



US 20060143729A1

(19) **United States**

(12) **Patent Application Publication**  
**Alexandrov et al.**

(10) **Pub. No.: US 2006/0143729 A1**

(43) **Pub. Date: Jun. 29, 2006**

(54) **NUCLEOTIDE SEQUENCES AND  
POLYPEPTIDES ENCODED THEREBY  
USEFUL FOR MODIFYING PLANT  
CHARACTERISTICS**

(75) Inventors: **Nickolai Alexandrov**, Thousand Oaks,  
CA (US); **Vyacheslav Brover**, Simi  
Valley, CA (US)

Correspondence Address:

**BIRCH STEWART KOLASCH & BIRCH  
PO BOX 747  
FALLS CHURCH, VA 22040-0747 (US)**

(73) Assignee: **CERES, INC.**, Thousand Oaks, CA

(21) Appl. No.: **11/174,307**

(22) Filed: **Jun. 30, 2005**

**Related U.S. Application Data**

(60) Provisional application No. 60/583,671, filed on Jun.  
30, 2004. Provisional application No. 60/583,781,  
filed on Jun. 30, 2004. Provisional application No.  
60/583,651, filed on Jun. 30, 2004.

**Publication Classification**

(51) **Int. Cl.**

**A01H 1/00** (2006.01)

**C12N 15/82** (2006.01)

**C12N 5/04** (2006.01)

(52) **U.S. Cl.** ..... **800/278**; 435/468; 435/419

(57) **ABSTRACT**

Isolated polynucleotides and polypeptides encoded thereby  
are described, together with the use of those products for  
making transgenic plants.

## NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY USEFUL FOR MODIFYING PLANT CHARACTERISTICS

### CROSS REFERENCE TO RELATED APPLICATION

[0001] This Nonprovisional application claims priority under 35 U.S.C. § 119(e) on U.S. Provisional Application No(s). 60/583,671; 60/583,781 and 60/583,651 filed on Jun. 30, 2004, the entire contents of which are hereby incorporated by reference.

### FIELD OF THE INVENTION

[0002] The present invention relates to isolated polynucleotides, polypeptides encoded thereby, and the use of those products for making transgenic plants or organisms, such as transgenic plants.

### BACKGROUND OF THE INVENTION

[0003] There are more than 300,000 species of plants. They show a wide diversity of forms, ranging from delicate liverworts, adapted for life in a damp habitat, to cacti, capable of surviving in the desert. The plant kingdom includes herbaceous plants, such as corn, whose life cycle is measured in months, to the giant redwood tree, which can live for thousands of years. This diversity reflects the adaptations of plants to survive in a wide range of habitats. This is seen most clearly in the flowering plants (phylum Angiospermophyta), which are the most numerous, with over 250,000 species. They are also the most widespread, being found from the tropics to the arctic.

[0004] The process of plant breeding involving man's intervention in natural breeding and selection is some 20,000 years old. It has produced remarkable advances in adapting existing species to serve new purposes. The world's economics was largely based on the successes of agriculture for most of these 20,000 years.

[0005] Plant breeding involves choosing parents, making crosses to allow recombination of gene (alleles) and searching for and selecting improved forms. Success depends on the genes/alleles available, the combinations required and the ability to create and find the correct combinations necessary to give the desired properties to the plant. Molecular genetics technologies are now capable of providing new genes, new alleles and the means of creating and selecting plants with the new, desired characteristics.

[0006] Plants specifically improved for agriculture, horticulture and other industries can be obtained using molecular technologies. As an example, great agronomic value can result from modulating the size of a plant as a whole or of any of its organs. The green revolution came about as a result of creating dwarf wheat plants, which produced a higher seed yield than taller plants because they could withstand higher levels and inputs of fertilizer and water.

[0007] Similarly, modulation of the size and stature of an entire plant, or a particular portion of a plant, allows production of plants better suited for a particular industry. For example, reductions in the height of specific ornamentals, crops and tree species can be beneficial by allowing easier harvesting. Alternatively, increasing height may be beneficial by providing more biomass. Other examples of

commercially desirable traits include increasing the length of the floral stems of cut flowers, increasing or altering leaf size and shape, enhancing the size of seeds and/or fruits, enhancing yields by specifically stimulating hormone (e.g. Brassinolide) synthesis and stimulating early flowering or evoking late flowering by altering levels of gibberellic acid or other hormones in specific cells. Changes in organ size and biomass also result in changes in the mass of constituent molecules such as secondary products.

[0008] To summarize, molecular genetic technologies provide the ability to modulate and manipulate growth, development and biochemistry of the entire plant as well as at the cell, tissue and organ levels. Thus, plant morphology, development and biochemistry are altered to maximize or minimize the desired plant trait.

### SUMMARY OF THE INVENTION

[0009] The present invention, therefore, relates to isolated polynucleotides, polypeptides encoded thereby, and the use of those products for making transgenic organisms, such as plants, bacteria, yeast, fungi and mammals, depending upon the desired characteristics.

[0010] In the field of agriculture and forestry efforts are constantly being made to produce plants with improved characteristics, such as increased overall yield or increased yield of biomass or chemical components, in particular in order to guarantee the supply of the constantly increasing world population with food and to guarantee the supply of reproducible raw materials. Conventionally, people try to obtain plants with an increased yield by breeding, but this is time-consuming and labor-intensive. Furthermore, appropriate breeding programs must be performed for each relevant plant species.

