



(86) Date de dépôt PCT/PCT Filing Date: 2005/07/27
(87) Date publication PCT/PCT Publication Date: 2006/02/09
(45) Date de délivrance/Issue Date: 2013/04/30
(85) Entrée phase nationale/National Entry: 2007/01/29
(86) N° demande PCT/PCT Application No.: US 2005/026641
(87) N° publication PCT/PCT Publication No.: 2006/015045
(30) Priorité/Priority: 2004/07/29 (US60/592,268)

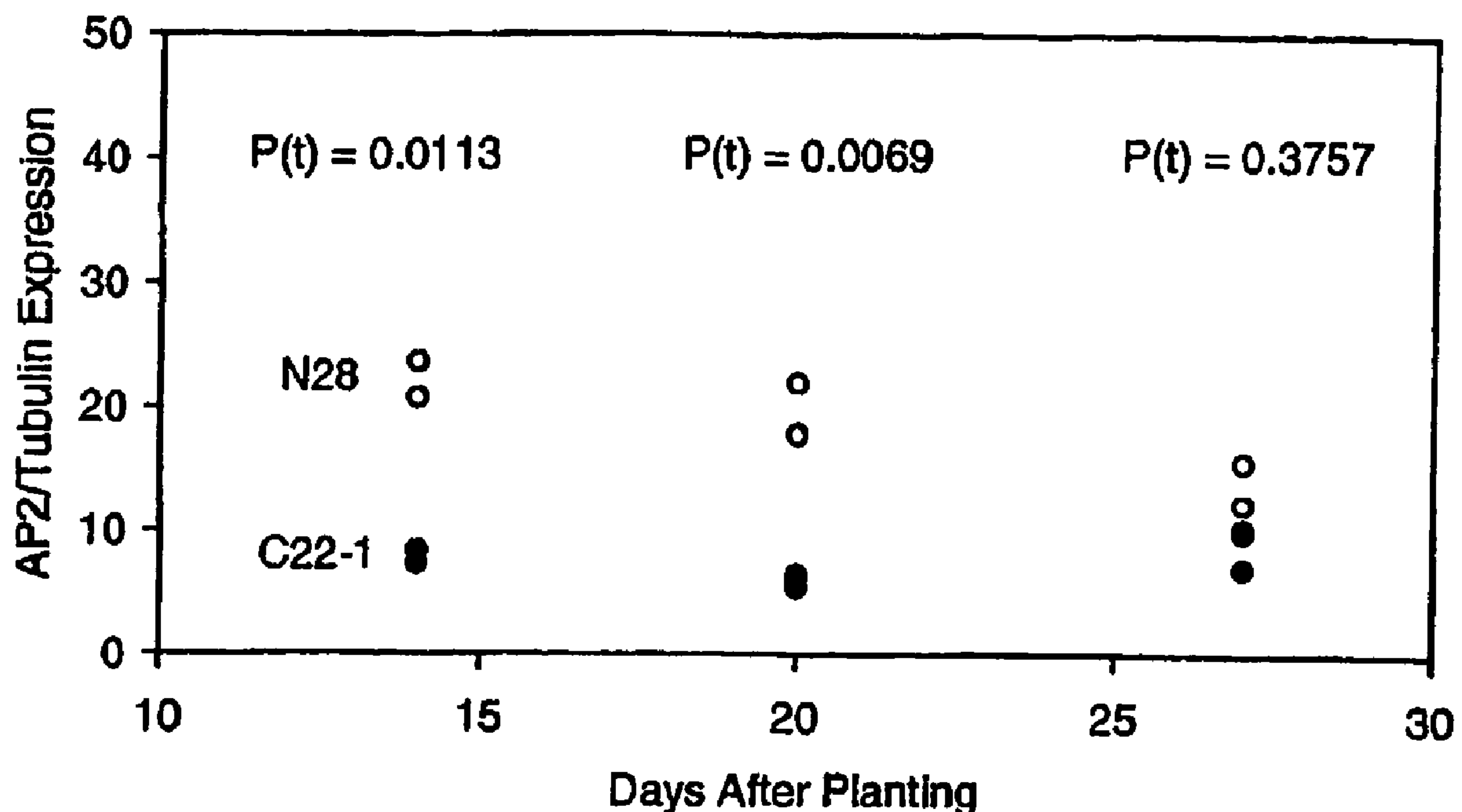
(51) Cl.Int./Int.Cl. *C12N 15/29* (2006.01),
A01H 5/00 (2006.01), *C07K 14/415* (2006.01),
C12N 15/82 (2006.01), *C12N 5/10* (2006.01)

(72) Inventeurs/Inventors:
TOMES, DWIGHT, US;
SALVI, SILVIO, IT;
MORGANTE, MICHELE, IT;
SPONZA, GIORGIO, IT;
BRUGGEMANN, EDWARD, US;
...

(73) Propriétaires/Owners:
PIONEER HI-BRED INTERNATIONAL, INC., US;
UNIVERSITY OF BOLOGNA, IT

(74) Agent: TORYS LLP

(54) Titre : PROCEDES ET COMPOSITIONS POUR MODULER LA FLORAISON ET LA MATURITE DES PLANTES
(54) Title: METHODS AND COMPOSITIONS FOR MODULATING FLOWERING AND MATURITY IN PLANTS



(57) Abrégé/Abstract:

The present invention provides compositions and methods for modulating flowering time in plants. Maize RAP2.7 nucleotide sequences are disclosed which upon overexpression cause later flowering and when inhibited cause earlier flowering. Also disclosed is a DNA sequence which acts as a regulator/enhancer of RAP2.7, termed VGT1. This sequence does not code for any known protein, but acts as either a RNAi element or a regulatory DNA or RNA element that either directly regulates expression of flowering genes such as Rap2.7 or specifically targets expression of other genes which control flowering genes such as Rap2.7. This element this can be used as a sequence-based marker to identify inbred and hybrids which have altered maturity. Methods for expressing these nucleotide sequences in a plant for modifying maturity and flowering in plants are provided as well as expression constructs, vectors, transformed cells and plants.

(72) Inventeurs(suite)/Inventors(continued): NIU, XIAOMU, US; LI, BAILIN, US; TUBEROSA, ROBERTO, IT

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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
9 February 2006 (09.02.2006)

PCT

(10) International Publication Number
WO 2006/015045 A2

(51) International Patent Classification:

C12N 15/82 (2006.01) *C07K 14/00* (2006.01)
C12N 15/29 (2006.01)

(21) International Application Number:

PCT/US2005/026641

(22) International Filing Date: 27 July 2005 (27.07.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/592,268 29 July 2004 (29.07.2004) US

(71) Applicants (for all designated States except US): **PIO-NEER HI-BRED INTERNATIONAL, INC.** [US/US]; 9550 White Oak Lane, Suite 100, Johnston, IA 50131-1014 (US). **UNIVERSITY OF BOLOGNA** [IT/IT]; Department of Agroenvironmental Sciences and Technology, Via Fanin, 44, I-40127 Bologna (IT).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **TOMES, Dwight** [US/US]; 2672 Brookview Lane, Van Meter, IA 50261 (US). **SALVI, Silvio**, [IT/IT]; Department of Agroenvironmental Sciences and Technology, Via Fanin, 44, I-40127 Bologna (IT). **MORGANTE, Michele** [IT/IT]; University

of Udine, Via delle Scienze 208, I-33100 Udine (IT). **SPONZA, Giorgio** [IT/IT]; Department of Agroenvironmental Sciences and Technology, Via Fanin, 44, I-40127 Bologna (IT). **BRUGGEMANN, Edward** [US/US]; 4000 Stonebridge Road, West Des Moines, IA 50265 (US). **NIU, Xiaomu** [CN/US]; 6473 N. Winwood Drive, Johnston, IA 50131 (US). **LI, Bailin** [US/US]; 549 Stenning Drive, Hockessin, DE 19707 (US). **TUBEROSA, Roberto** [IT/IT]; Department of Agroenvironmental Sciences and Technology, Via Fanin, 44, I-40127 Bologna (IT).

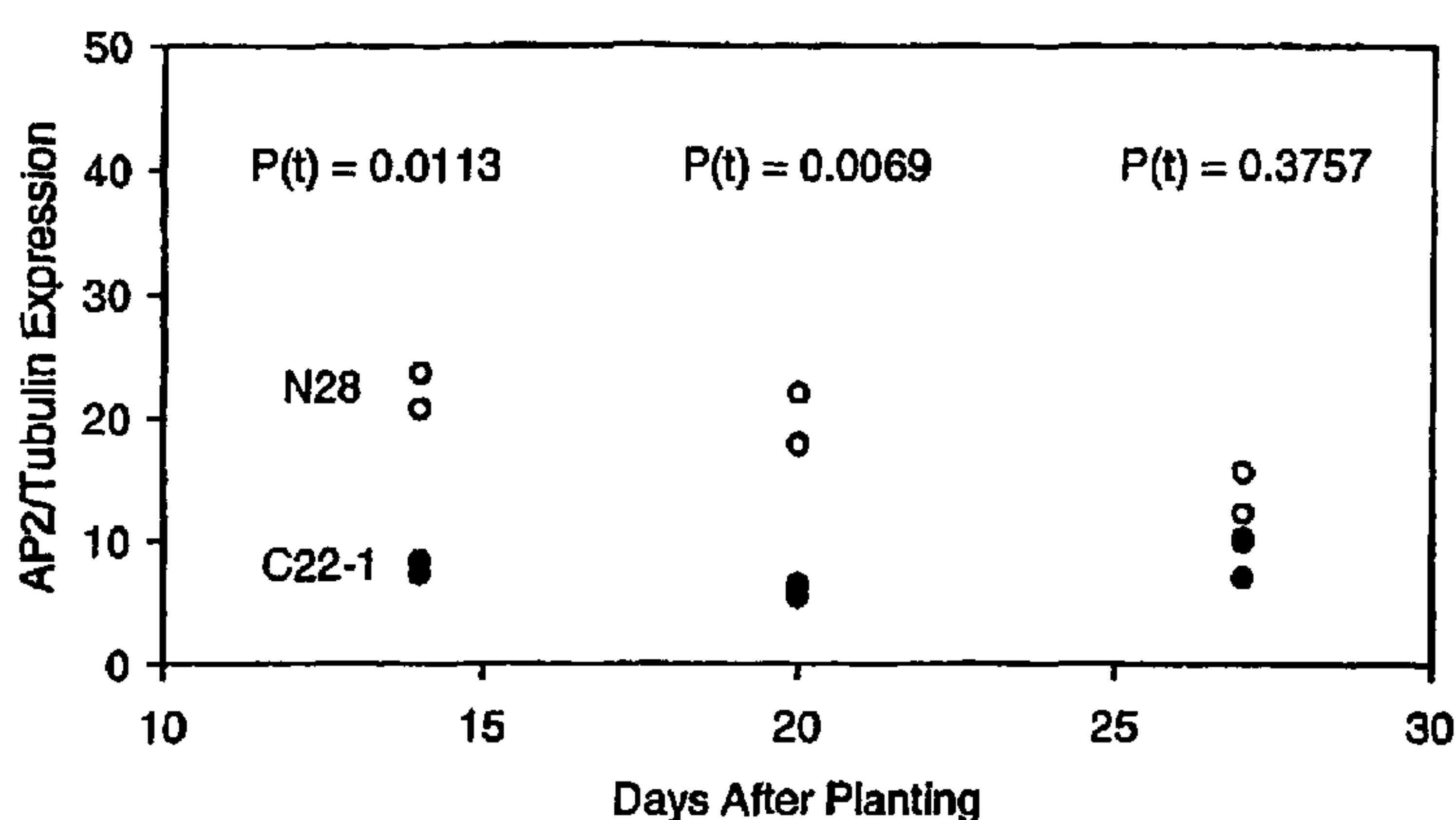
(74) Agent: **NEBEL, Heidi, S.**; McKee, Voorhees & Sease, PLC, 801 Grand Avenue, Suite 3200, Des Moines, IA 50309 (US).

(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,

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS FOR MODULATING FLOWERING AND MATURITY IN PLANTS



(57) Abstract: The present invention provides compositions and methods for modulating flowering time in plants. Maize RAP2.7 nucleotide sequences are disclosed which upon overexpression cause later flowering and when inhibited cause earlier flowering. Also disclosed is a DNA sequence which acts as a regulator/enhancer of RAP2.7, termed VGT1. This sequence does not code for any known protein, but acts as either a RNAi element or a regulatory DNA or RNA element that either directly regulates expression of flowering genes such as Rap2.7 or specifically targets expression of other genes which control flowering genes such as Rap2.7. This element this can be used as a sequence-based marker to identify inbred and hybrids which have altered maturity. Methods for expressing these nucleotide sequences in a plant for modifying maturity and flowering in plants are provided as well as expression constructs, vectors, transformed cells and plants.

WO 2006/015045 A2

WO 2006/015045 A2

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, LV, 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).

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.

DEMANDES OU BREVETS VOLUMINEUX

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

CECI EST LE TOME __1__ DE __2__

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__

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

Methods and Compositions for Modulating Flowering and Maturity in Plants

5

FIELD OF THE INVENTION

This invention is related to compositions and methods for affecting flowering time
10 in plants.

