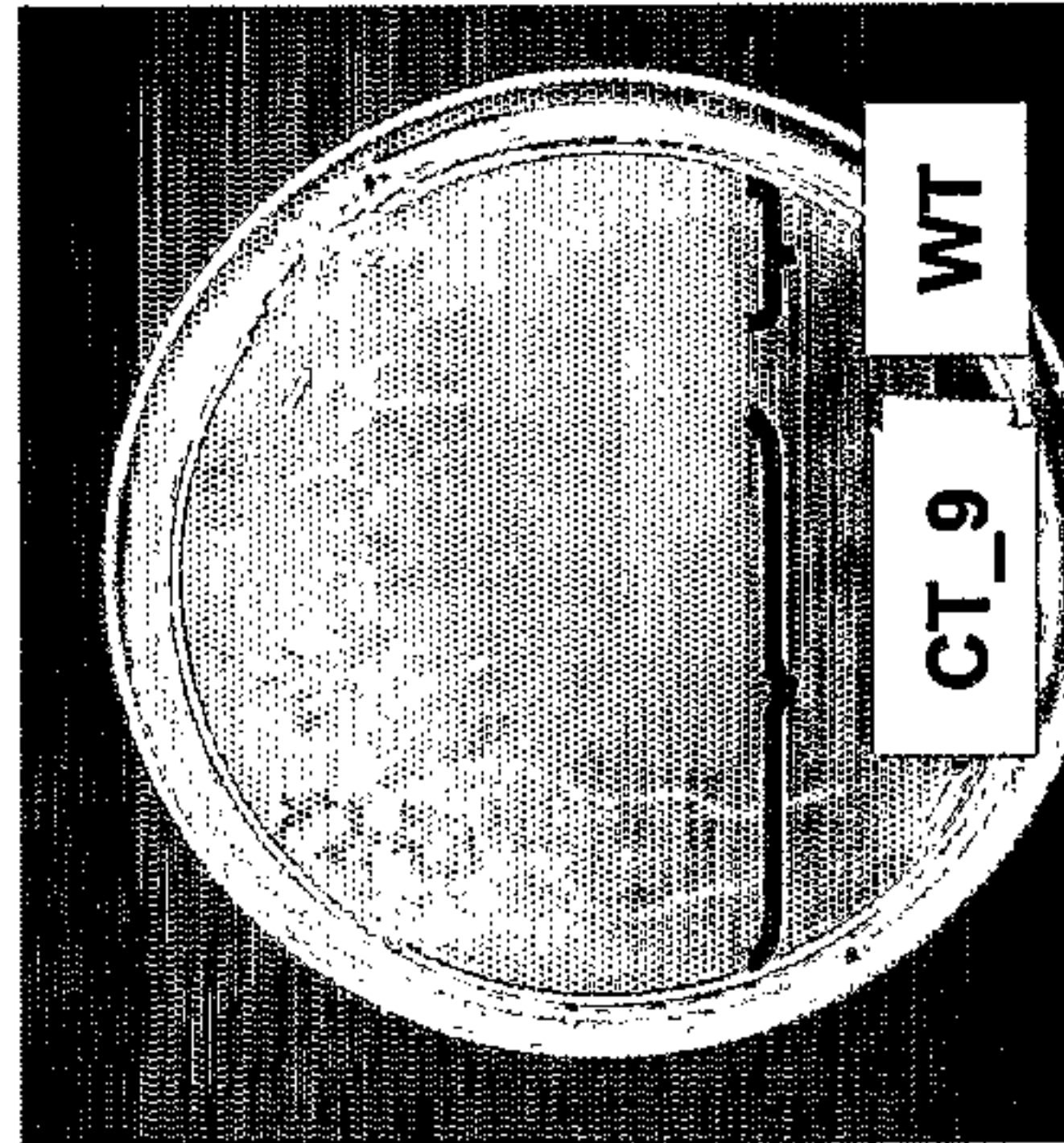


Fig. 5i

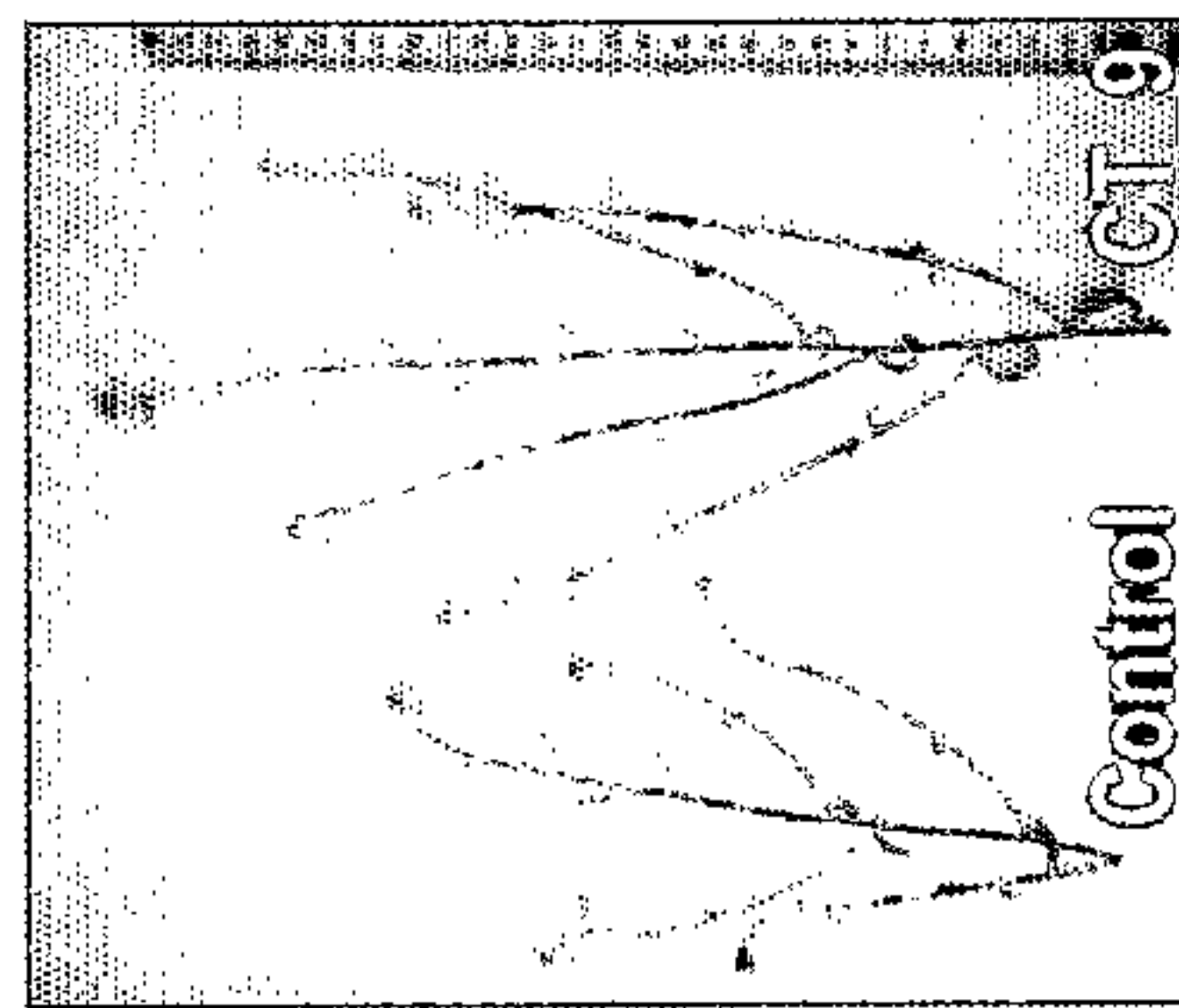