[0011] Recently, progress has been made by the genetic manipulation of plants. That is, by introducing into and expressing recombinant nucleic acid molecules in plants. Such approaches have the advantage of not usually being limited to one plant species, but being transferable to other plant species as well. EP-A 0 511 979, for example, discloses that the expression of a prokaryotic asparagine synthetase in plant cells *inter alia* leads to an increase in biomass production. Similarly, WO 96/21737 describes the production of plants with increased yield from the expression of deregulated or unregulated fructose-1,6-bisphosphatase due to an increased rate of the photosynthesis. Nevertheless, there still is a need for generally applicable processes that improve yield in plants interesting for agriculture or forestry purposes.

### DETAILED DESCRIPTION OF THE INVENTION

#### 1. Definitions

[0012] The following terms are utilized throughout this application:

[0013] Domain: Domains are fingerprints or signatures that can be used to characterize protein families and/or parts of proteins. Such fingerprints or signatures can comprise conserved (1) primary sequence, (2) secondary structure, and/or (3) three-dimensional conformation. Generally, each domain has been associated with either a family of proteins or motifs. Typically, these families

and/or motifs have been correlated with specific in-vitro and/or in-vivo activities. A domain can be any length, including the entirety of the sequence of a protein. Detailed descriptions of the domains, associated families and motifs, and correlated activities of the polypeptides of the instant invention are described below. Usually, the polypeptides with designated domain(s) can exhibit at least one activity that is exhibited by any polypeptide that comprises the same domain(s). Domains also define areas of non-coding sequences such as promoters and miRNAs.

[0014] Endogenous: The term “endogenous,” within the context of the current invention refers to any polynucleotide, polypeptide or protein sequence which is a natural part of a cell or organisms regenerated from said cell.

[0015] Exogenous: “Exogenous,” as referred to within, is any polynucleotide, polypeptide or protein sequence, whether chimeric or not, that is initially or subsequently introduced into the genome of an individual host cell or the organism regenerated from said host cell by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation (of dicots—e.g. Salomon et al. *EMBO J.* 3:141 (1984); Herrera-Estrella et al. *EMBO J.* 2:987 (1983); of monocots, representative papers are those by Escudero et al., *Plant J.* 10:355 (1996), Ishida et al., *Nature Biotechnology* 14:745 (1996), May et al., *Bio/Technology* 13:486 (1995)), biolistic methods (Armaleo et al., *Current Genetics* 17:97 (1990)), electroporation, in planta techniques, and the like. Such a plant containing the exogenous nucleic acid is referred to here as a T<sub>0</sub> for the primary transgenic plant and T<sub>1</sub> for the first generation. The term “exogenous” as used herein is also intended to encompass inserting a naturally found element into a non-naturally found location.

[0016] Gene: The term “gene,” as used in the context of the current invention, encompasses all regulatory and coding sequence contiguously associated with a single hereditary unit with a genetic function. Genes can include non-coding sequences that modulate the genetic function that include, but are not limited to, those that specify polyadenylation, transcriptional regulation, DNA conformation, chromatin conformation, extent and position of base methylation and binding sites of proteins that control all of these. Genes comprised of “exons” (coding sequences), which may be interrupted by “introns” (non-coding sequences), encode proteins. A gene’s genetic function may require only RNA expression or protein production, or may only require binding of proteins and/or nucleic acids without associated expression. In certain cases, genes adjacent to one another may share sequence in such a way that one gene will overlap the other. A gene can be found within the genome of an organism, artificial chromosome, plasmid, vector, etc., or as a separate isolated entity.

[0017] Heterologous sequences: “Heterologous sequences” are those that are not operatively linked or are not contiguous to each other in nature. For example, a promoter from corn is considered heterologous to an *Arabidopsis* coding region sequence. Also, a promoter from a gene encoding a growth factor from corn is considered heterologous to a sequence encoding the corn

receptor for the growth factor. Regulatory element sequences, such as UTRs or 3' end termination sequences that do not originate in nature from the same gene as the coding sequence originates from, are considered heterologous to said coding sequence. Elements operatively linked in nature and contiguous to each other are not heterologous to each other. On the other hand, these same elements remain operatively linked but become heterologous if other filler sequence is placed between them. Thus, the promoter and coding sequences of a corn gene expressing an amino acid transporter are not heterologous to each other, but the promoter and coding sequence of a corn gene operatively linked in a novel manner are heterologous.

[0018] Homologous gene: In the current invention, “homologous gene” refers to a gene that shares sequence similarity with the gene of interest. This similarity may be in only a fragment of the sequence and often represents a functional domain such as, examples including without limitation a DNA binding domain, a domain with tyrosine kinase activity, or the like. The functional activities of homologous genes are not necessarily the same.

[0019] Misexpression: The term “misexpression” refers to an increase or a decrease in the transcription of a coding region into a complementary RNA sequence as compared to the parental wild-type. This term also encompasses expression of a gene or coding region for a different time period as compared to the wild-type and/or from a non-natural location within the plant genome.

[0020] Percentage of sequence identity: “Percentage of sequence identity,” as used herein, is determined by comparing two optimally aligned sequences over a comparison window, where the fragment of the polynucleotide or amino acid sequence in the comparison window may comprise additions or deletions (e.g., gaps or overhangs) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Add. APL. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (USA)* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, PASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wis.), or by inspection. Given that two sequences have been identified for comparison, GAP and BESTFIT are preferably employed to determine their optimal alignment. Typically, the default values of 5.00 for gap weight and 0.30 for gap weight length are used. The term “substantial sequence identity” between polynucleotide or polypeptide sequences refers to polynucleotide or polypeptide comprising a sequence that has at least 80% sequence identity,

preferably at least 85%, more preferably at least 90% and most preferably at least 95%, even more preferably, at least 96%, 97%, 98% or 99% sequence identity compared to a reference sequence using the programs.