BACKGROUND OF THE INVENTION

Plants have two basic growth modes during their life cycles --- vegetative growth
and flower and seed growth. Above ground vegetative growth of the plant develops from
15 the apical meristem. This vegetative meristem gives rise to all of the leaves that are found
on the plant. The plant will maintain its vegetative growth pattern until the apical
meristem undergoes a change. This change actually alters the identity of the meristem
from a vegetative to an inflorescence meristem. The inflorescence meristem produces
small leaves before it next produces floral meristems. It is the floral meristem from which
20 the flower develops.

From a genetic perspective, two phenotypic changes that control vegetative and
floral growth are programmed in the plant. The first genetic change involves the switch
from the vegetative to the floral state. If this genetic change is not functioning properly,
then flowering will not occur. The second genetic event follows the commitment of the
25 plant to form flowers. The observation that the organs of the plant develop in a sequential
manner suggests that a genetic mechanism exists in which a series of genes are
sequentially turned on and off.

Flowering time is an important agronomic trait in cultivated plant species as it
determines in large measure the growing region of adaptation. Most angiosperm species
30 are induced to flower in response to environmental stimuli such as day length and
temperature, and internal cues, such as age. Genetic analysis revealed that there are several
types of mutants that alter flowering time.

Studies of two distantly related dicotyledons, *Arabidopsis thaliana* and *Antirrhinum*
majus, has led to the identification of three classes of homeotic genes, acting alone or in

combination to determine floral organ identity (Bowman, et al., Development, 112:1, 1991; Carpenter and Coen, Genes Devl., 4:1483, 1990; Schwarz-Sommer, et al., Science, 250:931, 1990). Several of these genes are transcription factors whose conserved DNA-binding domain has been designated the MADS box (Schwarz-Sommer, et al., supra).

5 Earlier acting genes that control the identity of flower meristems have also been characterized. Flower meristems are derived from inflorescence meristems in both Arabidopsis and Antirrhinum. Two factors that control the development of meristematic cells into flowers are known. In Arabidopsis, the factors are the products of the LEAFY *gene* (Weigel, et al., Cell 69:843, 1992) and the APETALA1 *gene* (Mandel, et al., Nature
10 360:273, 1992). When either of these genes is inactivated by mutation, structures combining the properties of flowers and inflorescence develop (Weigel, et al., supra; Irish and Sussex, Plant Cell, 2:741, 1990). In Antirrhinum, the homologue of the Arabidopsis LEAFY *gene* is FLORICAULA (Coen, et al., Cell, 63:1311, 1990) and that of the APETALA1 *gene* is SQUAMOSA (Huijser, et al., EMBO J., 11:1239, 1992). The latter
15 pair contains MADS box domains.

Genetic studies in *Arabidopsis thaliana* have identified five genes (APETALA1 (AP1), APETALA2 (AP2), APETALA3 (AP3), PISTILLATA (PI), and AGAMOUS (AG)) that are involved in the specification of floral organ identity. Mutations in these genes result in homeotic transformation of one organ type into another, much like the
20 homeotic selector genes in animal development. These five genes act in spatially localized domains in a flower and in different combinations to specify the development of the sepals, petals, stamens, and carpels. All five genes encode proteins which appear to function as transcription factors. Four of these proteins are members of the MADS domain family of dimeric transcription factors. MADS domain proteins are found in many
25 organisms including yeast, mammals, insects, and plants. The fifth protein, AP2, is a member of another class of DNA binding proteins which may be unique to plants

APETALA2 (AP2) plays an important role in the control of Arabidopsis flower and seed development and encodes a putative transcription factor that is distinguished by a novel DNA binding motif referred to as the AP2 domain. The AP2 domain containing or
30 RAP2 (related to AP2) family of proteins is encoded by a minimum of 12 genes in *Arabidopsis*. The *RAP2* genes encode two classes of proteins, AP2-like and EREBP-like, that are defined by the number of AP2 domains in each polypeptide as well as by two sequence motifs referred to as the YRG and RAYD elements that are located within each AP2 domain. *RAP2* genes are differentially expressed in flower, leaf, inflorescence stem,

and root. Moreover, the expression of at least three *RAP2* genes in vegetative tissues are controlled by *AP2*. Thus, unlike other floral homeotic genes, *AP2* is active during both reproductive and vegetative development.

Maize is a monocotyledonous plant species and belongs to the grass family. It is
5 unusual for a flowering plant as it has unisexual inflorescences. The male inflorescence (tassel) develops in a terminal position, whereas the female inflorescences (ears) grow in the axil of vegetative leaves. The inflorescences, as typical for grasses, are composed of spikelets. In the case of maize each spikelet contains two florets (the grass flower) enclosed by a pair of bracts (inner and outer glume). A number of genes have been
10 identified which modify flowering time in maize including *Id1* and *DLF*.

There is increasing incentive by those working in the field of plant biotechnology to successfully genetically engineer plants, including the major crop varieties. One genetic modification that would be economically desirable would be to accelerate the flowering time of a plant. Induction of flowering is often the limiting factor for growing crop plants.
15 One of the most important factors controlling induction of flowering is day length, which varies seasonally as well as geographically. There is a need to develop a method for controlling and inducing flowering in plants, regardless of the locale or the environmental conditions, thereby allowing production of crops, at any given time. Since most crop products (e.g. seeds, grains, fruits), are derived from flowers, such a method for controlling
20 flowering would be economically invaluable.

It is an object of the present invention to provide methods and compositions for affecting flowering time in plants.

It is yet another object of the invention to provide novel nucleotide sequences isolated from maize which encode proteins which affect flowering time in plants.

25 It is yet another object of the invention to provide maize *RAP2.7* genes which affect flowering time and internode length in maize.

It is yet another object of the invention to provide DNA regulatory factors which enhance/inhibit the ability of *RAP2.7* to regulate flowering time.

It is yet another object of the invention to provide methods and compositions
30 including nucleotide constructs, vectors, transgenic cells and plants with altered flowering characteristics as described herein.

It is yet another object of the invention to provide markers for identification of mutant plants which may have altered flowering time by the presence of marker *VGT1* sequences.

SUMMARY OF THE INVENTION

Compositions and methods involved in the modulation of flowering in plants are provided. The compositions include nucleic acid molecules isolated from maize which encode RAP2.7 proteins. Amino acid sequences of these proteins are also provided. Further, polynucleotides having nucleic acid sequences encoding maize RAP2.7 proteins are also provided. These proteins and the nucleotide sequences encoding them provide an opportunity to manipulate maturity of plants. When polynucleotide sequences encoding the RAP2.7 gene product are overexpressed flowering time is delayed, and when the product is inhibited flowering is earlier.

Typically maturity of a given plant is changed with crossing of earlier and later maturity plants. Agronomic traits such as yield, or lodging resistance are often lost in the process. The compositions of the invention provide for the ability to make significant changes in maturity while keeping other vegetative and reproductive characteristics similar using a transgenic approach.

The invention includes methods for manipulating the maturity of plants using polynucleotide sequences that were isolated from maize (*Zea mays*). These sequences alone, or in combination with other sequences, can be used to control plant maturity and thus area of adaptation. In another aspect of the present invention, nucleotide constructs such as expression cassettes and transformation vectors comprising the isolated nucleotide sequences are disclosed. The transformation vectors can be used to transform plants and express the flower modulation control genes in the transformed cells. In this manner, the maturity of plants as well as area of adaptation can be controlled. Transformed cells as well as regenerated transgenic plants and seeds containing and expressing the isolated polynucleotide sequences and protein products are also provided.

Also according to the invention, a novel DNA sequence termed (VGT1) has been identified which is a 2kb region on maize chromosome 8L. VGT1 acts as a CIS interaction type non-coding RNA sequence for maize RAP2.7. In mutants with early flowering, VGT1 may interact or repress the expression level of RAP2.7 causing down regulation of RAP2.7 and early flowering. Thus the invention also comprises nucleotide sequences encoding a VGT1 DNA factor which interacts either directly or indirectly with RAP2.7 in modulating the flowering time in plants. The invention includes nucleotide sequences, polymorphisms, "expression-type" constructs with the VGT1 sequences operably linked to promoters regions for transcription of the same, transgenic cells and plants with altered flowering time. The

VGT1 sequences and the alternate forms thereof may also be used as markers to identify plants with flowering that may be different from wild type.

For any of the sequences disclosed herein, the polynucleotide of the invention or at least 20 contiguous bases therefrom may be used as probes to isolate and identify similar genes in other plant species. The sequences disclosed may also be used to isolate regulatory elements and promoter sequences that are natively associated with the polynucleotides disclosed herein to give spatial and temporal expression of operatively linked sequences to flowering in plants.

This invention relates to the following:

<1> An isolated nucleic acid molecule that encodes a polypeptide having RAP2.7- like activity, wherein said nucleic acid molecule is:

- (a) a nucleic acid molecule comprising the sequence set forth in SEQ ID NOS: 1, 3, or 4;
- (b) a nucleic acid molecule comprising a sequence encoding the amino acid sequence set forth in SEQ ID NO: 2;
- (c) a nucleic acid molecule comprising a sequence having at least 90% sequence identity to the full length nucleotide sequence set forth in SEQ ID NOS: 1, 3, or 4; or
- (d) a nucleic acid molecule comprising a sequence of at least 1000 consecutive nucleic acids of any of (a) through (c),
wherein RAP2.7-like activity comprises modulation of onset of flowering time in plants.

<2> A vector comprising the nucleic acid molecule of <1>.

<3> A transgenic plant cell having stably incorporated in its genome the isolated nucleic acid molecule of <1>.

<4> The plant cell of <3>, wherein said plant cell is from a monocot plant.

<5> The plant cell of <4>, wherein said monocot plant is maize.

<6> A method for altering flowering time in a plant comprising:

(i) transforming a plant cell with a vector comprising a nucleic acid molecule operably linked to a promoter that regulates transcription of said nucleic acid molecule in a plant cell; wherein said nucleic acid molecule comprises:

- (a) a nucleotide sequence comprising the sequence set forth in SEQ ID NOS: 1, 3, or 4;
- (b) a nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 2; or
- (c) a nucleotide sequence having at least 90% sequence identity to the full length sequence of SEQ ID NOS: 1, 3, or 4, wherein said nucleotide sequence encodes a protein which regulates flowering in plants and

(ii) regenerating a stably transformed plant from the plant cell.

<7> The method of <6>, wherein said plant is a monocot.

<8> The method of <7>, wherein said monocot is maize.

<9> The method of <6> wherein said vector over- expresses RAP2.7.

<10> The method of <9> wherein said plant flowers later than a plant which does not have the vector incorporated therein.

<11> The method of <6> wherein said vector inhibits expression of RAP2.7.

<12> The method of <11> wherein said vector is an RNA interference vector.

<13> The method of <12> wherein said plant flowers earlier than a plant which does not have the RNA interference vector incorporated therein.

<14> A method for altering onset of maturity of a plant, said method comprising: transforming said plant with a vector comprising a heterologous nucleic acid molecule operably linked to a promoter that induces transcription of said heterologous nucleic acid molecule in a plant cell; and regenerating stably transformed plants, wherein said heterologous nucleic acid molecule comprises:

- (a) a nucleotide sequence comprising the sequence set forth in SEQ ID NOS: 1, 3 or 4;
- (b) a nucleotide sequence comprising at least 900 contiguous nucleotides of the sequence of SEQ ID NOS: 1, 3, or 4, wherein said nucleotide sequence encodes a protein having flowering regulatory activity; or
- (c) a nucleotide sequence having at least 90% sequence identity to the full length sequence of SEQ ID NOS: 1, 3, or 4, wherein said nucleotide sequence encodes a protein having flowering regulatory activity.

<15> The method of <14>, wherein said promoter is a constitutive promoter.

<16> The method of <14>, wherein said promoter is a tissue-preferred promoter.

<17> The method of <14>, wherein said promoter is an inducible promoter.

<18> The method of <14>, wherein said plant is a monocot.

<19> The method of <18>, wherein said monocot is maize.

<20> The method of <14> wherein said vector over- expresses RAP2.7.

<21> The method of <20> wherein said plant matures later than a plant which does not have the vector incorporated therein.

<22> The method of <14> wherein said vector inhibits expression of RAP2.7.

<23> The method of <22> wherein said vector is an RNA interference vector.

<24> The method of <22> wherein said plant matures earlier than a plant which does not have the RNA interference vector incorporated therein.

<25> An isolated polypeptide having flowering regulatory activity, wherein the polypeptide is:

- (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2;

- (b) a polypeptide encoded by a nucleic acid molecule comprising the sequence set forth in SEQ ID NOS: 1, 3, or 4;
- (c) a polypeptide encoded by a nucleic acid molecule that has at least 80% sequence identity to the full length sequence set forth in SEQ ID NOS: 1, 3, or 4;
- (d) a polypeptide comprising an amino acid sequence having at least 90% sequence identity to the full length sequence set forth in SEQ ID NO:2; or
- (e) a polypeptide comprising an amino acid sequence of at least 200 consecutive amino acids of any of (a) through (d).