Fig. 5h

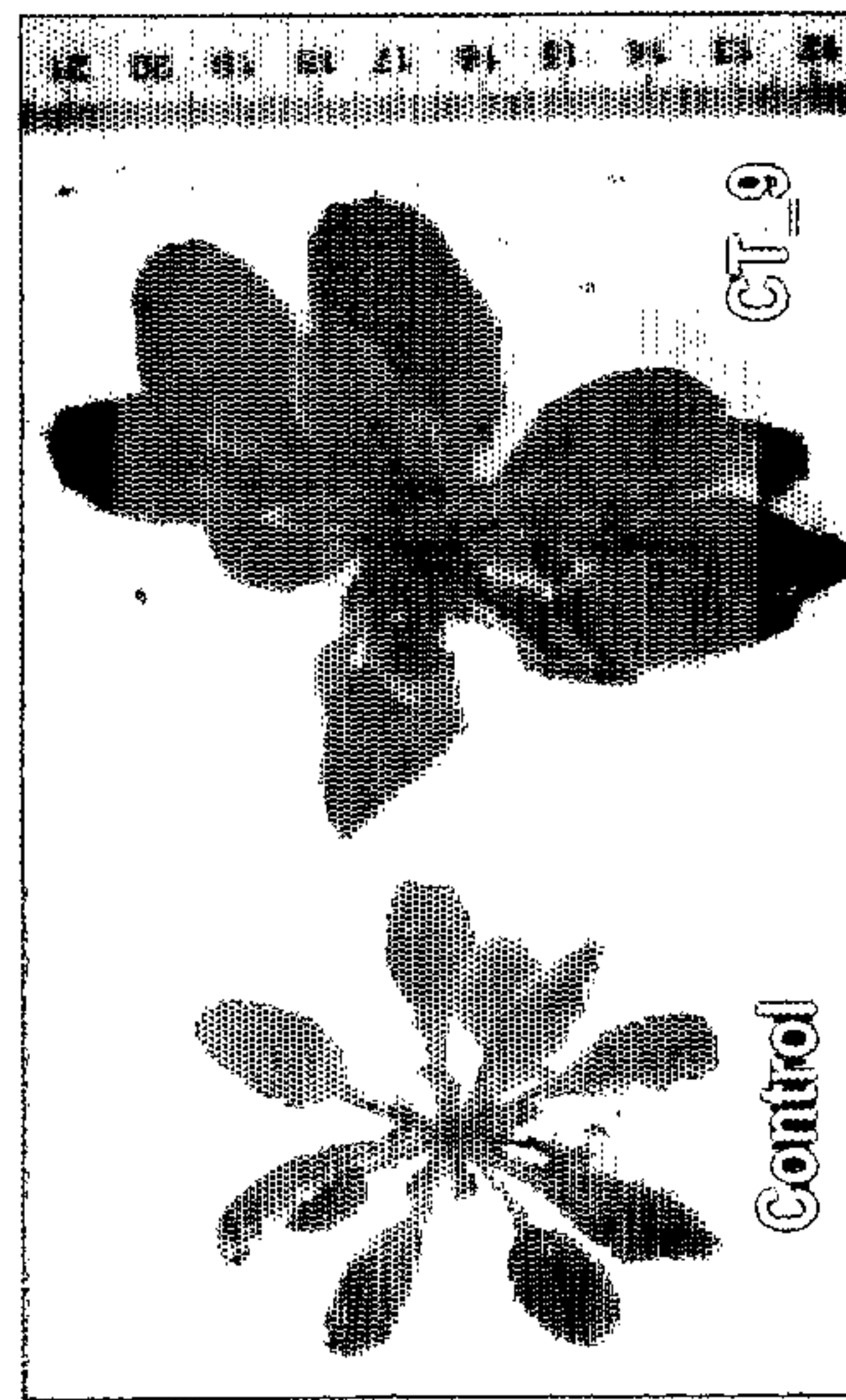


Fig. 5g

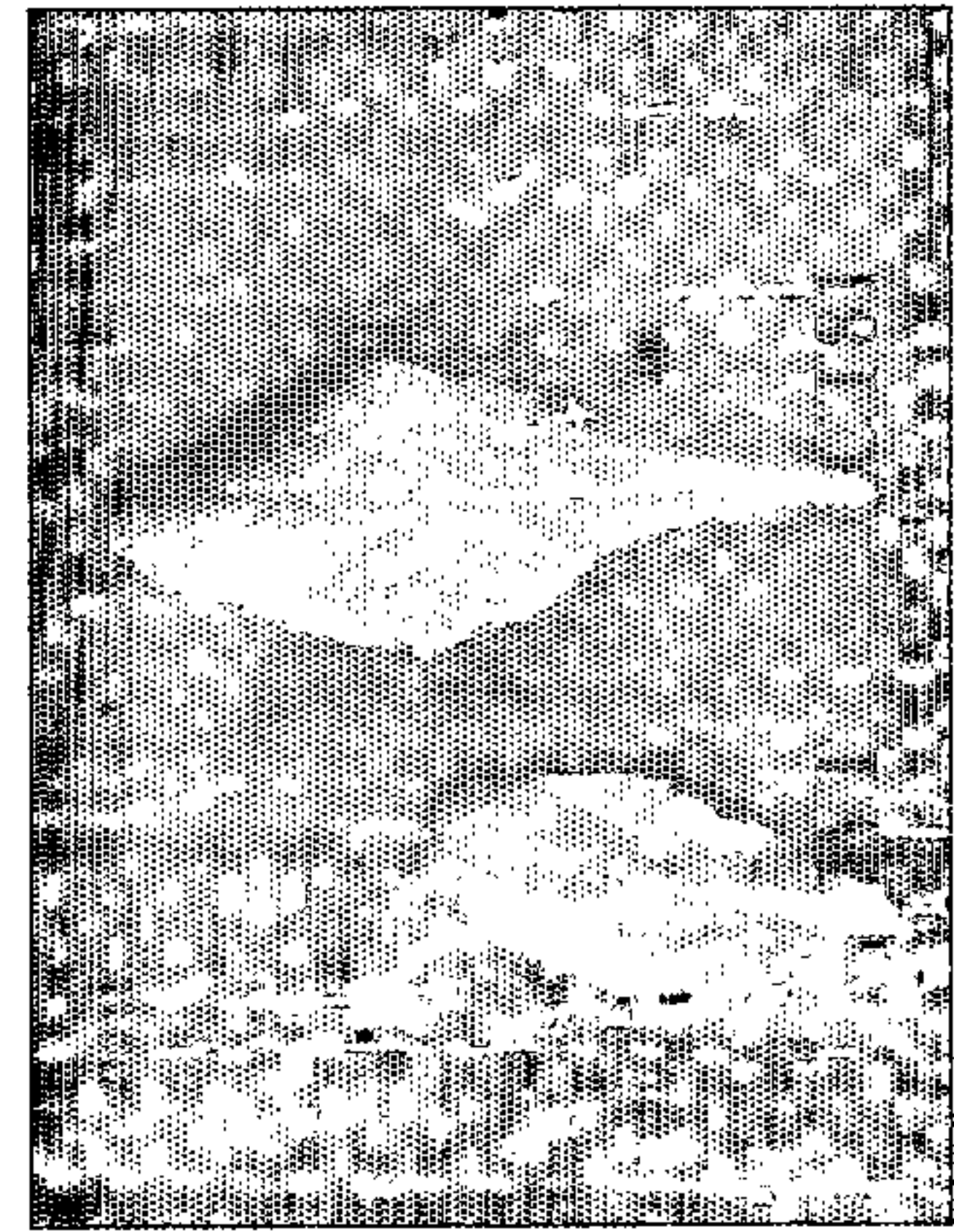