**[0021]** Regulatory Sequence: The term “regulatory sequence,” as used in the current invention, refers to any nucleotide sequence that influences transcription or translation initiation and rate, and stability and/or mobility of the transcript or polypeptide product. Regulatory sequences include, but are not limited to, promoters, promoter control elements, protein binding sequences, 5' and 3' UTRs, transcriptional start site, termination sequence, polyadenylation sequence, introns, certain sequences within a coding sequence, etc.

**[0022]** Stringency: “Stringency” as used herein is a function of probe length, probe composition (G+C content), and salt concentration, organic solvent concentration, and temperature of hybridization or wash conditions. Stringency is typically compared by the parameter  $T_m$ , which is the temperature at which 50% of the complementary molecules in the hybridization are hybridized, in terms of a temperature differential from  $T_m$ . High stringency conditions are those providing a condition of  $T_m-5^\circ\text{C}$ . to  $T_m-0^\circ\text{C}$ . Medium or moderate stringency conditions are those providing  $T_m-20^\circ\text{C}$ . to  $T_m-29^\circ\text{C}$ . Low stringency conditions are those providing a condition of  $T_m-40^\circ\text{C}$ . to  $T_m-48^\circ\text{C}$ . The relationship of hybridization conditions to  $T_m$  (in  $^\circ\text{C}$ .) is expressed in the mathematical equation

$$T_m = 81.5 - 16.6(\log_{10}[\text{Na}^+]) + 0.41(\% \text{ G+C}) - (600/N) \quad (1)$$

where N is the length of the probe. This equation works well for probes 14 to 70 nucleotides in length that are identical to the target sequence. The equation below for  $T_m$  of DNA-DNA hybrids is useful for probes in the range of 50 to greater than 500 nucleotides, and for conditions that include an organic solvent (formamide).

$$T_m = 81.5 + 16.6 \log \{[\text{Na}^+]/(1 + 0.7[\text{Na}^+])\} + 0.41(\% \text{ G+C}) - 500/L - 0.63(\% \text{ formamide}) \quad (2)$$

where L is the length of the probe in the hybrid. (P. Tijessen, “Hybridization with Nucleic Acid Probes” in Laboratory Techniques in Biochemistry and Molecular Biology, P. C. van der Vliet, ed., c. 1993 by Elsevier, Amsterdam.) The  $T_m$  of equation (2) is affected by the nature of the hybrid; for DNA-RNA hybrids  $T_m$  is  $10-15^\circ\text{C}$ . higher than calculated, for RNA-RNA hybrids  $T_m$  is  $20-25^\circ\text{C}$ . higher. Because the  $T_m$  decreases about  $1^\circ\text{C}$ . for each 1% decrease in homology when a long probe is used (Bonner et al., *J. Mol. Biol.* 81:123 (1973)), stringency conditions in polynucleotide hybridization reactions can be adjusted to favor hybridization of polynucleotides from identical genes or related family members.

**[0023]** Equation (2) is derived assuming equilibrium and therefore, hybridizations according to the present invention are most preferably performed under conditions of probe excess and for sufficient time to achieve equilibrium. The time required to reach equilibrium can be shortened by inclusion of a hybridization accelerator such as dextran sulfate or another high volume polymer in the hybridization buffer.

**[0024]** Stringency conditions can be selected during the hybridization reaction or after hybridization has occurred by

altering the salt and temperature conditions of the wash solutions used. The formulas shown above are equally valid when used to compute the stringency of a wash solution. Preferred wash solution stringencies lie within the ranges stated above; high stringency is  $5-8^\circ\text{C}$ . below  $T_m$ , medium or moderate stringency is  $26-29^\circ\text{C}$ . below  $T_m$  and low stringency is  $45-48^\circ\text{C}$ . below  $T_m$ .

**[0025]** Substantially free of: A composition containing A is “substantially free of” B when at least 85% by weight of the total A+B in the composition is A. Preferably, A comprises at least about 90% by weight of the total of A+B in the composition, more preferably at least about 95% or even 99% by weight. For example, a plant gene or DNA sequence can be considered substantially free of other plant genes or DNA sequences.

**[0026]** Translational start site: In the context of the current invention, a “translational start site” is usually an ATG in the cDNA transcript, more usually the first ATG. A single cDNA, however, may have multiple translational start sites.

**[0027]** Transcription start site: “Transcription start site” is used in the current invention to describe the point at which transcription is initiated. This point is typically located about 25 nucleotides downstream from a TFIID binding site, such as a TATA box. Transcription can initiate at one or more sites within the gene, and a single gene may have multiple transcriptional start sites, some of which may be specific for transcription in a particular cell-type or tissue.

**[0028]** Untranslated region (UTR): A “UTR” is any contiguous series of nucleotide bases that is transcribed, but is not translated. These untranslated regions may be associated with particular functions such as increasing mRNA message stability. Examples of UTRs include, but are not limited to polyadenylation signals, terminations sequences, sequences located between the transcriptional start site and the first exon (5' UTR) and sequences located between the last exon and the end of the mRNA (3' UTR).