<26> A nucleotide construct comprising:

a nucleic acid molecule encoding the polypeptide of parts (a) to (d) of <25>, wherein said nucleic acid molecule is operably linked to a promoter that drives expression in a host cell.

<27> A plant cell having stably incorporated in its genome the nucleotide construct of <26>.

<28> The plant cell of <27>, wherein said plant cell is from a monocot plant.

<29> The plant cell of <28>, wherein said monocot plant is maize.

DESCRIPTION OF THE FIGURES

Figure 1 is a graph depicting the levels of RAP2.7 expression at Day: 14, day 20 and day 27.

Figure 2 is an illustration of the over-expression vector used in transformation, showing the portion between the right and left T-DNA borders (RB, LB). The transformation vector for RAP2.7 over-expression, PHP20922, was created by electroporating the JT vector PHP20921 into *Agrobacterium*. The Invitrogen (Carlsbad, California) Gateway technology was used to create PHP20921. Specifically, the RAP2.7 coding region was first amplified by PCR with 5'-primer (ggggacaagttgtacaaaaagcaggctatgcagttggatctgaacgt) and 3'primer (ggggaccactttgtacaagaaagctgggttcagcggggatggtgatg). These primers contain Gateway attB recombination sites. The PCR product was confirmed by sequencing and cloned into a Gateway vector pDONR221 *via* a BP recombination reaction as described by the vendor (Invitrogen, Carlsbad, California). This resulted in the entry clone, PHP20923. Two other entry clones, PHP20830 containing the rice actin promoter, PHP20234 containing the pinII terminator, have been previously created using similar cloning strategies. A destination vector PHP20909 was also created from pDESTR4-R3 vector (Invitrogen, Carlsbad, California) by inserting an expression cassette of the CaMV35S promoter driving the herbicide resistance gene Bar followed by a pinII terminator. The four vectors, PHP20923, 20830, 20234 and 20909, were then used to create the JT vector, PHP20921, *via* a LR recombination reaction following vendor's instructions (Invitrogen, Carlsbad, California).

Figure 3 is the deduced amino acid sequence of RAP2.7 from C22-4 allele. The two putative AP2 domains are highlighted in bold, whereas the linker region between them

is italicized. The well conserved YRG and RAYD motifs are underlined, although there is an R to K substitution in the second RAYD motif.

Figure 4 is a GAP alignment between the genomic RAP2.7 from B73 compared to the sequence of RAP2.7 of MO17. Gap Weight: 50; Average Match: 10.000; Length Weight: 3; Average Mismatch: 0.000; Quality: 59308; Length:8586; Ratio: 7.730; Gaps: 41; Percent Similarity: 87.482, Percent Identity: 87.482. Vladutu, C., McLaughlin J. and Phillips R.L. 1999. Fine Mapping and Characterization of Linked Quantitative Trait Loci Involved in the Transition of the Maize Apical Meristem From Vegetative to Generative Structures. Genetics 153: 993-1007.

Figure 5 is an illustration of construction of the RAP2.7 gene fragments for the RNA interference vector. Fragment TR1 (Truncated1) was created by PCR using forward (ggatccgatctgaacgtggccgag) and reverse primers (gaattcctaggcagctgttcttctctctttg) corresponding to roughly 2/3 of the coding sequence starting from 9 bp downstream of ATG. Fragment IR1 (Inverted1) was generated similarly by PCR with forward (gcggccgcatctgaacgtggccgag) and reverse primers (gaattctgtgggactcccagcggcctgtgc) starting from the same position as TR1, although IR1 is only half the length of TR1.

Figure 6 is an illustration of the vector used in transformation showing only the portion between the right and left T-DNA borders (RB, LB). As described in Figure 5, fragments corresponding to truncated coding regions to be used for the RNA interference construct were generated by as TR1, flanked by BamH1 and EcoR1, and IR1, flanked by NotI and EcoR1 restriction sites. Vector PHP16501 containing the rice actin promoter was linearized with NotI and BamH1. In a 3-piece ligation, TR1 was cloned in downstream of NotI followed by IR1 in reverse orientation. The resulting cassette vector, PHP21767, was then cloned into a JT vector containing the CaMV35S promoter driving the herbicide resistance gene Bar. The transformation vector for RAP2.7 down regulation, PHP21842, was generated by electroporating PHP21798 into Agrobacterium.

Figure 7 is the data showing the fine genetic and physical mapping of VGT1. First row indicates physical distance (in kb) from Rap2.7, based on sequence derived from the relevant Mo17 BAC clone from library. Second row indicates the type of molecular marker. Third row indicates the name of the molecular markers. Rows from 4 to 21 indicate the genotype of parental lines (N28 and C22-4) and of the 17 segmental QTL-Nearly Isogenic Lines. The VGT1 column shows where the QTL was mapped. The last two columns provide the phenotypic scores for DPS (Days to Pollen Shed) and ND (plant node number).

Figure 8 is a graph showing the results of association mapping for markers developed at the Vgt1-Rap2.7 region and phenotypic data for flowering time collected in Bologna (2002 and 2003) among a set of 96 maize inbred lines. Statistical association is expressed as P from ANOVA tests.

5 Figure 9 is an alignment of the VGT1 sequences from all four lines, including a consensus sequence line N28 was identical to B73.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides, *inter alia*, compositions and methods for
10 manipulating flowering time in plants. As used herein the term "flowering time" or maturity shall mean the time at which a plant reaches physiological maturity and is capable of reproducing.

The compositions comprise nucleic acid molecules comprising sequences of plant genes and the polypeptides encoded thereby as well as regulatory factors which are non-
15 coding. Particularly, the nucleotide and amino acid sequences for a maize RAP2.7 (related to AP2 domain containing) gene are provided. Three RAP2.7 encoding nucleotide sequences are provided at SEQ ID NOS: 1 (cDNA), 3 (genomic), and 4 (genomic) with the corresponding protein at SEQ ID NO: 2. Three VGT1 nucleotide sequences are provided at SEQ ID NOS: 5, 6, and 7 with the consensus sequence at SEQ ID NO:8. As discussed in more detail below, the
20 sequences of the invention are involved in many basic biochemical pathways that regulate flowering time and maturity in plants. Thus, methods are provided for the expression of these sequences in a host plant to modulate plant flowering. Some of the methods involve stably transforming a plant with a nucleotide sequence capable of modulating plant flowering operably linked with a promoter capable of driving expression (or transcription) of a
25 nucleotide sequence in a plant cell.

Promoter and other regulatory elements which are natively associated with these genes can be easily isolated using the sequences and methods described herein with no more than routine experimentation. These sequences can be used to identify promoter, enhancer or other signaling signals in the regulatory regions of RAP2.7 encoding sequences. These regulatory
30 and promoter elements provide for temporal and spatial expression of operably linked sequences with flowering in a plant. Methods are provided for the regulated expression of a nucleotide sequence of interest that is operably linked to the promoter regulatory sequences disclosed herein. Nucleotide sequences operably linked to the promoter sequences are transformed into a plant cell. Exposure of the transformed plant to a stimulus such as the

timing of flowering induces transcriptional activation of the nucleotide sequences operably linked to these promoter regulatory sequences.

By "heterologous nucleotide sequence" is intended a sequence that is not naturally occurring with the referenced sequence. While the referenced nucleotide sequence is
5 heterologous to the promoter sequence or vice versa, it may be homologous, or native, or heterologous, or foreign, to the plant host.

By "operably linked" is intended a functional linkage between a promoter sequence and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means
10 that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame. In the case of a DNA regulatory factor the nucleic acid sequences are transcribed only.

A polypeptide is said to have RAP2.7-like activity when it has one or more of the properties of the native protein. It is within the skill in the art to assay protein activities
15 obtained from various sources to determine whether the properties of the proteins are the same. In so doing, one of skill in the art may employ any of a wide array of known assays including, for example, biochemical and/or pathological assays. For example, one of skill in the art could readily produce a plant transformed with a RAP2.7 polypeptide variant and assay a property of native RAP2.7 protein in that plant material to determine whether a particular RAP2.7
20 property was retained by the variant.

The compositions and methods of the invention are involved in biochemical pathways and as such may also find use in the activation or modulation of expression of other genes, including those involved in other aspects of flowering time.

Although there is some conservation among these genes, proteins encoded by
25 members of these gene families may contain different elements or motifs or sequence patterns that modulate or affect the activity, subcellular localization, and/or target of the protein in which they are found. Such elements, motifs, or sequence patterns may be useful in engineering novel enzymes for modulating gene expression in particular tissues. By "modulating" or "modulation" is intended that the level of expression of a gene may be
30 increased or decreased relative to genes driven by other promoters or relative to the normal or uninduced level of the gene in question.

According to the invention, overexpression of maize RAP2.7 caused flowering that was later than normal in plants as well as an increased number of leaves (nodes) produced prior to flowering resulting in taller plant stature. Inhibition of maize RAP2.7 caused

maturation or flowering that was earlier than normal. Also the presence of mutant VGT1 correlated with earlier flowering as well, thus leading to the concept that VGT1 acts as direct or indirect enhancer/regulator of RAP2.7. Expression of the proteins encoded by RAP2.7 encoding sequences or transcription of the VGT1 regulatory element can be used to modulate or regulate the expression of proteins in these flowering pathways and other directly or indirectly affected pathways. Hence, the compositions and methods of the invention find use in altering plant flowering and maturity. In other embodiments, fragments of the sequences are used to confer desired properties to synthetic constructs for use in regulating plant maturity and flowering.

The present invention provides for isolated nucleic acid molecules comprising nucleotide sequences encoding the amino acid sequence shown in SEQ ID NOS: 2 as well as their conservatively modified variants. Further provided are polypeptides having an amino acid sequence encoded by a nucleic acid molecule described herein, for example those polypeptides comprising the sequences set forth in SEQ ID NO: 1, 3 and 4, and fragments and variants thereof.

The present invention further provides for an isolated nucleic acid molecule comprising the sequences shown in SEQ ID NO: 1, 3, 4, 5, 6, 7, or 8.

The invention encompasses isolated or substantially purified nucleic acid or protein compositions. An "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In some embodiments, an "isolated" nucleic acid is free of sequences (such as other protein-encoding sequences) that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, 0.4 kb, 0.3 kb, 0.2 kb, or 0.1 kb, or 50, 40, 30, 20, or 10 nucleotides that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, 5%, (by dry weight) of contaminating protein. When the protein of the invention or biologically active portion thereof is recombinantly produced, culture medium may represent less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein-of-interest chemicals.

Fragments and variants of the disclosed nucleotide sequences which retain the functional properties of the encoded peptide or of the non-coding RNA are encompassed by the present invention. By "fragment" is intended a portion of the nucleotide sequence or a portion of the amino acid sequence and hence protein encoded thereby. Fragments of a nucleotide sequence may encode protein fragments that retain the biological activity of the native protein and hence affect flowering by retaining RAP2.7-like activity or may include portions of non-coding regulatory element which retain the RAP2.7 modulating activity of VGT1. Alternatively, fragments of a nucleotide sequence that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Thus, fragments of a nucleotide sequence may range from at least about 20 nucleotides, about 50 nucleotides, about 100 nucleotides, and up to the full-length nucleotide sequence encoding the proteins or regulating RAP2.7 of the invention.

A fragment of a RAP2.7 nucleotide sequence that encodes a biologically active portion of a RAP2.7 protein of the invention will encode at least 12, 25, 30, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, or 450, contiguous amino acids, or up to the total number of amino acids present in a full-length RAP2.7 protein of the invention.

A fragment of a VGT1 nucleotide sequence that encodes a biologically active non-transcribed RNA of the invention will encode at least 12, 25, 30, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, or 680 contiguous nucleotide bases or up to the total number of nucleotides present in a full-length VGT1 regulatory element of the invention.

Fragments of a RAP2.7 or VGT1 nucleotide sequence that are useful as hybridization probes or PCR primers generally need not encode a biologically active portion of a protein or RNA. Thus, a fragment of a RAP2.7 or VGT1 nucleotide sequence may encode a biologically active portion of a RAP2.7 protein or a biologically active non-coding RNA, or it may be a fragment that can be used as a hybridization probe or PCR primer using methods disclosed below. A biologically active portion of a RAP2.7 protein or VGT1 regulatory element can be prepared by isolating a portion of the RAP2.7 or VGT1 nucleotide sequences of the invention, expressing the encoded portion of the Rap2.7 or the active portion of the VGT1 regulatory element (e.g., by recombinant expression in vitro), and assessing the activity of the encoded portion of the RAP2.7 protein or the regulating ability of the regulatory element on RAP2.7. Nucleic acid molecules that are fragments of a RAP2.7 or VGT1 nucleotide sequence comprise at least 16, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750,

800, 850, 900, 950, 1000, 1050, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, or 2400 nucleotides, or up to the number of nucleotides present in a full RAP2.7 or VGT1 nucleotide sequence disclosed herein.