Fig. 5l

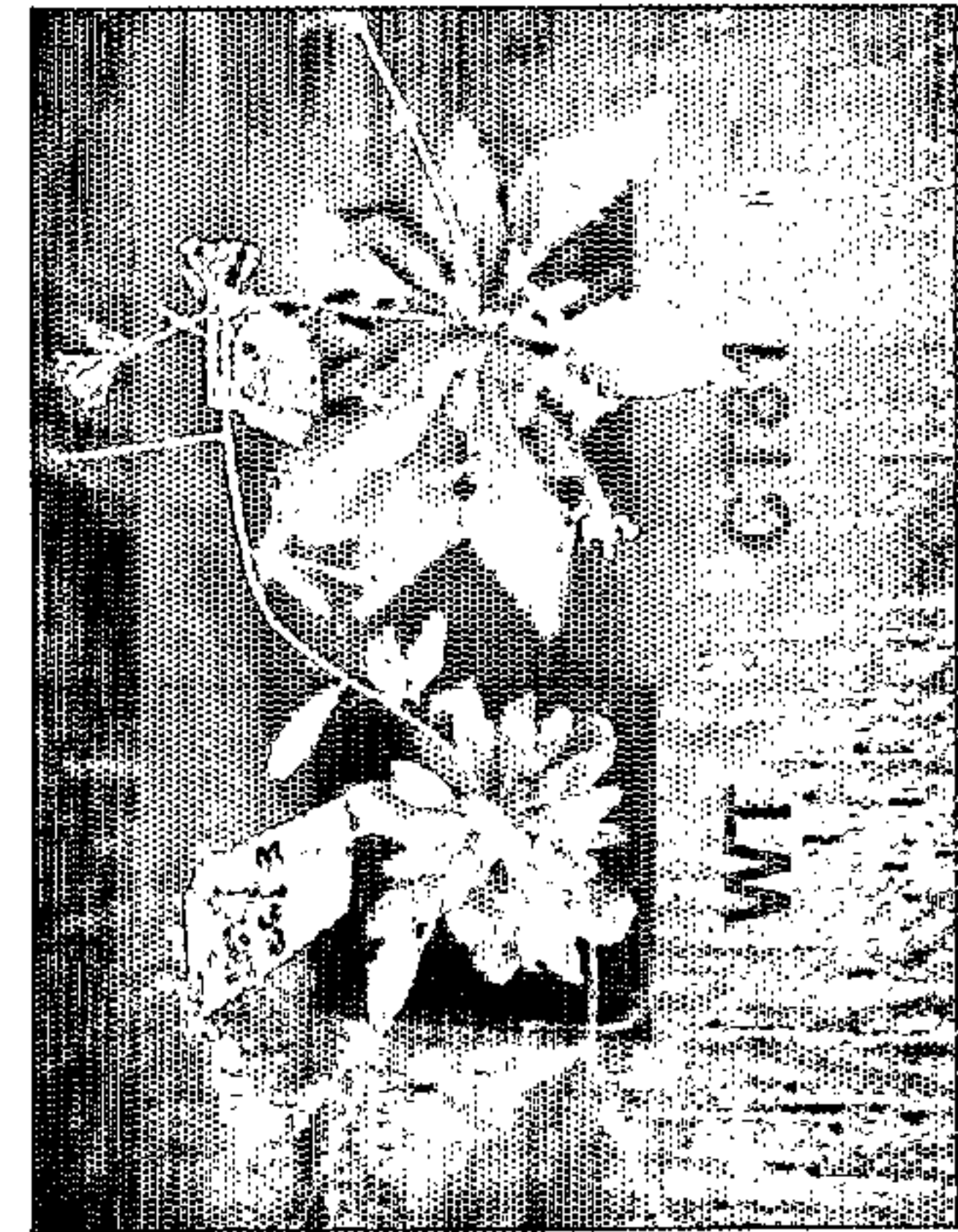


Fig. 5k

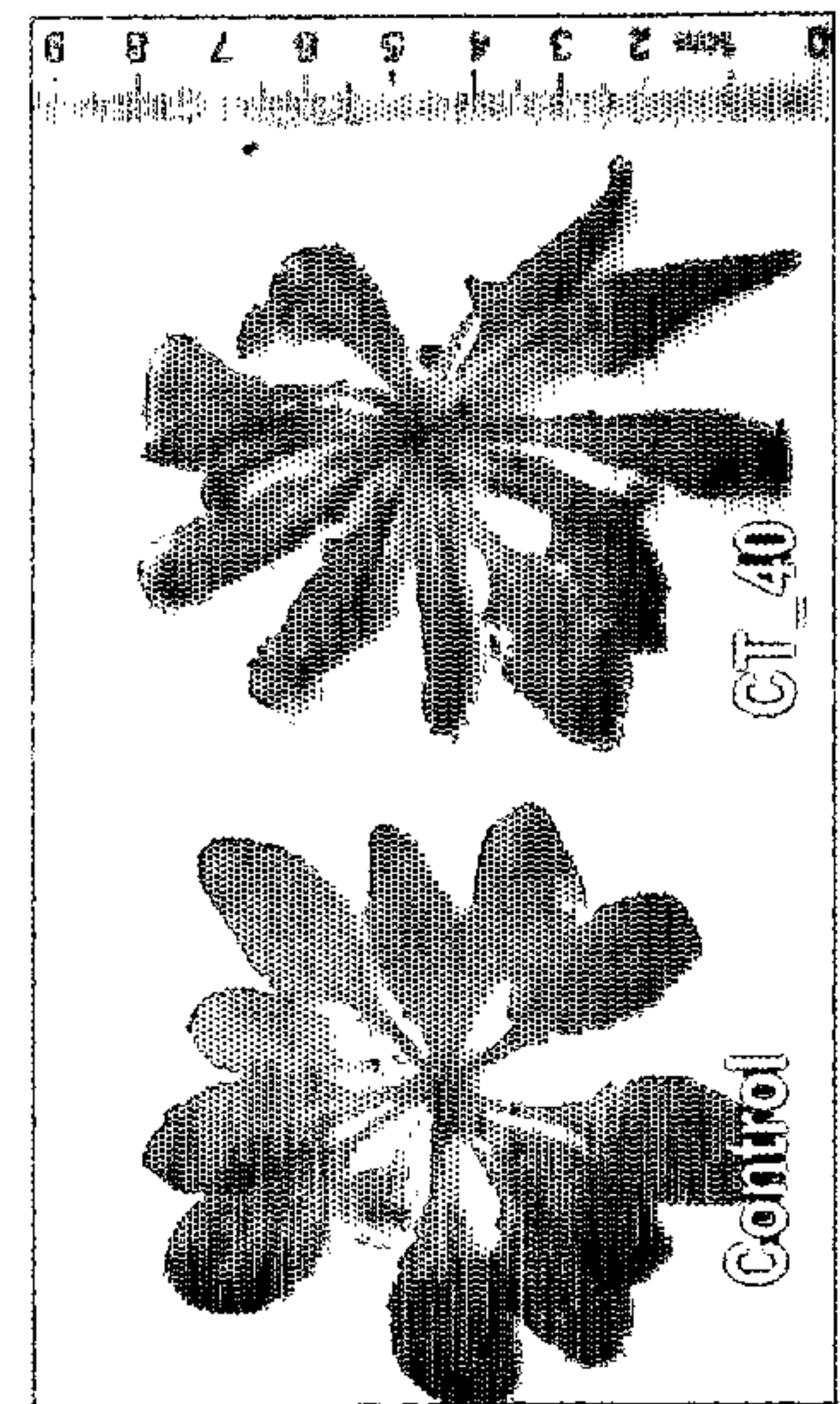