**[0029]** Variant: The term “variant” is used herein to denote a polypeptide or protein or polynucleotide molecule that differs from others of its kind in some way. For example, polypeptide and protein variants can consist of changes in amino acid sequence and/or charge and/or post-translational modifications (such as glycosylation, etc).

## 2. Important Characteristics of the Polynucleotides of the Invention

**[0030]** The genes and polynucleotides of the present invention are of interest because when they are misexpressed (i.e. when expressed at a non-natural location or in an increased amount) they produce plants with important modified characteristics as discussed below. These traits can be used to exploit or maximize plant products or to minimize undesirable characteristics. For example, an increase in plant height is beneficial in species grown or harvested for their main stem or trunk, such as ornamental cut flowers, fiber crops (e.g. flax, kenaf, hesperaloe, hemp) and wood producing trees. Increase in inflorescence thickness is also desirable for some ornamentals, while increases in the number, shape and size of leaves can lead to increased production/ harvest from leaf crops such as lettuce, spinach, cabbage and tobacco. Likewise, a decrease in plant height is beneficial in species that are particularly susceptible to lodging or uprooting due to wind stress.

[0031] The polynucleotides and polypeptides of the invention were isolated from *Arabidopsis thaliana*, corn, soybean, wheat, *Brassica* and others as noted in the Sequence Listing. The polynucleotides and polypeptides are useful to confer on transgenic plants the properties identified for each sequence in the relevant portion (miscellaneous feature section) of the Sequence Listing. The miscellaneous feature section of the sequence listing contains, for each sequence, a description of the domain or other characteristic from which the sequence has the function known in the art for other sequences. Some identified domains are indicated with "PFam Name", signifying that the pfam name and description can be found in the pfam database at <http://pfam.wustl.edu>. Other domains are indicated by reference to a "GI Number" from the public sequence database maintained by GenBank under the NCBI, including the non-redundant (NR) database.

[0032] The sequences of the invention can be applied to substrates for use in array applications such as, but not limited to, assays of global gene expression, under varying conditions of development, and growth conditions. The arrays are also used in diagnostic or forensic methods

[0033] The polynucleotides, or fragments thereof, can also be used as probes and primers. Probe length varies depending on the application. For use as primers, probes are 12-40 nucleotides, preferably 18-30 nucleotides long. For use in mapping, probes are preferably 50 to 500 nucleotides, preferably 100-250 nucleotides long. For Southern hybridizations, probes as long as several kilobases are used.

[0034] The probes and/or primers are produced by synthetic procedures such as the triester method of Matteucci et al. *J. Am. Chem. Soc.* 103:3185(1981) or according to Urdea et al. *Proc. Natl. Acad.* 80:7461 (1981) or using commercially available automated oligonucleotide synthesizers.

[0035] The polynucleotides of the invention can be utilized in a number of methods known to those skilled in the art as probes and/or primers to isolate and detect polynucleotides including, without limitation: Southern, Northern, Branched DNA hybridization assays, polymerase chain reaction microarray assays and variations thereof. Specific methods given by way of examples, and discussed below include:

[0036] Hybridization

[0037] Methods of Mapping

[0038] Southern Blotting

[0039] Isolating cDNA from Related Organisms

[0040] Isolating and/or Identifying Homologous and Orthologous Genes.

Also, the nucleic acid molecules of the invention can be used in other methods, such as high density oligonucleotide hybridizing assays, described, for example, in U.S. Pat. Nos. 6,004,753 and 5,945,306.

[0041] The polynucleotides or fragments thereof of the present invention can be used as probes and/or primers for detection and/or isolation of related polynucleotide sequences through hybridization. Hybridization of one nucleic acid to another constitutes a physical property that defines the polynucleotide of the invention and the identified related sequences. Also, such hybridization imposes struc-

tural limitations on the pair. A good general discussion of the factors for determining hybridization conditions is provided by Sambrook et al. ("Molecular Cloning, a Laboratory Manual, 2nd ed., c. 1989 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; see esp., chapters 11 and 12). Additional considerations and details of the physical chemistry of hybridization are provided by G. H. Keller and M. M. Manak "DNA Probes", 2<sup>nd</sup> Ed. pp. 1-25, c. 1993 by Stockton Press, New York, N.Y.

[0042] When using the polynucleotides to identify orthologous genes in other species, the practitioner will preferably adjust the amount of target DNA of each species so that, as nearly as is practical, the same number of genome equivalents are present for each species examined. This prevents faint signals from species having large genomes, and thus small numbers of genome equivalents per mass of DNA, from erroneously being interpreted as absence of the corresponding gene in the genome.

[0043] The probes and/or primers of the instant invention can also be used to detect or isolate nucleotides that are "identical" to the probes or primers. Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below.

[0044] Isolated polynucleotides within the scope of the invention also include allelic variants of the specific sequences presented in the Sequence Listing. The probes and/or primers of the invention are also used to detect and/or isolate polynucleotides exhibiting at least 80% sequence identity with the sequences of the Sequence Listing or fragments thereof. Related polynucleotide sequences can also be identified according to the methods described in U.S. Patent Publication 20040137466A1, dated Jul. 15, 2004 to Jofuku et al.

[0045] With respect to nucleotide sequences, degeneracy of the genetic code provides the possibility to substitute at least one nucleotide of the nucleotide sequence of a gene with a different nucleotide without changing the amino acid sequence of the polypeptide. Hence, the DNA of the present invention also has any base sequence that has been changed from a sequence in the Sequence Listing by substitution in accordance with degeneracy of genetic code. References describing codon usage include: Carels et al., *J. Mol. Evol.* 46: 45 (1998) and Fennoy et al., *Nucl. Acids Res.* 21(23): 5294 (1993).