By "variants" is intended substantially similar sequences. For nucleotide sequences, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the amino acid sequence of one of the polypeptides of the invention. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant nucleotide sequences also include synthetically-derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis but which still encode a RAP2.7 protein or VGT1 regulatory element of the invention. Generally, variants of a particular nucleotide sequence of the invention will have at least 40%, 50%, 60%, 70%, generally at least 75%, 80%, 85%, or about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to that particular nucleotide sequence as determined by sequence alignment programs described elsewhere herein using default parameters.

By "variant" protein is intended a protein derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Variant proteins encompassed by the present invention are biologically active, that is they continue to possess the desired biological activity of the native protein, hence they will continue to possess at least one activity possessed by the native RAP2.7 protein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Biologically active variants of a RAP2.7 native protein of the invention will have at least 40%, 50%, 60%, 70%, generally at least 75%, 80%, 85%, 86%, 87%, 88%, 89%, about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the amino acid sequence for the native protein as determined by sequence alignment programs described elsewhere herein using default parameters. A biologically active variant of a protein of the invention may differ from that protein by as few as 1-15 amino acid residues, as few as 1-10, , as few as 5, as few as 4, 3, 2, or even 1 amino acid residue. As used herein, reference to a particular nucleotide or amino acid sequence (a RAP2.7 or VGT1 sequence) shall include all modified variants as described supra.

The proteins of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of RAP2.7 proteins can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide
5 sequence alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Nad. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; U.S. Patent No. 4,873,192; Walker and Gastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect
10 biological activity of the protein of interest may be found in the model of Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.).

Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be made.

Thus, the nucleotide sequences of the invention include both naturally occurring
15 sequences as well as mutant forms. Likewise, the proteins of the invention encompass both naturally-occurring proteins as well as variations and modified forms thereof. Such variants whether protein or nucleotide will continue to possess the desired RAP2.7 or VGT1 -like activity. It is recognized that variants need not retain all of the activities and/or properties of the native RAP2.7 or VGT1. Obviously, the mutations that will be
20 made in the DNA encoding the RPA2.7 variant must not place the sequence out of reading frame and in some embodiments will not create complementary regions that could produce secondary mRNA structure. See, EP Patent Application Publication No. 75,444.

The deletions, insertions, and substitutions of the protein or nucleotide sequences encompassed herein are not expected to produce radical changes in the characteristics of the
25 RAP 2.7 protein or VGT1 regulatory element. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays. That is, the activity of RAP2.7 polypeptides or VGT1 can be evaluated by either a change in flowering time or maturity when the encoded protein or regulatory element is altered.

30 Variant nucleotide sequences and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different RAP2.7 or VGT1 coding sequences can be manipulated to create a new RAP2.7 or VGT1 possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence

polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. For example, using this approach, sequence motifs encoding a domain of interest may be shuffled between the RAP2.7 encoding polynucleotide of the invention and other known flowering genes to obtain a new gene coding for a protein with an improved property of interest, such as an increased K_m in the case of an enzyme. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389391; Cramer *et al.* (1997) *Nature Biotech.* 15:436-438; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347; Zhang *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Cramer *et al.* (1998) *Nature* 391:288-291; and U.S. Patent Nos. 5,605,793 and 5,837,458.

The compositions of the invention also include isolated nucleic acid molecules comprising the promoter nucleotide sequences natively associated with the RAP2.7 polynucleotides. By "promoter" is intended a regulatory region of DNA usually comprising a TATA box capable of directing RNA polymerase II to initiate RNA synthesis at the appropriate transcription initiation site for a particular coding sequence. A promoter may additionally comprise other recognition sequences generally positioned upstream or 5' to the TATA box, referred to as upstream promoter elements, which influence the transcription initiation rate.

The nucleotide sequences of the invention can be used to isolate corresponding sequences from other organisms, particularly other plants, more particularly other crop plants. In this manner, methods such as PCR, hybridization, and the like can be used to identify such sequences based on their sequence homology to the sequences set forth herein. Sequences isolated based on their sequence identity to the nucleotide sequences set forth herein or to fragments thereof are encompassed by the present invention. Such sequences include sequences that are orthologs of the disclosed sequences. By "orthologs" is intended genes derived from a common ancestral gene and which are found in different species as a result of speciation. Genes found in different species are considered orthologs when their nucleotide sequences and/or their encoded protein sequences share substantial identity as defined elsewhere herein. Functions of orthologs are often highly conserved among species. Thus, isolated sequences that have RAP2.7 or VGT1 activity or encode a RAP2.7 protein and which hybridize under stringent conditions to RAP2.7 or VGT1 sequences disclosed herein, or to fragments thereof, are encompassed by the present invention.

In a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any plant of interest. Methods for designing PCR primers and PCR cloning are generally known in the art and are disclosed in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press Plainview, New York). See also Innis *et al.*, eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis and Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); and Innis and Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, New York). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially-mismatched primers, and the like.

In hybridization techniques, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen organism. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labeled with a detectable group such as ^{32}P , or any other detectable marker. Thus, for example, probes for hybridization can be made by labeling synthetic oligonucleotides based on the disease-resistant sequences of the invention. Methods for preparation of probes for hybridization and for construction of cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

For example, an entire sequence disclosed herein, or one or more portions thereof, may be used as a probe capable of specifically hybridizing to corresponding flowering or maturity regulating sequences, including promoters and messenger RNAs. To achieve specific hybridization under a variety of conditions, such probes include sequences that are unique among flowering or maturity related sequences and may be at least about 10 or 15 or 17 nucleotides in length or at least about 20 or 22 or 25 nucleotides in length. Such probes may be used to amplify corresponding sequences from a chosen organism by PCR. This technique may be used to isolate additional coding sequences from a desired organism or as a diagnostic assay to determine the presence of coding sequences in an organism. Hybridization techniques include hybridization screening of plated DNA libraries (either

plaques or colonies; see, for example, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

Hybridization of such sequences may be carried out under stringent conditions. By "stringent conditions" or "stringent hybridization conditions" is intended conditions under which a probe will hybridize to its target sequence to a detectably greater degree than to other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different under different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences that are 100% complementary to the probe can be identified (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length or less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3. Incubation should be at a temperature of least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulfate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl, 0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C for 20 minutes. Optionally, wash buffers may comprise about 0.1% to about 1% SDS. Duration of hybridization is generally less than about 24 hours, usually about 4 to about 12 hours.

Specificity is typically a function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl (1984) *Anal. Biochem.* 138:267-284: $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target

sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1 °C for each 1% of mismatching; thus, T_m , hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T_m , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T_m of less than 45°C (aqueous solution) or 32°C (formamide solution), the SSC concentration may be increased so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 (Elsevier, New York); and Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Chapter 2 (Greene Publishing and Wiley-Interscience, New York). See Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

In general, sequences that encode a RAP2.7 protein or which encode a VGT1 regulatory element and which hybridize to the RAP2.7 or VGT1 sequences disclosed herein will be at least about 70% homologous, and even about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 98%, 99% or more homologous with the disclosed sequences. That is, the sequence identity of the sequences may be from about 70% or 75%, and even about 80%, 85%, 87%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical, or higher, so that the sequences may differ by only 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 amino acid residue or by 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 nucleic acid.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence," (b) "comparison window," (c) "sequence identity," (d) "percentage of sequence identity," and (e) "substantial identity."

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified

sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent sequence identity between any two sequences can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988) *CABIOS* 4:11-17; the local homology algorithm of Smith *et al.* (1981) *Adv. Appl. Math.* 2:482; the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453; the search-for-similarity-method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444-2448; the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins *et al.* (1988) *Gene* 73:237-244 (1988); Higgins *et al.* (1989) *CABIOS* 5:151-153; Corpet *et al.* (1988) *Nucleic Acids Res.* 16:10881-90; Huang *et al.* (1992) *CABIOS* 8:155-65; and Pearson *et al.* (1994) *Meth. Mol. Biol.* 24:307-331. The ALIGN program is based on the algorithm of Myers and Miller (1988) *supra*. A PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be

WO 2006/015045

PCT/US2005/026641

used with the ALIGN program when comparing amino acid sequences. The BLAST programs of Altschul *et al* (1990) *J. Mol. Biol.* 215:403 are based on the algorithm of Karlin and Altschul (1990) *supra*. BLAST nucleotide searches can be performed with the BLASTN program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the invention. BLAST protein searches can be performed with the BLASTX program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to a protein or polypeptide of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.* (1997) *supra*. When utilizing BLAST, Gapped BLAST, or PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins) can be used (See National Center for Biotechnology Information (NCBI) website). Alignment may also be performed manually by inspection.

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP version 10 using the following parameters: % identity using GAP Weight of 50 and Length Weight of 3; % similarity using Gap Weight of 12 and Length Weight of 4, or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10. GAP uses the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453, to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension

penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 200. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 or greater.

5 GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually
10 match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

15 (c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid
20 substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have
25 "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score
30 between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

 (d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may

comprise additions or deletions (i.e., gaps) 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
5 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.

(e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 80%, 85%, 90%, 95%, or higher
10 sequence identity compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for
15 these purposes normally means sequence identity of at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or at least 95% or higher sequence identity.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_{rr}) for the specific
20 sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C lower than the T_m , depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is
25 created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross-reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a peptide indicates that a
30 peptide comprises a sequence with at least 70%, 75%, 80%, 83%, 85%, 88%, 90%, 93%, 95%, 96%, 97%, 98%, or 99% or higher sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (1970) *J Mol. Biol.* 48:443453. An indication that two peptide sequences are substantially identical is that one peptide is

immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. Peptides that are "substantially similar" share sequences as noted above except that residue positions that are not identical may differ by conservative amino acid changes.

Methods for altering flowering time in a plant are provided. In some embodiments, the methods involve stably transforming a plant with a DNA construct comprising a nucleotide sequence of the invention operably linked to a promoter that drives expression (or transcription) in a plant. While the choice of promoter will depend on the desired timing and location of expression of the nucleotide sequences, desirable promoters include constitutive and tissue specific promoters. These methods may find use in agriculture, particularly in changing the maturity of a particular crop plant to alter its area of adaptation. Thus, transformed plants, plant cells, plant tissues and seeds thereof are provided by the present invention.

In another embodiment, the methods of the present invention involve identifying phenotypes associated with an altered flowering time by loss of RAP2.7 or VGT1 activity in plants that contain transposon insertions within the nucleotide sequences herein.

In some embodiments, the nucleic acid molecules comprising RAP2.7 or VGT1 sequences of the invention are provided in expression cassettes or nucleotide constructs for expression/transcription in the plant of interest. Such cassettes will include 5' and 3' regulatory sequences operably linked to a RAP2.7 or VGT1 nucleotide sequence of the invention. By "operably linked" is intended a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. The cassette may additionally contain at least one additional nucleotide sequence to be cotransformed into the organism. Alternatively, the additional sequence(s) can be provided on multiple expression cassettes or nucleotide construct.

Such an expression cassette or nucleotide construct is provided with a plurality of restriction sites for insertion of the RAP2.7 or VGT1 sequence to be under the transcriptional regulation of the regulatory regions. The expression cassette or nucleotide construct may additionally contain selectable marker genes. The expression cassette will include in the 5'-3' direction of transcription, a transcriptional and if necessary a translational initiation region, a RAP2.7 or VGT1 nucleotide sequence of the invention, and a transcriptional and if necessary, translational termination region

WO 2006/015045

PCT/US2005/026641

functional in plants. The transcriptional initiation region, or promoter, may be native or analogous or foreign or heterologous to the plant host. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. By "foreign" is intended that the transcriptional initiation region is not found in the native plant into which the
5 transcriptional initiation region is introduced. As used herein, a "chimeric gene" comprises a coding sequence operably linked to a transcription initiation region that is heterologous to the coding sequence.

While it may be preferable to regulate RAP2.7 or VGT1 sequences using heterologous promoters, the native promoter sequences may be used. Such constructs
10 would change expression levels of the RAP2.7 or amounts of the VGT1 regulatory element present in the plant or plant cell. Thus, the phenotype of the plant or plant cell is altered.

The termination region may be native with the transcriptional initiation region, may be native with the operably linked DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid
15 of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acid Res.*
20 15:9627-9639. Where appropriate, the gene(s) may be optimized for increased expression in the transformed plant. That is, the genes can be synthesized using plant-preferred codons for improved expression. Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Patent Nos. 5,380,831, and 5,436,391, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498.

25

Additional sequence modifications are known to enhance gene expression in a cellular host. These include elimination of sequences encoding spurious polyadenylation signals, exon-intron splice site signals, transposon-like repeats, and other such well characterized sequences that may be deleterious to gene expression.
30 The G-C content of the sequence may be adjusted to enhance expression in a given host cell. When possible, the sequence is modified to avoid predicted hairpin secondary mRNA structures.

The expression cassettes/nucleotide constructs may additionally contain 5' leader sequences in the expression cassette construct. Such leader sequences can act to

enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein *et al.* (1989) *Proc. Nat'l. Acad. Sci. USA* 86:6126-6130); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Allison *et al.* (1986); MDMV leader (Maize Dwarf Mosaic Virus); *Virology* 154:9-20), and human immunoglobulin heavy-chain binding protein (BiP), (Macejak *et al.* (1991) *Nature* 353:90-94); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4) (Jobling *et al.* (1987) *Nature* 325:622-625); tobacco mosaic virus leader (TMV) (Gallie *et al.* (1989) in *Molecular Biology of RNA*, ed. Cech (Liss, New York), pp. 237-256); and maize chlorotic mottle virus leader (MCMV) (Lommel *et al.* (1991) *Virology* 81:382-385). See also, Della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968. Other methods known to enhance translation can also be utilized, for example, introns, and the like.