Fig. 5j



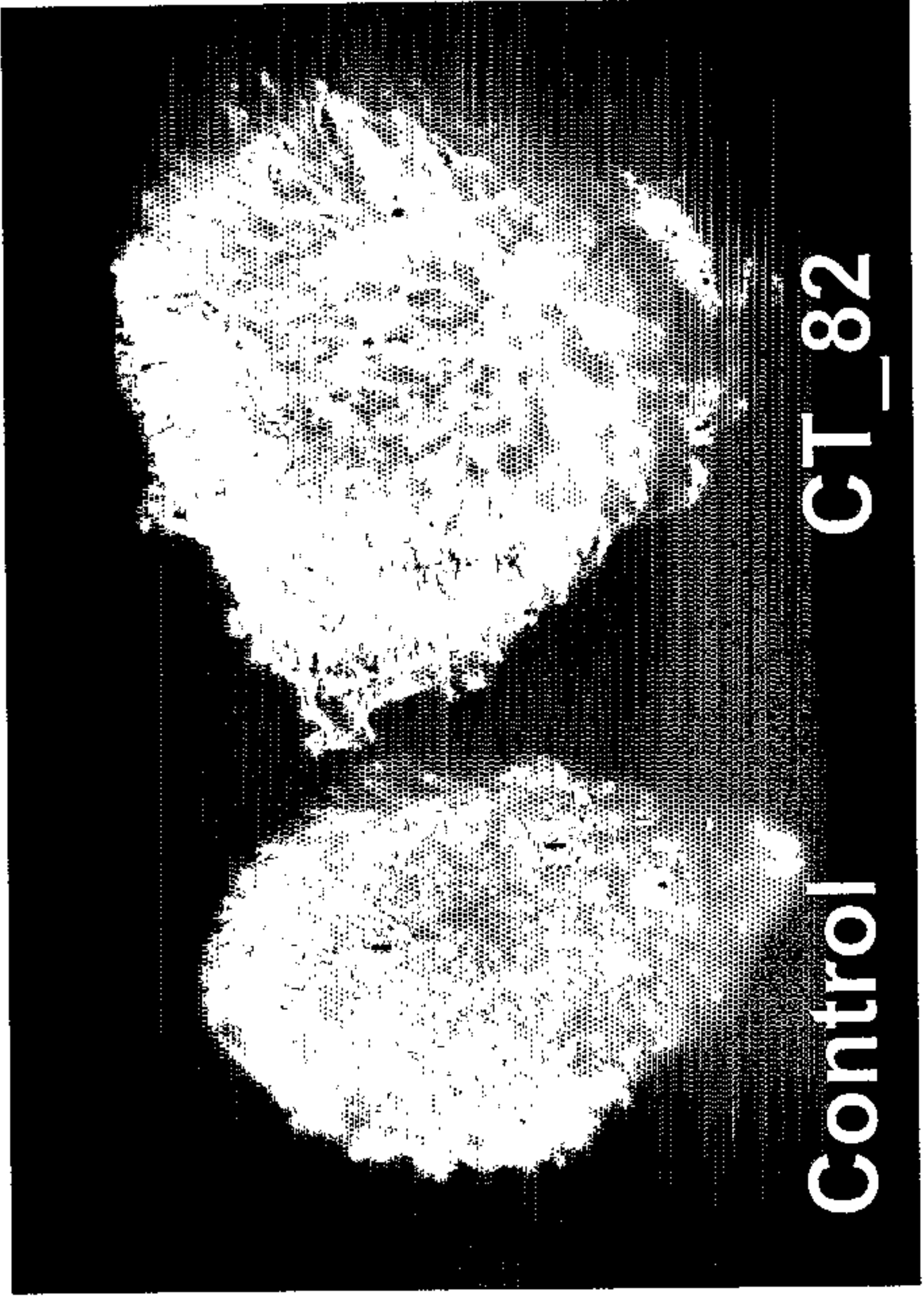


Fig. 6f

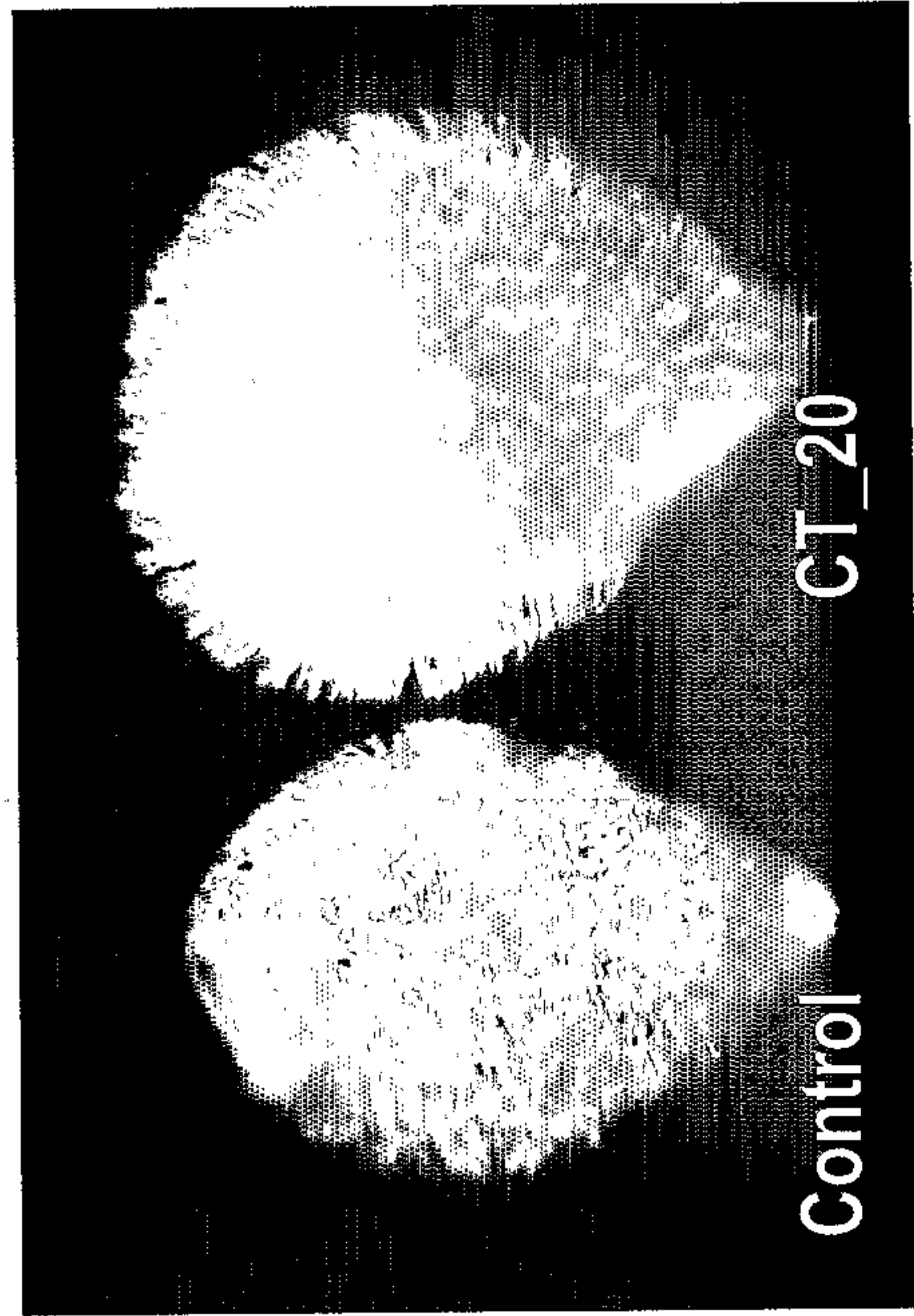