[0046] The polynucleotides of the invention are also used to create various types of genetic and physical maps of the genome of corn, *Arabidopsis*, soybean, rice, wheat, or other plants. Some are absolutely associated with particular phenotypic traits, allowing construction of gross genetic maps. Creation of such maps is based on differences or variants, generally referred to as polymorphisms, between different parents used in crosses. Common methods of detecting polymorphisms that can be used are restriction fragment length polymorphisms (RFLPs, single nucleotide polymorphisms (SNPs) or simple sequence repeats (SSRs).

[0047] The use of RFLPs and of recombinant inbred lines for such genetic mapping is described for *Arabidopsis* by Alonso-Blanco et al. (*Methods in Molecular Biology*, vol. 82, "Arabidopsis Protocols", pp. 137-146, J. M. Martinez-

Zapater and J. Salinas, eds., c. 1998 by Humana Press, Totowa, N.J.) and for corn by Burr ("Mapping Genes with Recombinant Inbreds", pp. 249-254. In Freeling, M. and V. Walbot (Ed.), *The Maize Handbook*, c. 1994 by Springer-Verlag New York, Inc.: New York, N.Y., USA; Berlin Germany; Burr et al. *Genetics* (1998) 118: 519; Gardiner, J. et al., (1993) *Genetics* 134: 917). This procedure, however, is not limited to plants and is used for other organisms (such as yeast) or for individual cells.

[0048] The polynucleotides of the present invention are also used for simple sequence repeat (SSR) mapping. Rice SSR mapping is described by Morgante et al. (*The Plant Journal* (1993) 3: 165), Panaud et al. (*Genome* (1995) 38: 1170); Senior et al. (*Crop Science* (1996) 36: 1676), Taramino et al. (*Genome* (1996) 39: 277) and Ahn et al. (*Molecular and General Genetics* (1993) 241: 483-90). SSR mapping is achieved using various methods. In one instance, polymorphisms are identified when sequence specific probes contained within a polynucleotide flanking an SSR are made and used in polymerase chain reaction (PCR) assays with template DNA from two or more individuals of interest. Here, a change in the number of tandem repeats between the SSR-flanking sequences produces differently sized fragments (U.S. Pat. No. 5,766,847). Alternatively, polymorphisms are identified by using the PCR fragment produced from the SSR-flanking sequence specific primer reaction as a probe against Southern blots representing different individuals (U. H. Refseth et al., (1997) *Electrophoresis* 18: 1519).

[0049] The polynucleotides of the invention can further be used to identify certain genes or genetic traits using, for example, known AFLP technologies, such as in EP0534858 and U.S. Pat. No. 5,878,215.

[0050] The polynucleotides of the present invention are also used for single nucleotide polymorphism (SNP) mapping.

[0051] Genetic and physical maps of crop species have many uses. For example, these maps are used to devise positional cloning strategies for isolating novel genes from the mapped crop species. In addition, because the genomes of closely related species are largely syntenic (i.e. they display the same ordering of genes within the genome), these maps are used to isolate novel alleles from relatives of crop species by positional cloning strategies.

[0052] The various types of maps discussed above are used with the polynucleotides of the invention to identify Quantitative Trait Loci (QTLs). Many important crop traits, such as the solids content of tomatoes, are quantitative traits and result from the combined interactions of several genes. These genes reside at different loci in the genome, often times on different chromosomes, and generally exhibit multiple alleles at each locus. The polynucleotides of the invention are used to identify QTLs and isolate specific alleles as described by de Vicente and Tanksley (*Genetics* 134:585 (1993)). Once a desired allele combination is identified, crop improvement is accomplished either through biotechnological means or by directed conventional breeding programs (for review see Tanksley and McCouch, *Science* 277:1063 (1997)). In addition to isolating QTL alleles in present crop species, the polynucleotides of the invention are also used to isolate alleles from the corresponding QTL of wild relatives.

[0053] In another embodiment, the polynucleotides are used to help create physical maps of the genome of corn,

*Arabidopsis* and related species. Where polynucleotides are ordered on a genetic map, as described above, they are used as probes to discover which clones in large libraries of plant DNA fragments in YACs, BACs, etc. contain the same polynucleotide or similar sequences, thereby facilitating the assignment of the large DNA fragments to chromosomal positions. Subsequently, the large BACs, YACs, etc. are ordered unambiguously by more detailed studies of their sequence composition (e.g. Marra et al. (1997) *Genomic Research* 7:1072-1084) and by using their end or other sequences to find the identical sequences in other cloned DNA fragments. The overlapping of DNA sequences in this way allows building large contigs of plant sequences to be built that, when sufficiently extended, provide a complete physical map of a chromosome. Sometimes the polynucleotides themselves provide the means of joining cloned sequences into a contig. All scientific and patent publications cited in this paragraph are hereby incorporated by reference.

[0054] U.S. Pat. Nos. 6,287,778 and 6,500,614, both hereby incorporated by reference, describe scanning multiple alleles of a plurality of loci using hybridization to arrays of oligonucleotides. These techniques are useful for each of the types of mapping discussed above.

[0055] Following the procedures described above and using a plurality of the polynucleotides of the present invention, any individual is genotyped. These individual genotypes are used for the identification of particular cultivars, varieties, lines, ecotypes and genetically modified plants or can serve as tools for subsequent genetic studies involving multiple phenotypic traits.