In those instances where it is desirable to have the expressed product of the heterologous nucleotide sequence of interest directed to a particular organelle, such as the chloroplast or mitochondrion, or secreted at the cell's surface or extracellularly, the expression cassette may further comprise a coding sequence for a transit peptide. Such transit peptides are well known in the art and include, but are not limited to, the transit peptide for the acyl carrier protein, the small subunit of RUBISCO, plant EPSP synthase, and the like.

In preparing the expression cassette/nucleotide construct, the various DNA fragments may be manipulated so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g., transitions and transversions, may be involved.

Generally, the expression cassette/nucleotide construct will comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). See generally, Yarranton (1992) *Curr. Opin. Biotech.*

3:506-511; Christopherson *et al.* (1992) *Proc. Nad. Acad. Sci. USA* 89:6314-6318; Yao *et al.* (1992) *Cell* 71:63-72; Reznikoff (1992) *Mol. Microbiol.* 6:2419-2422; Barkley *et al.* (1980) in *The Operon*, pp. 177-220; Hu *et al.* (1987) *Cell* 48:555-566; Brown *et al.* (1987) *Cell* 49:603-612; Figge *et al.* (1988) *Cell* 52:713-722; Deuschle *et al.* (1989) *Proc. Nad. Acad. Sci. USA* 86:5400-5404; Fuerst *et al.* (1989) *Proc. Nad. Acad. Sci. USA* 86:2549-2553; Deuschle *et al.* (1990) *Science* 248:480-483; Gossen (1993) Ph.D. Thesis, University of Heidelberg; Reines *et al.* (1993) *Proc. Nad. Acad. Sci. USA* 90:1917-1921; Labow *et al.* (1990) *Mol. Cell. Biol.* 10:3343-3356; Zambretti *et al.* (1992) *Proc. Nad. Acad. Sci. USA* 89:3952-3956; Baim *et al.* (1991) *Proc. Nad. Acad. Sci. USA* 88:5072-5076; Wyborski *et al.* (1991) *Nucleic Acids Res.* 19:4647-4653; Hillenand-Wissman (1989) *Topics Mol. Struc. Biol.* 10:143-162; Degenkolb *et al.* (1991) *Antimicrob. Agents Chemother.* 35:1591-1595; Kleinschmidt *et al.* (1988) *Biochemistry* 27:1094-1104; Bonin (1993) Ph.D. Thesis, University of Heidelberg; Gossen *et al.* (1992) *Proc. Nad. Acad. Sci. USA* 89:5547-5551; Oliva *et al.* (1992) *Antimicrob. Agents Chemother.* 36:913-919; Hlavka *et al.* (1985) *Handbook of Experimental Pharmacology*, Vol. 78 (Springer-Verlag, Berlin); Gill *et al.* (1988) *Nature* 334:721-724.

The above list of selectable marker genes is not meant to be limiting. Any selectable marker gene can be used in the present invention. A number of promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. That is, the nucleic acids can be combined with constitutive, tissue-preferred, or other promoters for expression in plants. Constitutive promoters include, for example, the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO 99/43838 and U.S. Patent No. 6,072,050; the core CaMV 35S promoter (Odell *et al.* (1985) *Nature* 313:810-812); rice actin (McElroy *et al.* (1990) *Plant Cell* 2:163-171); ubiquitin (Christensen *et al.* (1989) *Plant Mol. Biol.* 12:619-632 and Christensen *et al.* (1992) *Plant Mol. Biol.* 18:675-689); pEMU (Last *et al.* (1991) *Theor. Appl. Genet.* 81:581-588); MAS (Velten *et al.* (1984) *EMBO J.* 3:2723-2730); ALS promoter (U.S. Patent No. 5,659,026), and the like. Other constitutive promoters include, for example, U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; and 6,177,611.

Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. Depending upon the objective, the promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression, or a chemical-repressible

WO 2006/015045

PCT/US2005/026641

promoter, where application of the chemical represses gene expression. Chemical-inducible promoters are known in the art and include, but are not limited to: the maize In2-2 promoter, which is activated by benzenesulfonamide herbicide safeners; the maize GST promoter, which is activated by hydrophobic electrophilic compounds that are used as pre-emergent herbicides; and the tobacco PR-1 a promoter, which is activated by salicylic acid. Other chemical-regulated promoters of interest include steroid-responsive promoters. See, for example, the glucocorticoid-inducible promoter in Schena *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425 and McNellis *et al.* (1998) *Plant J.* 14(2):247-257) and tetracycline-inducible and tetracycline-repressible promoters (for example, Gatz *et al.* (1991) *Mol. Gen. Genet.* 227:229-237, and U.S. Patent Nos. 5,814,618 and 5,789,156).

Tissue-preferred promoters can be utilized to target enhanced gene expression within a particular plant tissue. Tissue-preferred promoters include Yamamoto *et al.* (1997) *Plant J* 12(2):255-265; Kawamata *et al.* (1997) *Plant Cell Physiol.* 38(7):792-803; Hansen *et al.* (1997) *Mol. Gen Genet.* 254(3):337-343; Russell *et al.* (1997) *Transgenic Res.* 6(2):157-168; Rinehart *et al.* (1996) *Plant Physiol.* 112(3):1331-1341; Van Camp *et al.* (1996) *Plant Physiol.* 112(2):525-535; Canevascini *et al.* (1996) *Plant Physiol.* 112(2):513-524; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5):773-778; Lam (1994) *Results Probl. Cell Differ.* 20:181-196; Orozco *et al.* (1993) *Plant Mol Biol.* 23(6):1129-1138; Matsuoka *et al.* (1993) *Proc Natl. Acad. Sci. USA* 90(20):9586-9590; and Guevara-Garcia *et al.* (1993) *Plant J.* 4(3):495-505. Such promoters can be modified, if necessary, for weak expression. Leaf-specific promoters are known in the art. See, for example, Yamamoto *et al.* (1997) *Plant J.* 12(2):255-265; Kwon *et al.* (1994) *Plant Physiol.* 105:357-67; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5):773-778; Gotor *et al.* (1993) *Plant J* 3:509-18; Orozco *et al.* (1993) *Plant Mol. Biol.* 23(6):1129-1138; and Matsuoka *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90(20):9586-9590.

Where low level expression is desired, weak promoters will be used. Generally, by "weak promoter" is intended a promoter that drives expression of a coding sequence at a low level. By low level is intended at levels of about 1/1000 transcripts to about 1/100,000 transcripts to about 1/500,000 transcripts per cell. Alternatively, it is recognized that weak promoters also include promoters that are expressed in only a few cells and not in others to give a total low level of expression. Where a promoter is expressed at unacceptably high levels, portions of the promoter sequence can be deleted or modified to decrease expression

WO 2006/015045

PCT/US2005/026641

levels. Such weak constitutive promoters include, for example, the core promoter of the *Rsyn7* promoter (WO 99/43838 and U.S. Patent No. 6,072,050), the core 35S CaMV promoter, and the like. Other constitutive promoters include, for example, U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142. See also,
5 U.S. Patent No. 6,177,611.

As used herein, "vector" refers to a molecule such as a plasmid, cosmid or bacterial phage for introducing a nucleotide construct and/or expression cassette into a host cell. Cloning vectors typically contain one or a small number of restriction endonuclease recognition sites at which foreign nucleotide sequences can be inserted in a determinable
10 fashion without loss of essential biological function of the vector, as well as a marker gene that is suitable for use in the identification and selection of cells transformed with the cloning vector. Marker genes typically include genes that provide tetracycline resistance, hygromycin resistance or ampicillin resistance.

The methods of the invention involve introducing a nucleotide construct into a plant. By "introducing" is intended presenting to the plant the nucleotide construct in such a manner
15 that the construct gains access to the interior of a cell of the plant. The methods of the invention do not depend on a particular method for introducing a nucleotide construct to a plant, only that the nucleotide construct gains access to the interior of at least one cell of the plant. Methods for introducing nucleotide constructs into plants are known in the art
20 including, but not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

By "stable transformation" is intended that the nucleotide construct introduced into a plant integrates into the genome of the plant and is capable of being inherited by progeny thereof. By "transient transformation" is intended that a nucleotide construct introduced
25 into a plant does not integrate into the genome of the plant.

The nucleotide constructs of the invention may be introduced into plants by contacting plants with a virus or viral nucleic acids. Generally, such methods involve incorporating a nucleotide construct of the invention within a viral DNA or RNA molecule. It is recognized that the RAP2.7 protein of the invention may be initially synthesized as part
30 of a viral polyprotein, which later may be processed by proteolysis in vivo or in vitro to produce the desired recombinant protein. Further, it is recognized that promoters of the invention also encompass promoters utilized for transcription by viral RNA polymerases. Methods for introducing nucleotide constructs into plants and expressing a protein encoded therein, involving viral DNA or RNA molecules, are known in the art. See, for example,

U.S. Patent Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367 and 5,316,931.

A variety of other transformation protocols are contemplated in the present invention. Transformation protocols as well as protocols for introducing nucleotide sequences into plants may vary depending on the type of plant or plant cell, *i. e.*, monocot or dicot, targeted for transformation. Suitable methods of introducing nucleotide sequences into plant cells and subsequent insertion into the plant genome include microinjection (Crossway et al. (1986) *Biotechniques* 4:320-334), electroporation (Riggs et al. (1986) *Proc. Nad. Acad. Sci. USA* 83:5602-5606, Agrobacterium-mediated transformation (Townsend et al., U.S. Patent No. 5,563,055; Zhao et al., U.S. Patent No. 5,981,840), direct gene transfer (Paszkowski et al. (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, Sanford et al., U.S. Patent No. 4,945,050; Tomes et al., U.S. Patent No. 5,879,918; Tomes et al., U.S. Patent No. 5,886,244; Bidney et al., U.S. Patent No. 5,932,782; Tomes et al. (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, eds. Gamborg and Phillips (Springer-Verlag, Berlin); McCabe et al. (1988) *Biotechnology* 6:923-926); and *Lecl* transformation (WO 00/28058, published May 18, 2000). Also see Weissinger et al. (1988) *Ann. Rev. Genet.* 22:421-477; Sanford et al. (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou et al. (1988) *Plant Physiol.* 87:671-674 (soybean); McCabe et al. (1988) *Bio/Technology* 6:923-926 (soybean); Finer and McMullen (1991) *In Vitro Cell Dev. Biol.* 27P:175-182 (soybean); Singh et al. (1998) *Theor. Appl. Genet.* 96:319-324 (soybean); Datta et al. (1990) *Biotechnology* 8:736-740 (rice); Klein et al. (1988) *Proc. Nad. Acad. Sci. USA* 85:43054309 (maize); Klein et al. (1988) *Biotechnology* 6:559-563 (maize); Tomes, U.S. Patent No. 5,240,855; Buising et al., U.S. Patent Nos. 5,322,783 and 5,324,646; Tomes et al. (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg (Springer-Verlag, Berlin) (maize); Klein et al. (1988) *Plant Physiol.* 91:440-444 (maize); Fromm et al. (1990) *Biotechnology* 8:833-839 (maize); Hooykaas-Van Slogteren et al. (1984) *Nature (London)* 311:763-764; Bowen et al., U.S. Patent No. 5,736,369 (cereals); Bytebier et al. (1987) *Proc. Nad. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet et al. (1985) in *The Experimental Manipulation of Ovule Tissues*, ed. Chapman et al. (Longman, New York), pp. 197-209 (pollen); Kaeppler et al. (1990) *Plant Cell Reports* 9:415-418 and Kaeppler et al. (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation);

WO 2006/015045

PCT/US2005/026641

D'Halluin *et al.* (1992) *Plant Cell* 4:1495-1505 (electroporation); Li *et al.* (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford (1995) *Annals of Botany* 75:407-413 (rice); Osjoda *et al.* (1996) *Nature Biotechnology* 14:745-750 (maize via *Agrobacterium tumefaciens*).

5 The cells that have been transformed may be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having constitutive expression of the desired phenotypic characteristic identified. Two or more
10 generations may be grown to ensure that constitutive expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure constitutive expression of the desired phenotypic characteristic has been achieved.

 The present invention may be used for transformation of any plant species,
15 including, but not limited to, monocots and dicots. Examples of plant species of interest include, but are not limited to, corn (*Zea mays*), *Brassica spp.* (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea spp.*),
20 coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus spp.*), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa spp.*), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum spp.*), oats, barley, vegetables, ornamentals, and conifers.
30

 Vegetables include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus spp.*), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea

(*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum.

5 Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir
10 (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow cedar (*Chamaecyparis nootkatensis*). Plants of the present invention may be crop plants (for example, alfalfa, sunflower, *Brassica*, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, *etc.*), corn or soybean plants.