Fig. 6c

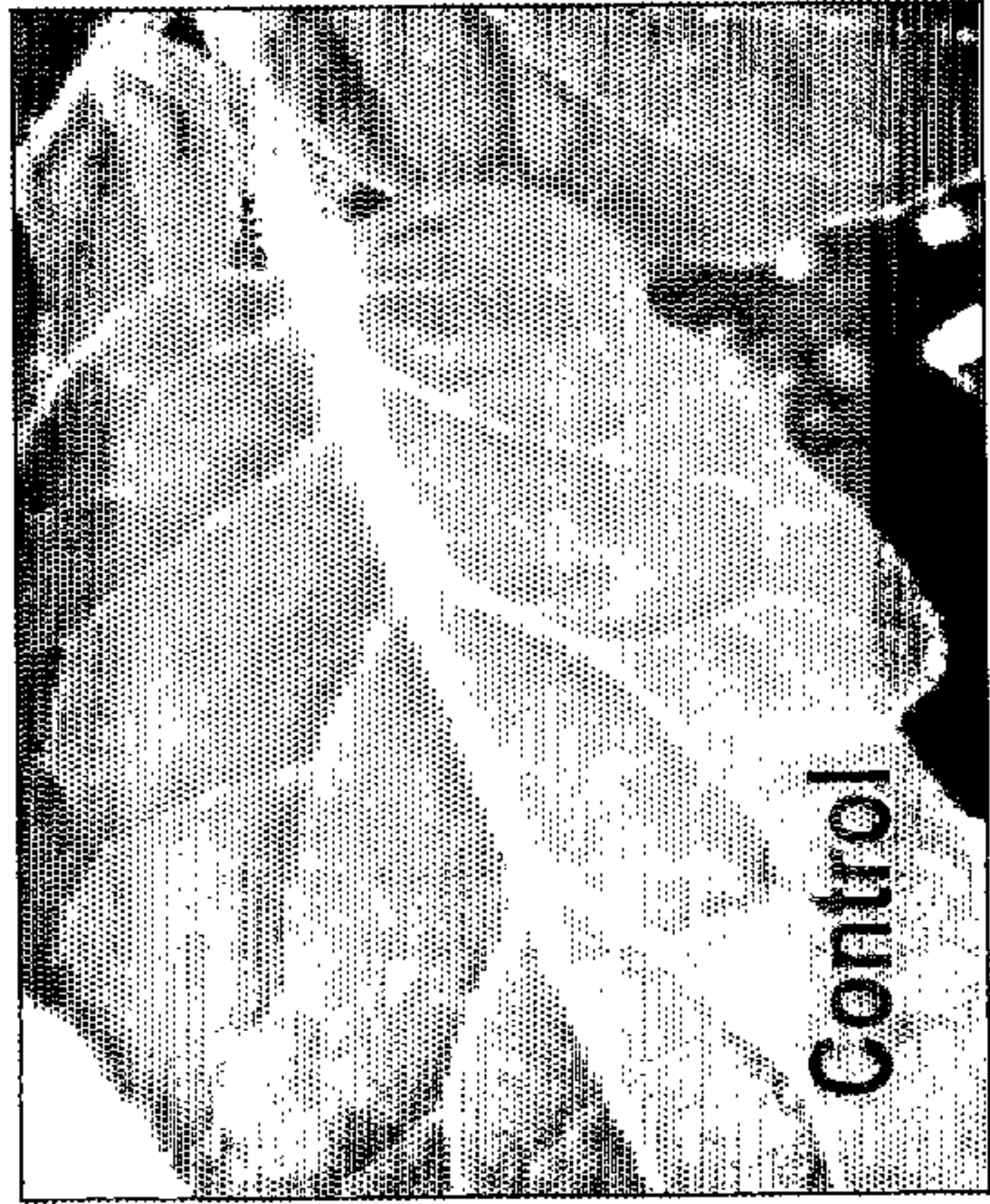


Fig. 6a



Fig. 6e