[0056] Identification and isolation of orthologous genes from closely related species and alleles within a species is particularly desirable because of their potential for crop improvement. Many important crop traits, result from the combined interactions of the products of several genes residing at different loci in the genome. Generally, alleles at each of these loci make quantitative differences to the trait. Once a more favorable allele combination is identified, crop improvement is accomplished either through biotechnological means or by directed conventional breeding programs (Tanksley et al. *Science* 277:1063(1997)).

#### 4. Use of the Genes to Make Transgenic Plants

[0057] To use the sequences of the present invention or a combination of them or parts and/or mutants and/or fusions and/or variants of them, recombinant DNA constructs are prepared which comprise the polynucleotide sequences of the invention inserted into a vector, and which are suitable for transformation of plant cells. The construct is made using standard recombinant DNA techniques (Sambrook et al. 1989) and is introduced to the species of interest by *Agrobacterium*-mediated transformation or by other means of transformation as referenced below.

[0058] The vector backbone is any of those typical in the art such as plasmids (such as Ti plasmids), viruses, artificial chromosomes, BACs, YACs and PACs and vectors of the sort described by

[0059] (a) BAC: Shizuya et al., *Proc. Natl. Acad. Sci. USA* 89: 8794-8797 (1992); Hamilton et al., *Proc. Natl. Acad. Sci. USA* 93: 9975-9979 (1996);

[0060] (b) YAC: Burke et al., *Science* 236:806-812 (1987);

[0061] (c) PAC: Sternberg N. et al., *Proc Natl Acad Sci USA*. January;87(1):103-7 (1990);

[0062] (d) Bacteria-Yeast Shuttle Vectors: Bradshaw et al., *Nucl Acids Res* 23: 4850-4856 (1995);

[0063] (e) Lambda Phage Vectors: Replacement Vector, e.g., Frischauf et al., *J. Mol Biol* 170: 827-842 (1983); or Insertion vector, e.g., Huynh et al., In: Glover N M (ed) *DNA Cloning: A practical Approach*, Vol. 1 Oxford: IRL Press (1985); T-DNA gene fusion vectors: Walden et al., *Mol Cell Biol* 1: 175-194 (1990); and

[0064] (g) Plasmid vectors: Sambrook et al., *infra*.

[0065] Typically, the construct comprises a vector containing a sequence of the present invention with any desired transcriptional and/or translational regulatory sequences, such as promoters, UTRs, and 3' end termination sequences. Vectors can also include origins of replication, scaffold attachment regions (SARs), markers, homologous sequences, introns, etc. The vector may also comprise a marker gene that confers a selectable phenotype on plant cells. The marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron, glyphosate or phosphinotricin.

[0066] A plant promoter fragment is used that directs transcription of the gene in all tissues of a regenerated plant and/or is a constitutive promoter. Alternatively, the plant promoter directs transcription of a sequence of the invention in a specific tissue (tissue-specific promoter) or is otherwise under more precise environmental control (inducible promoter).

[0067] If proper polypeptide production is desired, a polyadenylation region at the 3'-end of the coding region is typically included. The polyadenylation region is derived from the natural gene, from a variety of other plant genes, or from T-DNA, synthesized in the laboratory.

#### Transformation

[0068] Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, e.g. Weising et al., *Ann. Rev. Genet.* 22:421 (1988); and Christou, *Euphytica*, v. 85, n.1-3:13-27, (1995).

[0069] The person skilled in the art knows processes for the transformation of monocotyledonous and dicotyledonous plants. A variety of techniques are available for introducing DNA into a plant host cell. These techniques comprise transformation of plant cells by DNA injection, DNA electroporation, use of bolistics methods, protoplast fusion and via T-DNA using *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*, as well as further possibilities, or other bacterial hosts for Ti plasmid vectors. See for example, Broothaerts et al., *Gene Transfer to Plants by Diverse Species of Bacteria*, *Nature*, Vol. 433, pp. 629-633, 10 Feb. 2005.

[0070] DNA constructs of the invention are introduced into the cell or the genome of the desired plant host by a variety of conventional techniques. For example, the DNA

construct is introduced using techniques such as electroporation, microinjection and polyethylene glycol precipitation of plant cell protoplasts or protoplast fusion. Electroporation techniques are described in Fromm et al. *Proc. Natl Acad. Sci. USA* 82:5824 (1985). Microinjection techniques are known in the art and well described in the scientific and patent literature. The plasmids do not have to fulfill specific requirements for use in DNA electroporation or DNA injection into plant cells. Simple plasmids such as pUC derivatives can be used.

[0071] The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *EMBO J.* 3:2717 (1984). Introduction of foreign DNA using protoplast fusion is described by Willmitzer (Willmitzer, L., 1993 *Transgenic plants*. In: *Biotechnology, A Multi-Volume Comprehensive Treatise* (H. J. Rehm, G. Reed, A. Pühler, P. Stadler, eds.), Vol. 2, 627-659, VCH Weinheim-New York-Basel-Cambridge).

[0072] Alternatively, the DNA constructs of the invention are introduced directly into plant tissue using ballistic methods, such as DNA particle bombardment. Ballistic transformation techniques are described in Klein et al. *Nature* 327:773 (1987). Introduction of foreign DNA using ballistics is described by Willmitzer (Willmitzer, L., 1993 *Transgenic plants*. In: *Biotechnology, A Multi-Volume Comprehensive Treatise* (H. J. Rehm, G. Reed, A. Pühler, P. Stadler, eds.), Vol. 2, 627-659, VCH Weinheim-New York-Basel-Cambridge).