Plants of particular interest include grain plants that provide seeds of interest, oil-
15 seed plants, and leguminous plants. Seeds of interest include grain seeds, such as corn, wheat, barley, rice, sorghum, rye, *etc.* Oil-seed plants include cotton, soybean, safflower, sunflower, *Brassica*, maize, alfalfa, palm, coconut, *etc.* Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, *etc.*

20 It is recognized that with these nucleotide sequences, gene silencing such as antisense constructions complementary to at least a portion of the messenger RNA (mRNA) for RAP2.7 or VGT1 sequences can be constructed. Antisense nucleotides are constructed to hybridize with the corresponding mRNA. Modifications of the antisense sequences may be made as long as the sequences hybridize to and interfere with expression of the corresponding mRNA. In
25 this manner, antisense constructions having 70%, 80%, 85%, 90%, 95% or more sequence identity to the corresponding antisense sequences may be used. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, or greater may be used.

Gene silencing refers to posttranscriptional interference with gene expression.

30 Techniques such as antisense, co-suppression, and RNA interference (RNAi), for example, have been shown to be effective in gene silencing. (For reviews, see Arndt and Rank, Genome 40(6):785-797, 1997; Turner and Schuch, Journal of Chemical Technology and Biotechnology 75(10):869-882, 2000; Klink and Wolniak, Journal of Plant Growth Regulation 19(4):371-384, 2000)

Antisense technology can be used to control gene expression through antisense DNA or RNA or through double- or triple-helix formation. Antisense techniques are discussed, for example, in - Okano, J. *Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988).

5 Triple helix formation is discussed in, for instance Lee *et al.*, *Nucleic Acids Research* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991). The methods are based on binding of a polynucleotide to a complementary DNA or RNA. For example, the 5' coding portion of a polynucleotide that encodes the mature polypeptide of the present invention may be used to design an antisense RNA oligonucleotide of from about
10 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription, thereby preventing transcription and the production of cytokinin biosynthetic enzymes. The antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA molecule into cytokinin biosynthetic enzymes. The oligonucleotides described above can also be delivered to cells
15 such that the antisense RNA or DNA may be expressed in vivo to inhibit production of cytokinin biosynthetic enzymes. The DNAs of this invention may also be employed to co-suppress or silence the cytokinin metabolic enzyme genes; for example, as described in PCT Patent Application Publication WO 98/36083.

The RAP 2.7 nucleotide sequence operably linked to the regulatory elements herein
20 can be an antisense sequence for a targeted gene. By "antisense DNA nucleotide sequence" is intended a sequence that is in inverse orientation to the 5'-to-3' normal orientation of that nucleotide sequence. When delivered into a plant cell, expression of the antisense DNA sequence prevents normal expression of the DNA nucleotide sequence for the targeted gene. The antisense nucleotide sequence encodes an RNA transcript that is
25 complementary to and capable of hybridizing with the endogenous messenger RNA (mRNA) produced by transcription of the DNA nucleotide sequence for the targeted gene. In this case, production of the native protein encoded by the targeted gene is inhibited to achieve a desired phenotypic response. Thus the regulatory sequences disclosed herein can be operably linked to antisense DNA sequences to reduce or inhibit expression of a
30 native protein in the plant seed.

It is also recognized that the level and/or activity of the polypeptide may be modulated by employing a polynucleotide that is not capable of directing, in a transformed plant, the expression of a protein or an RNA. For example, the polynucleotides of the invention may be used to design polynucleotide constructs that can be employed in

methods for altering or mutating a genomic nucleotide sequence in an organism. Such polynucleotide constructs include, but are not limited to, RNA:DNA vectors, RNA:DNA mutational vectors, RNA:DNA repair vectors, mixed-duplex oligonucleotides, self-complementary RNA:DNA oligonucleotides, and recombinogenic oligonucleobases. Such nucleotide constructs and methods of use are known in the art. See, U.S. Patent Nos. 5,565,350; 5,731,181; 5,756,325; 5,760,012; 5,795,972; and 5,871,984.

See also, WO 98/49350, WO 99/07865, WO 99/25821, and Beetham *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96:8774-8778.

It is therefore recognized that methods of the present invention do not depend on the incorporation of the entire polynucleotide into the genome, only that the plant or cell thereof is altered as a result of the introduction of the polynucleotide into a cell. In one embodiment of the invention, the genome may be altered following the introduction of the polynucleotide into a cell. For example, the polynucleotide, or any part thereof, may incorporate into the genome of the plant. Alterations to the genome of the present invention include, but are not limited to, additions, deletions, and substitutions of nucleotides into the genome. While the methods of the present invention do not depend on additions, deletions, and substitutions of any particular number of nucleotides, it is recognized that such additions, deletions, or substitutions comprises at least one nucleotide.

Methods are provided to reduce or eliminate the activity of a RAP2.7 polypeptide of the invention by transforming a plant cell with an expression cassette that expresses a polynucleotide that inhibits the expression of the RAP2.7 polypeptide. The polynucleotide may inhibit the expression of the Rap2.7 polypeptide directly, by preventing translation of the RAP2.7 messenger RNA, or indirectly, by encoding a polypeptide that inhibits the transcription or translation of a RAP2.7 gene encoding a RAP2.7 polypeptide. Methods for inhibiting or eliminating the expression of a gene in a plant are well known in the art, and any such method may be used in the present invention to inhibit the expression of a RAP2.7 polypeptide.

In some embodiments of the present invention, a plant is transformed with an expression cassette that is capable of expressing a polynucleotide that inhibits the expression of a RAP2.7 polypeptide of the invention. The term "expression" as used herein refers to the biosynthesis of a gene product, including the transcription and/or translation of said gene product. For example, for the purposes of the present invention, an expression cassette capable of expressing a polynucleotide that inhibits the expression of at

least one RAP2.7 polypeptide is an expression cassette capable of producing an RNA molecule that inhibits the transcription and/or translation of at least one RAP2.7 polypeptide of the invention. The "expression" or "production" of a protein or polypeptide from a DNA molecule refers to the transcription and translation of the coding sequence to produce the protein or polypeptide, while the "expression" or "production" of a protein or polypeptide from an RNA molecule refers to the translation of the RNA coding sequence to produce the protein or polypeptide.

Examples of polynucleotides that inhibit the expression of a RAP2.7 polypeptide are given below.

i. Sense Suppression/Cosuppression

In some embodiments of the invention, inhibition of the expression of a RAP2.7 polypeptide may be obtained by sense suppression or cosuppression. For cosuppression, an expression cassette is designed to express an RNA molecule corresponding to all or part of a messenger RNA encoding a RAP2.7 polypeptide in the "sense" orientation. Over expression of the RNA molecule can result in reduced expression of the native gene. Accordingly, multiple plant lines transformed with the cosuppression expression cassette are screened to identify those that show the greatest inhibition of RAP2.7 polypeptide expression.

The polynucleotide used for cosuppression may correspond to all or part of the sequence encoding the RAP2.7 polypeptide, all or part of the 5' and/or 3' untranslated region of a RAP2.7 polypeptide transcript, or all or part of both the coding sequence and the untranslated regions of a transcript encoding a RAP2.7 polypeptide. In some embodiments where the polynucleotide comprises all or part of the coding region for the RAP2.7 polypeptide, the expression cassette is designed to eliminate the start codon of the polynucleotide so that no protein product will be translated.

Cosuppression may be used to inhibit the expression of plant genes to produce plants having undetectable protein levels for the proteins encoded by these genes. See, for example, Broin *et al.* (2002) *Plant Cell* 14:1417-1432. Cosuppression may also be used to inhibit the expression of multiple proteins in the same plant. See, for example, U.S. Patent No. 5,942,657. Methods for using cosuppression to inhibit the expression of endogenous genes in plants are described in Flavell *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3490-3496; Jorgensen *et al.* (1996) *Plant Mol. Biol.* 31:957-973; Johansen and Carrington (2001) *Plant Physiol.* 126:930-938; Broin *et al.* (2002) *Plant Cell* 14:1417-1432;

Stoutjesdijk *et al.* (2002) *Plant Physiol.* 129:1723-1731; Yu *et al.* (2003) *Phytochemistry* 63:753-763; and U.S. Patent Nos. 5,034,323, 5,283,184, and 5,942,657.

The efficiency of cosuppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the sense sequence and 5' of the polyadenylation signal. See, U.S. Patent Publication No. 20020048814.

Typically, such a nucleotide sequence has substantial sequence identity to the sequence of the transcript of the endogenous gene, optimally greater than about 65% sequence identity, more optimally greater than about 85% sequence identity, most optimally greater than about 95% sequence identity. See, U.S. Patent Nos. 5,283,184 and 5,034,323.

ii. Antisense Suppression

In some embodiments of the invention, inhibition of the expression of the RAP2.7 polypeptide may be obtained by antisense suppression. For antisense suppression, the expression cassette is designed to express an RNA molecule complementary to all or part of a messenger RNA encoding the RAP2.7 polypeptide. Over expression of the antisense RNA molecule can result in reduced expression of the native gene. Accordingly, multiple plant lines transformed with the antisense suppression expression cassette are screened to identify those that show the greatest inhibition of RAP2.7 polypeptide expression.

The polynucleotide for use in antisense suppression may correspond to all or part of the complement of the sequence encoding the RAP2.7 polypeptide, all or part of the complement of the 5' and/or 3' untranslated region of the RAP2.7 transcript, or all or part of the complement of both the coding sequence and the untranslated regions of a transcript encoding the RAP2.7 polypeptide. In addition, the antisense polynucleotide may be fully complementary (i.e., 100% identical to the complement of the target sequence) or partially complementary (i.e., less than 100% identical to the complement of the target sequence) to the target sequence. Antisense suppression may be used to inhibit the expression of multiple proteins in the same plant. See, for example, U.S. Patent No. 5,942,657. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, 300, 400, 450, 500, 550, or greater may be used. Methods for using antisense suppression to inhibit the expression of endogenous genes in plants are described, for example, in Liu *et al.* (2002) *Plant Physiol.* 129:1732-1743 and U.S. Patent Nos. 5,759,829 and 5,942,657.

Efficiency of

antisense suppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the antisense sequence and 5' of the polyadenylation signal. See, U.S. Patent Publication No. 20020048814.

5 iii. *Double-Stranded RNA Interference*

In some embodiments of the invention, inhibition of the expression of a RAP2.7 polypeptide may be obtained by double-stranded RNA (dsRNA) interference. For dsRNA interference, a sense RNA molecule like that described above for cosuppression and an antisense RNA molecule that is fully or partially complementary to the sense RNA
10 molecule are expressed in the same cell, resulting in inhibition of the expression of the corresponding endogenous messenger RNA.

Expression of the sense and antisense molecules can be accomplished by designing the expression cassette to comprise both a sense sequence and an antisense sequence. Alternatively, separate expression cassettes may be used for the sense and antisense
15 sequences. Multiple plant lines transformed with the dsRNA interference expression cassette or expression cassettes are then screened to identify plant lines that show the greatest inhibition of RAP2.7 polypeptide expression. Methods for using dsRNA interference to inhibit the expression of endogenous plant genes are described in Waterhouse *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964, Liu *et al.* (2002)
20 *Plant Physiol.* 129:1732-1743, and WO 99/49029, WO 99/53050, WO 99/61631, and WO 00/49035.

iv. *Hairpin RNA Interference and Intron-Containing Hairpin RNA Interference*

In some embodiments of the invention, inhibition of the expression of one or a
25 RAP2.7 polypeptide may be obtained by hairpin RNA (hpRNA) interference or intron-containing hairpin RNA (ihpRNA) interference. These methods are highly efficient at inhibiting the expression of endogenous genes. See, Waterhouse and Helliwell (2003) *Nat. Rev. Genet.* 4:29-38 and the references cited therein.

For hpRNA interference, the expression cassette is designed to express an RNA
30 molecule that hybridizes with itself to form a hairpin structure that comprises a single-stranded loop region and a base-paired stem. The base-paired stem region comprises a sense sequence corresponding to all or part of the endogenous messenger RNA encoding the gene whose expression is to be inhibited, and an antisense sequence that is fully or partially complementary to the sense sequence. Thus, the base-paired stem region of the

WO 2006/015045

PCT/US2005/026641

molecule generally determines the specificity of the RNA interference. hpRNA molecules are highly efficient at inhibiting the expression of endogenous genes, and the RNA interference they induce is inherited by subsequent generations of plants. See, for example, Chuang and Meyerowitz (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; 5 Stoutjesdijk *et al.* (2002) *Plant Physiol.* 129:1723-1731; and Waterhouse and Helliwell (2003) *Nat. Rev. Genet.* 4:29-38. Methods for using hpRNA interference to inhibit or silence the expression of genes are described, for example, in Chuang and Meyerowitz (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk *et al.* (2002) *Plant Physiol.* 129:1723-1731; Waterhouse and Helliwell (2003) *Nat. Rev. Genet.* 4:29-38; Pandolfini *et al.* *BMC Biotechnology* 3:7, and U.S. Patent Publication No. 20030175965. 10

A transient assay for the efficiency of hpRNA constructs to silence gene expression *in vivo* has been described by Panstruga *et al.* (2003) *Mol. Biol. Rep.* 30:135-140.