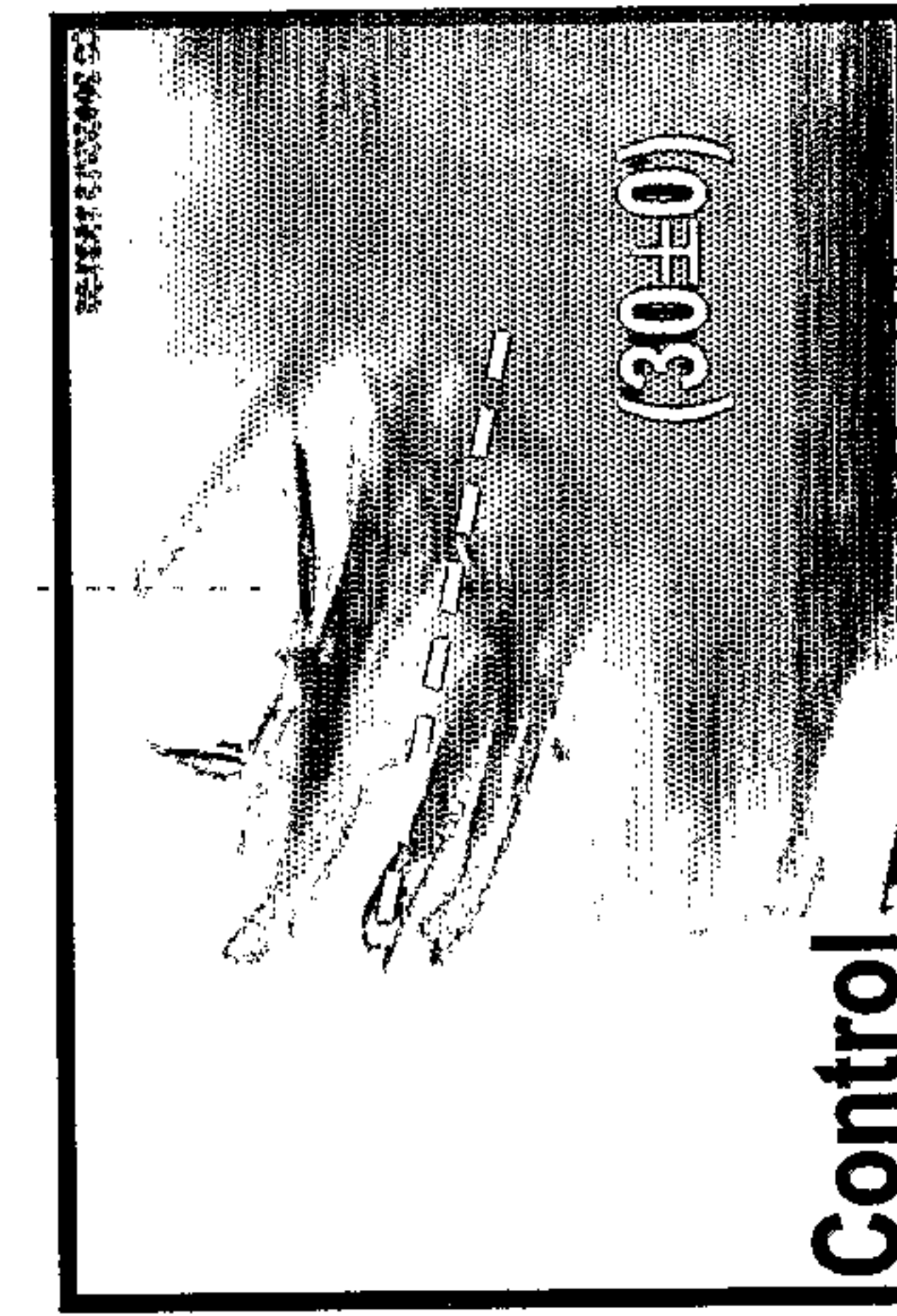


Fig. 6d



Fig. 6b



7/7

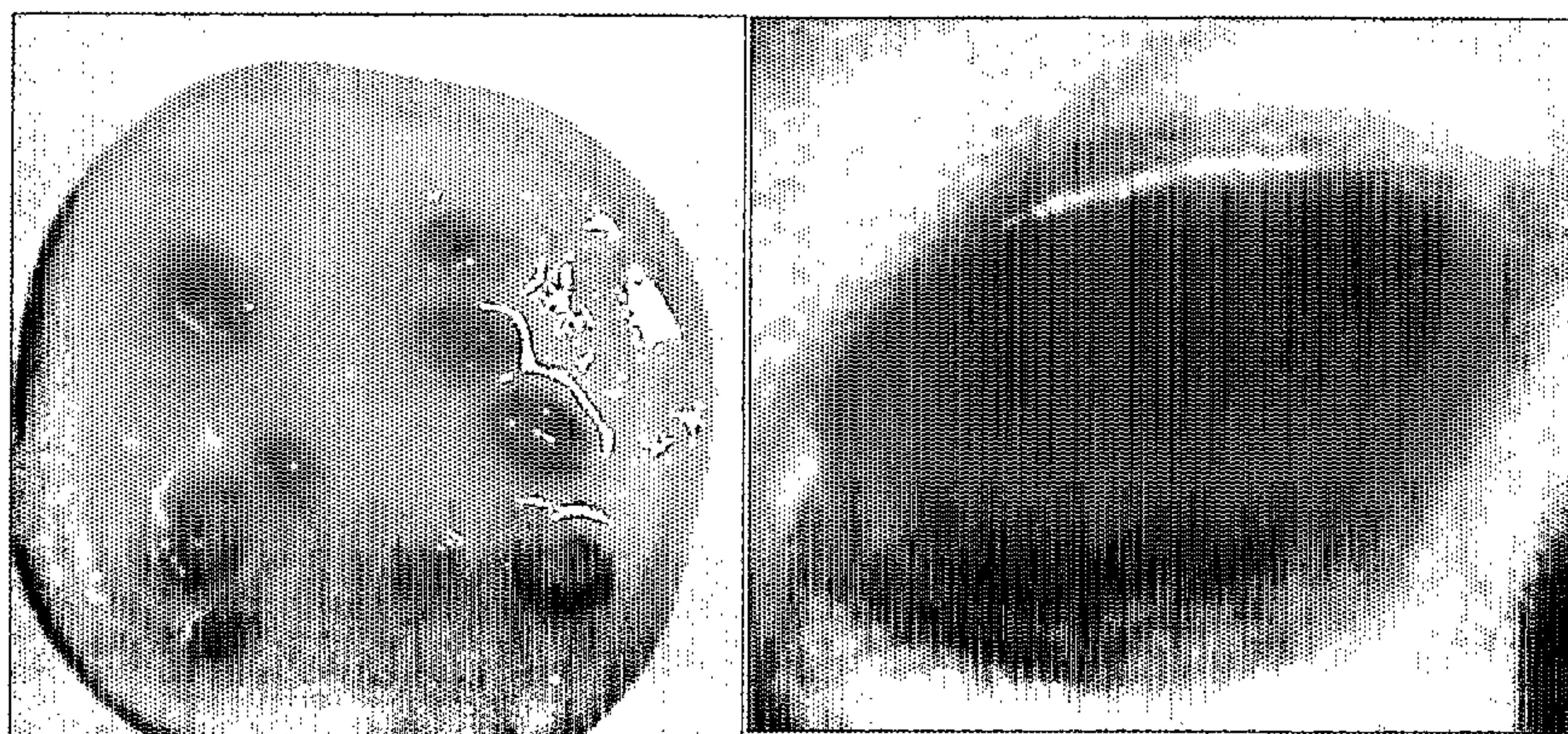