[0073] DNA constructs are also introduced with the help of *Agrobacteria*. The use of *Agrobacteria* for plant cell transformation is extensively examined and sufficiently disclosed in the specification of EP-A 120 516, and in Hoekema (In: *The Binary Plant Vector System* Offsetdrukkerij Kanters B. V., Alblasterdam (1985), Chapter V), Fraley et al. (*Crit. Rev. Plant. Sci.* 4, 1-46) and DePicker et al. (*EMBO J.* 4 (1985), 277-287). Using this technique, the DNA constructs of the invention are combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host direct the insertion of the construct and adjacent marker(s) into the plant cell DNA when the cell is infected by the bacteria (McCormac et al., 1997, *Mol. Biotechnol.* 8:199; Hamilton, 1997, *Gene* 200:107; Salomon et al., 1984 *EMBO J.* 3:141; Herrera-Estrella et al., 1983 *EMBO J.* 2:987). *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary or co-integrate vectors, are well described in the scientific literature. See, for example Hamilton, C M., *Gene* 200:107 (1997); Müller et al. *Mol. Gen. Genet.* 207:171 (1987); Komari et al. *Plant J.* 10:165 (1996); Venkateswarlu et al. *Biotechnology* 9:1103 (1991) and Gleave, A P., *Plant Mol. Biol.* 20:1203 (1992); Graves and Goldman, *Plant Mol. Biol.* 7:34 (1986) and Gould et al., *Plant Physiology* 95:426 (1991).

[0074] For plant cell T-DNA transfer of DNA, plant organs, e.g. inflorescences, plant explants, plant cells that have been cultured in suspension or protoplasts are co-cultivated with *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes* or other suitable T-DNA hosts. Whole plants are regenerated from the infected plant material or seeds generated from infected plant material using a suitable medium that contains antibiotics or biocides for the selection

of transformed cells or by spraying the biocide on plants to select the transformed plants. Plants obtained in this way are then examined for the presence of the DNA introduced. The transformation of dicotyledonous plants via Ti-plasmid-vector systems and *Agrobacterium tumefaciens* is well established.

[0075] Monocotyledonous plants are also transformed by means of *Agrobacterium* based vectors (See. Chan et al., Plant Mol. Biol. 22 (1993), 491-506; Hiei et al., Plant J. 6 (1994), 271-282; Deng et al., Science in China 33 (1990), 28-34; Wilmink et al., Plant Cell Reports 11 (1992), 76-80; May et al., Bio/Technology 13 (1995), 486-492; Conner and Domisse; Int. J. Plant Sci. 153 (1992), 550-555; Ritchie et al., Transgenic Res. 2 (1993), 252-265). Maize transformation in particular is described in the literature (see, for example, WO95/06128, EP 0 513 849; EP 0 465 875; Fromm et al., Biotechnology 8 (1990), 833-844; Gordon-Kamm et al., Plant Cell 2 (1990), 603-618; Koziel et al., Biotechnology 11 (1993), 194-200). In EP 292 435 and in Shillito et al. (1989, Bio/Technology 7, 581) fertile plants are obtained from a mucus-free, soft (friable) maize callus. Prioli and Söndahl (1989, Bio/Technology 7, 589) also report regenerating fertile plants from maize protoplasts of the maize Cateto inbred line, Cat 100-1.

[0076] Other cereal species have also been successfully transformed, such as barley (Wan and Lemaux, see above; Ritala et al., see above) and wheat (Nehra et al., 1994, Plant J. 5, 285-297).

[0077] Alternatives to *Agrobacterium* transformation for plants are ballistics, protoplast fusion, electroporation of partially permeabilized cells and use of glass fibers (See Wan and Lemaux, Plant Physiol. 104 (1994), 37-48; Vasil et al., Bio/Technology 11 (1993), 1553-1558; Ritala et al., Plant Mol. Biol. 24 (1994), 317-325; Spencer et al., Theor. Appl. Genet. 79 (1990), 625-631)).

[0078] Introduced DNA is usually stable after integration into the plant genome and is transmitted to the progeny of the transformed cell or plant. Generally the transformed plant cell contains a selectable marker that makes the transformed cells resistant to a biocide or an antibiotic such as kanamycin, G 418, bleomycin, hygromycin, phosphinotricin or others. Therefore, the individually chosen marker should allow the selection of transformed cells from cells lacking the introduced DNA.

[0079] The transformed cells grow within the plant in the usual way (McCormick et al., 1986, Plant Cell Reports 5, 81-84) and the resulting plants are cultured normally. Transformed plant cells obtained by any of the above transformation techniques are cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker that has been introduced together with the desired nucleotide sequences.

[0080] Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture* in "Handbook of Plant Cell Culture," pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1988. Regeneration also occurs from

plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467 (1987). Regeneration of monocots (rice) is described by Hosoyama et al. (*Biosci. Biotechnol. Biochem.* 58:1500 (1994)) and by Ghosh et al. (*J. Biotechnol.* 32:1 (1994)). Useful and relevant procedures for transient expression are also described in U.S. application Ser. No. 60/537,070 filed on Jan. 16, 2004 and PCT Application No. PCT/US2005/001153 filed on Jan. 14, 2005.

[0081] After transformation, seeds are obtained from the plants and used for testing stability and inheritance. Generally, two or more generations are cultivated to ensure that the phenotypic feature is stably maintained and transmitted.