For ihpRNA, the interfering molecules have the same general structure as for 15 hpRNA, but the RNA molecule additionally comprises an intron that is capable of being spliced in the cell in which the ihpRNA is expressed. The use of an intron minimizes the size of the loop in the hairpin RNA molecule following splicing, and this increases the efficiency of interference. See, for example, Smith *et al.* (2000) *Nature* 407:319-320. In fact, Smith *et al.* show 100% suppression of endogenous gene expression using ihpRNA-mediated interference. Methods for using ihpRNA interference to inhibit the expression of 20 endogenous plant genes are described, for example, in Smith *et al.* (2000) *Nature* 407:319-320; Wesley *et al.* (2001) *Plant J.* 27:581-590; Wang and Waterhouse (2001) *Curr. Opin. Plant Biol.* 5:146-150; Waterhouse and Helliwell (2003) *Nat. Rev. Genet.* 4:29-38; Helliwell and Waterhouse (2003) *Methods* 30:289-295, and U.S. Patent Publication No. 25 20030180945.

The expression cassette for hpRNA interference may also be designed such that the sense sequence and the antisense sequence do not correspond to an endogenous RNA. In this embodiment, the sense and antisense sequence flank a loop sequence that comprises a nucleotide sequence corresponding to all or part of the endogenous messenger RNA of the 30 target gene. Thus, it is the loop region that determines the specificity of the RNA interference. See for example, WO 02/00904.

v. *Amplicon-Mediated Interference*

Amplicon expression cassettes comprise a plant virus-derived sequence that contains all or part of the target gene but generally not all of the genes of the native virus. The viral sequences present in the transcription product of the expression cassette allow the transcription product to direct its own replication. The transcripts produced by the amplicon may be either sense or antisense relative to the target sequence (i.e., the messenger RNA for the RAP2.7 polypeptide). Methods of using amplicons to inhibit the expression of endogenous plant genes are described, for example, in Angell and Baulcombe (1997) *EMBO J.* 16:3675-3684, Angell and Baulcombe (1999) *Plant J.* 20:357-362, and U.S. Patent No. 6,646,805.

vi. *Ribozymes*

In some embodiments, the polynucleotide expressed by the expression cassette of the invention is catalytic RNA or has ribozyme activity specific for the messenger RNA of the RAP2.7 polypeptide. Thus, the polynucleotide causes the degradation of the endogenous messenger RNA, resulting in reduced expression of the RAP2.7 polypeptide. This method is described, for example, in U.S. Patent No. 4,987,071.

vii. *Small Interfering RNA or Micro RNA*

In some embodiments of the invention, inhibition of the expression of a RAP2.7 polypeptide may be obtained by RNA interference by expression of a gene encoding a micro RNA (miRNA). miRNAs are regulatory agents consisting of about 22 ribonucleotides. miRNA are highly efficient at inhibiting the expression of endogenous genes. See, for example Javier *et al.* (2003) *Nature* 425: 257-263.

For miRNA interference, the expression cassette is designed to express an RNA molecule that is modeled on an endogenous miRNA gene. The miRNA gene encodes an RNA that forms a hairpin structure containing a 22-nucleotide sequence that is complementary to another endogenous gene (target sequence). For suppression of RAP2.7 expression, the 22-nucleotide sequence is selected from a RAP2.7 transcript sequence and contains 22 nucleotides of said RAP2.7 sequence in sense orientation and 21 nucleotides of a corresponding antisense sequence that is complementary to the sense sequence.

miRNA molecules are highly efficient at inhibiting the expression of endogenous genes, and the RNA interference they induce is inherited by subsequent generations of plants.

The availability of reverse genetics systems, which are well-known in the art, makes the generation and isolation of down-regulated or null mutants feasible, given the availability of a defined nucleic acid sequence, as provided herein. One such system (the Trait Utility System for Corn, i.e., *TUSC*) is based on successful systems from other organisms (Ballinger et al., Proc. Natl. Acad. Sci. USA, 86, 9402-9406 (1989); Kaiser et al. (1990), Proc. Natl. Acad. Sci. USA, 87, 1686-1690; and Rushforth et al., Mol. Cell. Biol., 13, 902-910 (1993)). The central feature of the system is to identify Mu transposon insertions within a DNA sequence of interest in anticipation that at least some of these insertion alleles will be mutants. See US Patents Numbers 6,300,542 and 5,962,764. To develop the system, DNA was collected from a large population of Mutator transposon stocks that were then self-pollinated to produced F2 seed. To find Mu transposon insertions within a specific DNA sequence, the collection of DNA samples is screened via PCR using a gene-specific primer and a primer that anneals to the inverted repeats of Mu transposons. A PCR product is expected only when the template DNA comes from a plant that contains a Mu transposon insertion within the target gene. Once such a DNA sample is identified, F2 seed from the corresponding plant is screened for a transposon insertion allele. Transposon insertion mutations of the *an1* gene have been obtained via the *TUSC* procedure (Bensen et al. (1995)). This system is applicable to other plant species, at times modified in accordance with knowledge and skills reasonably attributed to ordinary artisans.

The use of the term "nucleotide constructs" herein is not intended to limit the present invention to nucleotide constructs comprising DNA. Those of ordinary skill in the art will recognize that nucleotide constructs, particularly polynucleotides and oligonucleotides, comprised of ribonucleotides and combinations of ribonucleotides and deoxyribonucleotides may also be employed in the methods disclosed herein. Thus, the nucleotide constructs of the present invention encompass all nucleotide constructs that can be employed in the methods of the present invention for transforming plants including, but not limited to, those comprised of deoxyribonucleotides, ribonucleotides, and combinations thereof. Such deoxyribonucleotides and ribonucleotides include both naturally occurring molecules and synthetic analogues. The nucleotide constructs of the invention also encompass all forms of nucleotide constructs including, but not limited to, single-stranded forms, double-stranded forms, hairpins, stem-and-loop structures, and the like.

Furthermore, it is recognized that the methods of the invention may employ a nucleotide construct that is capable of directing, in a transformed plant, the expression of at least one protein, or at least one RNA, such as, for example, an antisense RNA that is complementary to at least a portion of an mRNA. Typically such a nucleotide construct is comprised of a coding sequence for a protein or an RNA operably linked to 5' and 3' transcriptional regulatory regions. Alternatively, it is also recognized that the methods of the invention may employ a nucleotide construct that is not capable of directing, in a transformed plant, the expression of a protein or an RNA.

In addition, it is recognized that methods of the present invention do not depend on the incorporation of the entire nucleotide construct into the genome. Rather, the methods of the present invention only require that the plant or cell thereof is altered as a result of the introduction of the nucleotide construct into a cell. In one embodiment of the invention, the genome may be altered following the introduction of the nucleotide construct into a cell. or example, the nucleotide construct, or any part thereof, may incorporate into the genome of the plant. Alterations to the genome of the present invention include, but are not limited to, additions, deletions, and substitutions of nucleotides in the genome. While the methods of the present invention do not depend on additions, deletions, or substitutions of any particular number of nucleotides, it is recognized that such additions, deletions, or substitutions comprise at least one nucleotide.

In certain embodiments the nucleic acid sequences of the present invention can be used in combination ("stacked") with other polynucleotide sequences of interest in order to create plants with a desired phenotype. The combinations generated can include multiple copies of any one or more of the polynucleotides of interest. The polynucleotides of the present invention may be stacked with any gene or combination of genes to produce plants with a variety of desired trait combinations, including but not limited to traits desirable for animal feed such as high oil genes (*e.g.*, U.S. Patent No. 6,232,529); balanced amino acids (*e.g.* hordothionins (U.S. Patent Nos. 5,990,389; 5,885,801; 5,885,802; and 5,703,409); barley high lysine (Williamson et al. (1987) *Eur. J. Biochem.* 165:99-106; and WO 98/20122); and high methionine proteins (Pedersen et al. (1986) *J. Biol. Chem.* 261:6279; Kirihaara et al. (1988) *Gene* 71:359; and Musumura et al. (1989) *Plant Mol. Biol.* 12: 123)); increased digestibility (*e.g.*, modified storage proteins (U.S. Application Serial No. 10/053,410, filed November 7, 2001); and thioredoxins (U.S. Application Serial No. 10/005,429, filed December 3, 2001)).

The polynucleotides of the present invention can also be stacked with traits

desirable for insect, disease or herbicide resistance (*e.g.*, *Bacillus thuringiensis* toxic proteins (U.S. Patent Nos. 5,366,892; 5,747,450; 5,737,514; 5,723,756; 5,593,881; Geiser et al (1986) *Gene* 48:109); lectins (Van Damme et al. (1994) *Plant Mol. Biol.* 24:825); fumonisin detoxification genes (U.S. Patent No. 5,792,931); avirulence and disease resistance genes (Jones et al. (1994) *Science* 266:789; Martin et al. (1993) *Science* 262:1432; Mindrinos et al. (1994) *Cell* 78:1089); acetolactate synthase (ALS) mutants that lead to herbicide resistance such as the S4 and/or Hra mutations; inhibitors of glutamine synthase such as phosphinothricin or basta (*e.g.*, bar gene); and glyphosate resistance (EPSPS gene)); and traits desirable for processing or process products such as high oil (*e.g.*, U.S. Patent No. 6,232,529); modified oils (*e.g.*, fatty acid desaturase genes (U.S. Patent No. 5,952,544; WO 94/11516)); modified starches (*e.g.*, ADPG pyrophosphorylases (AGPase), starch synthases (SS), starch branching enzymes (SBE) and starch debranching enzymes (SDBE)); and polymers or bioplastics (*e.g.*, U.S. patent No. 5,602,321; beta-ketothiolase, polyhydroxybutyrate synthase, and acetoacetyl-CoA reductase (Schubert et al. (1988) *J. Bacteriol.* 170:5837-5847) facilitate expression of polyhydroxyalkanoates (PHAs)).

One could also combine the polynucleotides of the present invention with polynucleotides affecting agronomic traits such as male sterility (*e.g.*, see U.S. Patent No. 5,583,210), stalk strength, flowering time, or transformation technology traits such as cell cycle regulation or gene targeting (*e.g.* WO 99/61619; WO 00/17364; WO 99/25821).

These stacked combinations can be created by any method, including but not limited to cross breeding plants by any conventional or TopCross methodology, or genetic transformation. If the traits are stacked by genetically transforming the plants, the polynucleotide sequences of interest can be combined at any time and in any order. For example, a transgenic plant comprising one or more desired traits can be used as the target to introduce further traits by subsequent transformation. The traits can be introduced simultaneously in a co-transformation protocol with the polynucleotides of interest provided by any combination of transformation cassettes. For example, if two sequences will be introduced, the two sequences can be contained in separate transformation cassettes (trans) or contained on the same transformation cassette (cis). Expression of the sequences of interest can be driven by the same promoter or by different promoters. In certain cases, it may be desirable to introduce a transformation cassette that will suppress the expression of a polynucleotide of interest. This may be accompanied by any combination of other

suppression cassettes or overexpression cassettes to generate the desired combination of traits in the plant.

The nucleotide constructs of the invention also encompass nucleotide constructs that may be employed in methods for altering or mutating a genomic nucleotide sequence in an organism, including, but not limited to, chimeric vectors, chimeric mutational vectors, chimeric repair vectors, mixed-duplex oligonucleotides, self-complementary chimeric oligonucleotides, and recombinogenic oligonucleobases. Such nucleotide constructs and methods of use, such as, for example, chimeraplasty, are known in the art. Chimeraplasty involves the use of such nucleotide constructs to introduce site-specific changes into the sequence of genomic DNA within an organism. eg, U.S. Patent Nos. 5,565,350; 5,731,181; 5,756,325; 5,760,012; 5,795,972; and 5,871,984..

See also, WO 98/49350, WO 99/07865, WO 99/25821, and Beetham et al. (1999) *Proc. Natl. Acad. Sci. USA* 96:8774-8778.

The following examples are offered by way of illustration and not by way of limitation.

15

Other embodiments of the invention include the use of VGT1 and its alternate forms described herein as markers to screen for and identify plants which may have altered maturity. The invention thus relates to genetic markers for plants with altered maturity. The markers represent polymorphic variants of the non-coding regulatory element VGT1 that are associated with RAP2.7 regulation and thus provides a method of genotyping plants to determine those more likely to have flowering time that is altered from wildtype.

Thus, the invention relates to genetic markers and methods of identifying those markers in plants, whereby the plant is more likely to have a maturity that is earlier than normal by means of a mutant VGT1 which does not down regulate RAP2.7 appropriately.

Any method of identifying the presence or absence of these markers may be used, including, for example, single-strand conformation polymorphism (SSCP) analysis, base excision sequence scanning (BESS), RFLP analysis, heteroduplex analysis, denaturing gradient gel electrophoresis, and temperature gradient electrophoresis, allelic PCR, ligase chain reaction direct sequencing, mini sequencing, nucleic acid hybridization, micro-array-type detection of the VGT1 regulatory element.

The following is a general overview of techniques which can be used to assay for the polymorphisms of the invention.

In the present invention, a sample of genetic material is obtained from a plant.