Fig. 7a

Fig. 7b

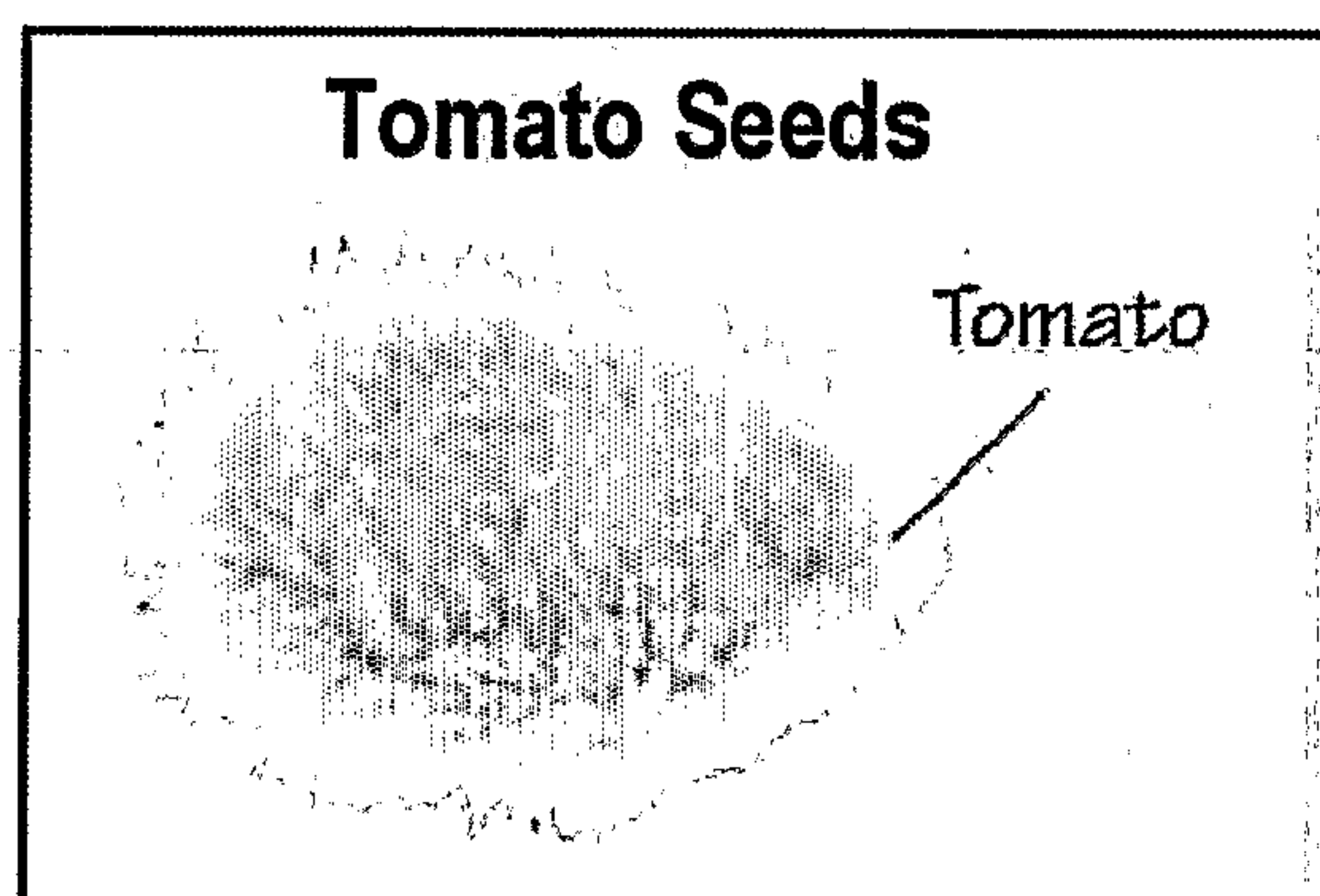


Fig. 8a

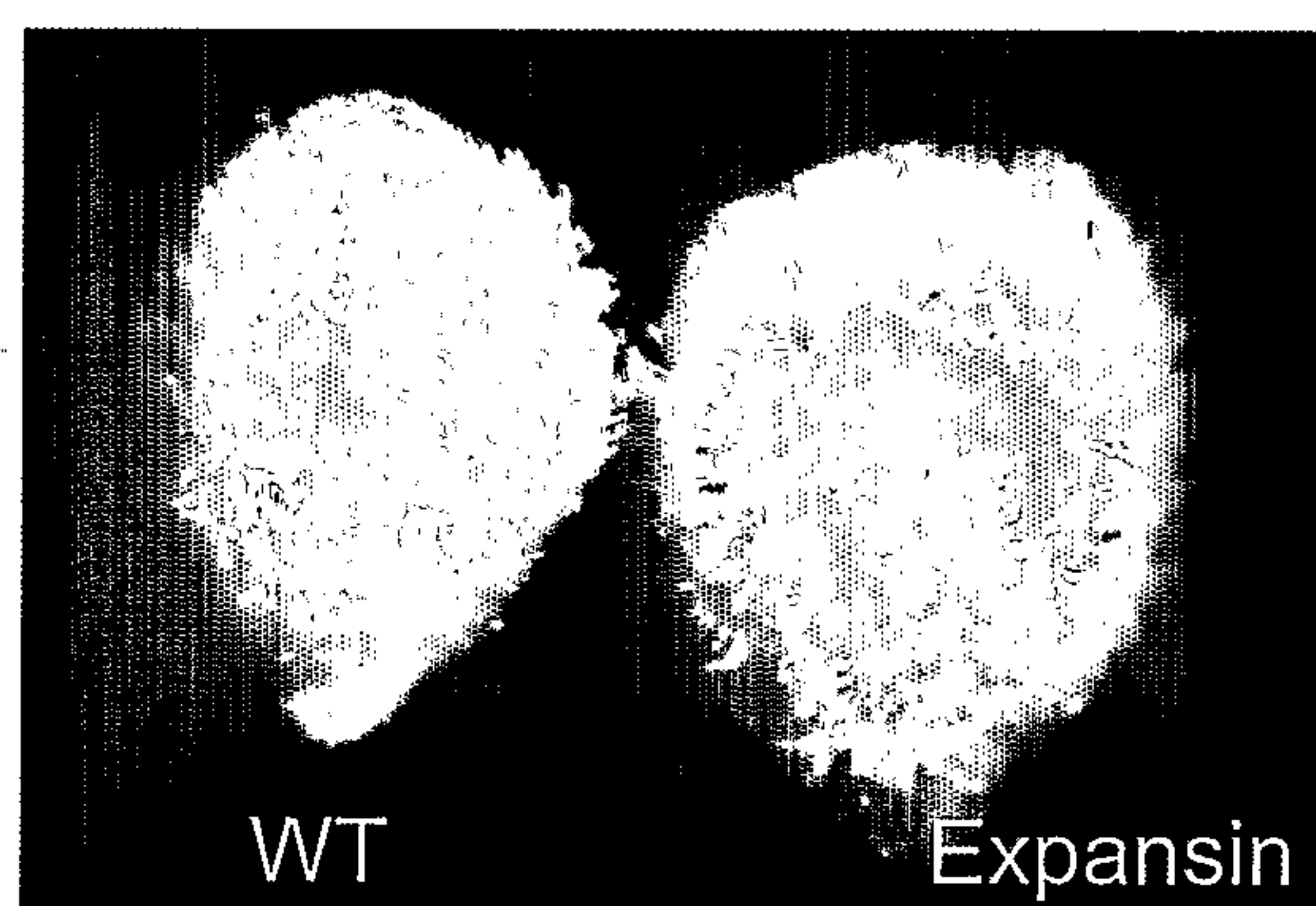