[0082] One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

[0083] The nucleotide sequences according to the invention generally encode an appropriate protein from any organism, in particular from plants, fungi, bacteria or animals. The sequences preferably encode proteins from plants or fungi. Preferably, the plants are higher plants, in particular starch or oil storing useful plants, such as potato or cereals such as rice, maize, wheat, barley, rye, triticale, oat, millet, etc., as well as spinach, tobacco, sugar beet, soya, cotton etc.

[0084] In principle, the process according to the invention can be applied to any plant. Therefore, monocotyledonous as well as dicotyledonous plant species are particularly suitable. The process is preferably used with plants that are interesting for agriculture, horticulture and/or forestry. Examples are vegetable plants such as cucumber, melon, pumpkin, eggplant, zucchini, tomato, spinach, cabbage species, peas, beans, etc., as well as fruits such as pears, apples, etc.

[0085] Thus, the invention has use over a broad range of plants, preferably higher plants, pertaining to the classes of Angiospermae and Gymnospermae. Plants of the subclasses of the Dicotyledonae and the Monocotyledonae are particularly suitable. Dicotyledonous plants belong to the orders of the Magnoliales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papaverales, Sarraceniales, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, Santales, Rafflesiales, Celastrales, Euphorbiales, Rhamniales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginiales, Scrophulariales, Campanulales, Rubiales, Dipsacales, and Asterales. Monocotyledonous plants belong to the orders of the Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales, and Orchidales. Plants belonging to the class of the Gymnospermae are Pinales, Ginkgoales, Cycadales and Gnetales.

[0086] The method of the invention is preferably used with plants that are interesting for agriculture, horticulture, biomass for bioconversion and/or forestry. Examples are tobacco, oilseed rape, sugar beet, potato, tomato, cucumber, pepper, bean, pea, citrus fruit, apple, pear, berries, plum, melon, eggplant, cotton, soybean, sunflower, rose, poinsettia, petunia, guayule, cabbage, spinach, alfalfa, artichoke, corn, wheat, rye, barley, grasses such as switch grass or turf grass, millet, hemp, banana, poplar, eucalyptus trees, conifers.

[0087] The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

[0088] Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.

---

### SEQUENCE LISTING

The patent application contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US20060143729A1>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

---

What is claimed is:

1. An isolated nucleic acid molecule comprising:
  - a) a nucleic acid having a nucleotide sequence which encodes an amino acid sequence exhibiting at least 85% sequence identity to an amino acid sequence in Sequence Listing;
  - b) a nucleic acid which is a complement of a nucleotide sequence according to paragraph (a);
  - c) a nucleic acid which is the reverse of the nucleotide sequence according to subparagraph (a), such that the reverse nucleotide sequence has a sequence order which is the reverse of the sequence order of the nucleotide sequence according to subparagraph (a); or
  - d) a nucleic acid capable of hybridizing to a nucleic acid according to any one of paragraphs (a)-(c), under conditions that permit formation of a nucleic acid duplex at a temperature from about 40° C. and 48° C. below the melting temperature of the nucleic acid duplex.
2. The isolated nucleic acid molecule according to claim 1, which has the nucleotide sequence according to any polynucleotide sequence in the Sequence Listing.
3. The isolated nucleic acid molecule according to claim 1, wherein said amino acid sequence comprises any polypeptide sequence in the Sequence Listing.
4. A vector construct comprising:
  - a) a first nucleic acid having a regulatory sequence capable of causing transcription and/or translation in a plant; and
  - b) a second nucleic acid having the sequence of the isolated nucleic acid molecule according to claim 1;
 

wherein said first and second nucleic acids are operably linked and wherein said second nucleic acid is heterologous to any element in said vector construct.
5. The vector construct according to claim 4, wherein said first nucleic acid is native to said second nucleic acid.
6. The vector construct according to claim 4, wherein said first nucleic acid is heterologous to said second nucleic acid.
7. A host cell comprising an isolated nucleic acid molecule according to claim 1 wherein said nucleic acid molecule is flanked by exogenous sequence.
8. A host cell comprising a vector construct according to claim 4.
9. An isolated polypeptide comprising an amino acid sequence exhibiting at least 85% sequence identity of an amino acid sequence of the Sequence Listing.
10. A method of introducing an isolated nucleic acid into a host cell comprising:
  - a) providing an isolated nucleic acid molecule according to claim 1; and
  - b) contacting said isolated nucleic acid with said host cell under conditions that permit insertion of said nucleic acid into said host cell.
11. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to claim 4.
12. A method for detecting a nucleic acid in a sample which comprises:
  - a) providing an isolated nucleic acid molecule according to claim 1;
  - b) contacting said isolated nucleic acid molecule with a sample under conditions which permit a comparison of the sequence of said isolated nucleic acid molecule with the sequence of DNA in said sample; and
  - c) analyzing the result of said comparison.
13. A host cell or organism which comprises a nucleic acid molecule according to claim 1 which is exogenous or heterologous to said plant or plant cell.
14. A host cell or organism which comprises a vector construct according to claim 4.
15. A host cell or organism according to claim 13, which is a plant, plant cell, plant material or seed of a plant.

**16.** A host cell or organism according to claim 14, which is a plant, plant cell, plant material or seed of a plant.

**17.** A plant which has been regenerated from a plant cell or seed according to claims **15**.

**18.** A plant which has been regenerated from a plant cell or seed according to claims **16**.

\* \* \* \* \*