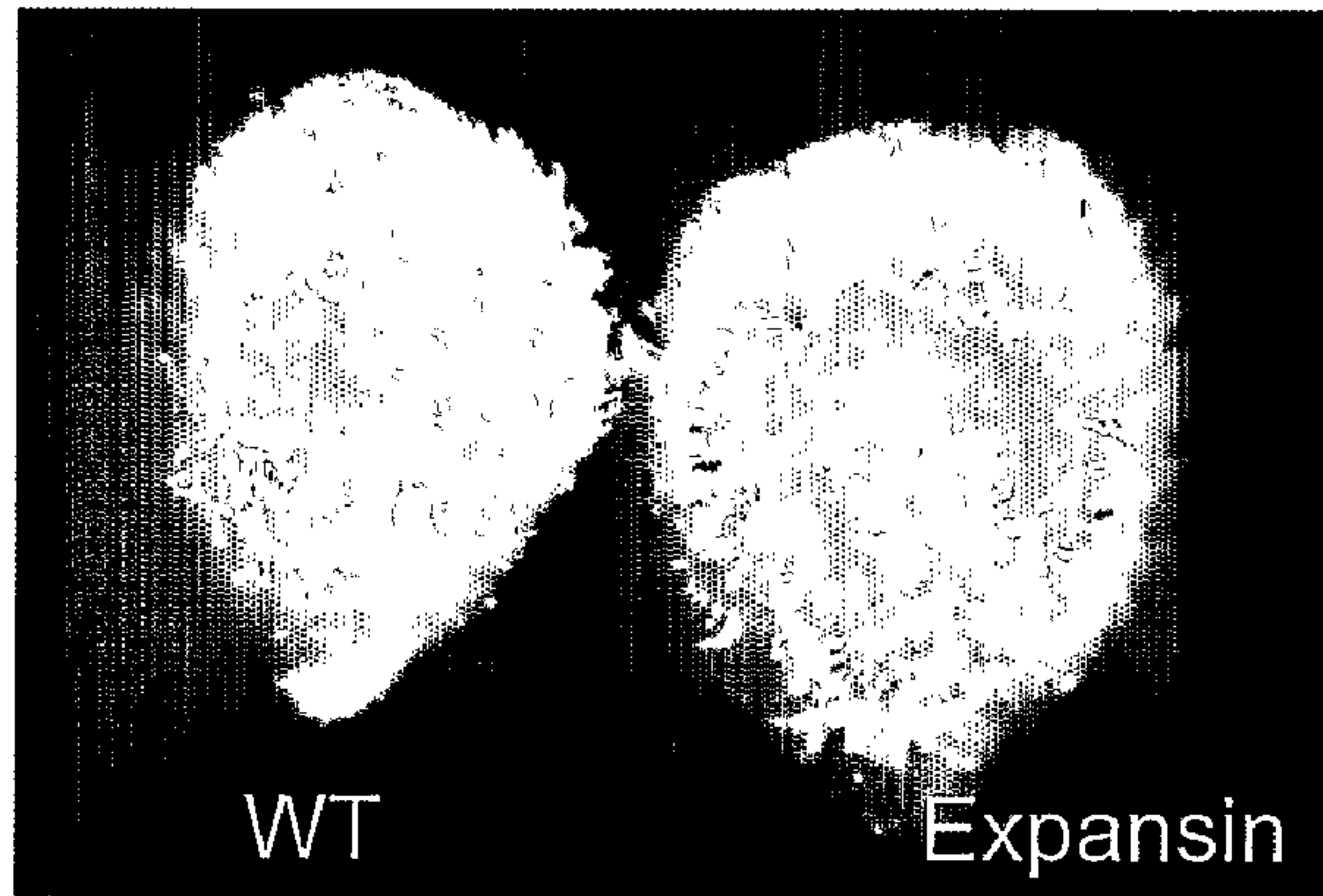


Fig. 8b



# Tomato Seeds

Tomato



a

b