Isolation and Amplification of Nucleic Acid

Samples of genomic DNA are isolated from any convenient source including any
5 suitable cell or tissue sample with intact interphase nuclei or metaphase cells. The cells
can be obtained from solid tissue as from a fresh or preserved plant part or from a tissue
sample. The sample can contain compounds which are not naturally intermixed with the
biological material such as preservatives, anticoagulants, buffers, fixatives, nutrients,
antibiotics, or the like.

10 Methods for isolation of genomic DNA from these various sources are described
in, for example, Kirby, *DNA Fingerprinting, An Introduction*, W.H. Freeman & Co. New
York (1992). Genomic DNA can also be isolated from cultured primary or secondary
cell cultures or from transformed cell lines derived from any of the aforementioned
tissue samples.

15 Samples of RNA can also be used. RNA can be isolated from tissues as
described in Sambrook et al., *supra*. RNA can be total cellular RNA, mRNA, poly A+
RNA, or any combination thereof. For best results, the RNA is purified, but can also be
unpurified cytoplasmic RNA. RNA can be reverse transcribed to form DNA which is
then used as the amplification template, such that the PCR indirectly amplifies a specific
20 population of RNA transcripts. See, e.g., Sambrook, *supra*, Kawasaki et al., Chapter 8 in
PCR Technology, (1992) *supra*, and Berg et al., *Hum. Genet.* 85:655-658 (1990).

PCR Amplification

The most common means for amplification is polymerase chain reaction (PCR),
25 as described in U.S. Pat. Nos. 4,683,195; 4,683,202; and 4,965,188,

Tissues should be roughly minced using a sterile,
disposable scalpel and a sterile needle (or two scalpels) in a 5 mm Petri dish. Procedures
for removing paraffin from tissue sections are described in a variety of specialized
handbooks well known to those skilled in the art.

30 To amplify a target nucleic acid sequence in a sample by PCR, the sequence must
be accessible to the components of the amplification system. Kits for the extraction of
high-molecular weight DNA for PCR include a Genomic Isolation Kit A.S.A.P.
(Boehringer Mannheim, Indianapolis, Ind.), Genomic DNA Isolation System (GIBCO
BRL, Gaithersburg, Md.), Elu-Quik DNA Purification Kit (Schleicher & Schuell, Keene,

N.H.), DNA Extraction Kit (Stratagene, LaJolla, Calif.), TurboGen Isolation Kit (Invitrogen, San Diego, Calif.), and the like. Use of these kits according to the manufacturer's instructions is generally acceptable for purification of DNA prior to practicing the methods of the present invention.

5 The concentration and purity of the extracted DNA can be determined by spectrophotometric analysis of the absorbance of a diluted aliquot at 260 nm and 280 nm. After extraction of the DNA, PCR amplification may proceed. The first step of each cycle of the PCR involves the separation of the nucleic acid duplex formed by the primer extension. Once the strands are separated, the next step in PCR involves hybridizing the
10 separated strands with primers that flank the target sequence. The primers are then extended to form complementary copies of the target strands. For successful PCR amplification, the primers are designed so that the position at which each primer hybridizes along a duplex sequence is such that an extension product synthesized from one primer, when separated from the template (complement), serves as a template for the
15 extension of the other primer. The cycle of denaturation, hybridization, and extension is repeated as many times as necessary to obtain the desired amount of amplified nucleic acid.

 In a particularly useful embodiment of PCR amplification, strand separation is achieved by heating the reaction to a sufficiently high temperature for a sufficient time to
20 cause the denaturation of the duplex but not to cause an irreversible denaturation of the polymerase (see U.S. Pat. No. 4,965,188). Typical heat denaturation involves temperatures ranging from about 80°C to 105°C for times ranging from seconds to minutes. Strand separation, however, can be accomplished by any suitable denaturing method including physical, chemical, or enzymatic means.
25 Strand separation may be induced by a helicase, for example, or an enzyme capable of exhibiting helicase activity. For example, the enzyme RecA has helicase activity in the presence of ATP. The reaction conditions suitable for strand separation by helicases are known in the art (see Kuhn Hoffman-Berling, 1978, *CSH-Quantitative Biology*, 43:63-67; and Radding, 1982, *Ann. Rev. Genetics* 16:405-436).

30

 Template-dependent extension of primers in PCR is catalyzed by a polymerizing agent in the presence of adequate amounts of four deoxyribonucleotide triphosphates (typically dATP, dGTP, dCTP, and dTTP) in a reaction medium comprised of the appropriate salts, metal cations, and pH buffering systems. Suitable polymerizing agents

are enzymes known to catalyze template-dependent DNA synthesis. In some cases, the target regions may encode at least a portion of a protein expressed by the cell. In this instance, mRNA may be used for amplification of the target region. Alternatively, PCR can be used to generate a cDNA library from RNA for further amplification, the initial
5 template for primer extension is RNA. Polymerizing agents suitable for synthesizing a complementary, copy-DNA (cDNA) sequence from the RNA template are reverse transcriptase (RT), such as avian myeloblastosis virus RT, Moloney murine leukemia virus RT, or *Thermus thermophilus* (Tth) DNA polymerase, a thermostable DNA polymerase with reverse transcriptase activity marketed by Perkin Elmer Cetus, Inc.

10 Typically, the genomic RNA template is heat degraded during the first denaturation step after the initial reverse transcription step leaving only DNA template. Suitable polymerases for use with a DNA template include, for example, *E. coli* DNA polymerase I or its Klenow fragment, T4 DNA polymerase, Tth polymerase, and *Taq* polymerase, a heat-stable DNA polymerase isolated from *Thermus aquaticus* and commercially
15 available from Perkin Elmer Cetus, Inc. The latter enzyme is widely used in the amplification and sequencing of nucleic acids. The reaction conditions for using *Taq* polymerase are known in the art and are described in Gelfand, 1989, *PCR Technology*, *supra*.

20 Allele Specific PCR

Allele-specific PCR differentiates between target regions differing in the presence of absence of a variation or polymorphism. PCR amplification primers are chosen which bind only to certain alleles of the target sequence. This method is described by Gibbs, *Nucleic Acid Res.* 17:12427-2448 (1989).

25 Allele Specific Oligonucleotide Screening Methods

Further diagnostic screening methods employ the allele-specific oligonucleotide (ASO) screening methods, as described by Saiki et al., *Nature* 324:163-166 (1986). Oligonucleotides with one or more base pair mismatches are generated for any particular
30 allele. ASO screening methods detect mismatches between variant target genomic or PCR amplified DNA and non-mutant oligonucleotides, showing decreased binding of the oligonucleotide relative to a mutant oligonucleotide. Oligonucleotide probes can be designed so that under low stringency, they will bind to both polymorphic forms of the allele, but at high stringency, bind to the allele to which they correspond. Alternatively,

stringency conditions can be devised in which an essentially binary response is obtained, i.e., an ASO corresponding to a variant form of the target gene will hybridize to that allele, and not to the wild-type allele.

5 Ligase Mediated Allele Detection Method

Target regions of a test subject's DNA can be compared with target regions in unaffected and affected family members by ligase-mediated allele detection. See Landegren et al., *Science* 241:107-1080 (1988). Ligase may also be used to detect point mutations in the ligation amplification reaction described in Wu et al., *Genomics* 4:560-10 569 (1989). The ligation amplification reaction (LAR) utilizes amplification of specific DNA sequence using sequential rounds of template dependent ligation as described in Wu, *supra*, and Barany, *Proc. Nat. Acad. Sci.* 88:189-193 (1990).

Denaturing Gradient Gel Electrophoresis

15 Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. DNA molecules melt in segments, termed melting domains, under conditions of increased temperature or denaturation. Each20 melting domain melts cooperatively at a distinct, base-specific melting temperature (T_m). Melting domains are at least 20 base pairs in length, and may be up to several hundred base pairs in length.

Differentiation between alleles based on sequence specific melting domain differences can be assessed using polyacrylamide gel electrophoresis, as described in25 Chapter 7 of Erlich, ed., *PCR Technology*, "Principles and Applications for DNA Amplification", W.H. Freeman and Co., New York (1992).

Generally, a target region to be analyzed by denaturing gradient gel electrophoresis is amplified using PCR primers flanking the target region. The amplified30 PCR product is applied to a polyacrylamide gel with a linear denaturing gradient as described in Myers et al., *Meth. Enzymol.* 155:501-527 (1986), and Myers et al., in *Genomic Analysis, A Practical Approach*, K. Davies Ed. IRL Press Limited, Oxford, pp. 95-139 (1988).

The

electrophoresis system is maintained at a temperature slightly below the T_m of the melting domains of the target sequences.

In an alternative method of denaturing gradient gel electrophoresis, the target sequences may be initially attached to a stretch of GC nucleotides, termed a GC clamp, as described in Chapter 7 of Erlich, *supra*. Preferably, at least 80% of the nucleotides in the GC clamp are either guanine or cytosine. Preferably, the GC clamp is at least 30 bases long. This method is particularly suited to target sequences with high T_m 's.

Generally, the target region is amplified by the polymerase chain reaction as described above. One of the oligonucleotide PCR primers carries at its 5' end, the GC clamp region, at least 30 bases of the GC rich sequence, which is incorporated into the 5' end of the target region during amplification. The resulting amplified target region is run on an electrophoresis gel under denaturing gradient conditions as described above. DNA fragments differing by a single base change will migrate through the gel to different positions, which may be visualized by ethidium bromide staining.

Temperature Gradient Gel Electrophoresis

Temperature gradient gel electrophoresis (TGGE) is based on the same underlying principles as denaturing gradient gel electrophoresis, except the denaturing gradient is produced by differences in temperature instead of differences in the concentration of a chemical denaturant. Standard TGGE utilizes an electrophoresis apparatus with a temperature gradient running along the electrophoresis path. As samples migrate through a gel with a uniform concentration of a chemical denaturant, they encounter increasing temperatures. An alternative method of TGGE, temporal temperature gradient gel electrophoresis (TTGE or tTGGE) uses a steadily increasing temperature of the entire electrophoresis gel to achieve the same result. As the samples migrate through the gel the temperature of the entire gel increases, leading the samples to encounter increasing temperature as they migrate through the gel. Preparation of samples, including PCR amplification with incorporation of a GC clamp, and visualization of products are the same as for denaturing gradient gel electrophoresis.

Single-Strand Conformation Polymorphism Analysis

Target sequences or alleles at the VGT1 loci can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single-stranded PCR products, as described in

Orita et al., *Proc. Nat. Acad. Sci.* 85:2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single-stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. Thus, electrophoretic mobility of single-stranded amplification products can detect base-sequence difference between alleles or target sequences.

Chemical or Enzymatic Cleavage of Mismatches

Differences between target sequences can also be detected by differential chemical cleavage of mismatched base pairs, as described in Grompe et al., *Am. J. Hum. Genet.* 48:212-222 (1991). In another method, differences between target sequences can be detected by enzymatic cleavage of mismatched base pairs, as described in Nelson et al., *Nature Genetics* 4:11-18 (1993). Briefly, genetic material from a plant and an affected family member may be used to generate mismatch free heterohybrid DNA duplexes. As used herein, "heterohybrid" means a DNA duplex strand comprising one strand of DNA from one plant, and a second DNA strand from another plant, usually a plant differing in the phenotype for the trait of interest. Positive selection for heterohybrids free of mismatches allows determination of small insertions, deletions or other polymorphisms that may be associated with VGT1 polymorphisms.

Non-gel Systems

Other possible techniques include non-gel systems such as TAQMAN™ (Perkin Elmer). In this system, oligonucleotide PCR primers are designed that flank the mutation in question and allow PCR amplification of the region. A third oligonucleotide probe is then designed to hybridize to the region containing the base subject to change between different alleles of the gene. This probe is labeled with fluorescent dyes at both the 5' and 3' ends. These dyes are chosen such that while in this proximity to each other the fluorescence of one of them is quenched by the other and cannot be detected. Extension by *Taq* DNA polymerase from the PCR primer positioned 5' on the template relative to the probe leads to the cleavage of the dye attached to the 5' end of the annealed probe through the 5' nuclease activity of the *Taq* DNA polymerase. This removes the quenching effect allowing detection of the fluorescence from the dye at the 3' end of the probe. The discrimination between different DNA sequences arises through the fact that if the hybridization of the probe to the template molecule is not

complete, i.e., there is a mismatch of some form, the cleavage of the dye does not take place. Thus, only if the nucleotide sequence of the oligonucleotide probe is completely complementary to the template molecule to which it is bound will quenching be removed. A reaction mix can contain two different probe sequences each designed against different
5 alleles that might be present thus allowing the detection of both alleles in one reaction.

Yet another technique includes an Invader Assay, which includes isothermal amplification that relies on a catalytic release of fluorescence. See Third Wave Technology at world wide web at twt.com.

10 Non-PCR Based DNA screening

Hybridization probes are generally oligonucleotides which bind through complementary base pairing to all or part of a target nucleic acid. Probes typically bind target sequences lacking complete complementarity with the probe sequence depending on the stringency of the hybridization conditions. The probes are preferably labeled
15 directly or indirectly, such that by assaying for the presence or absence of the probe, one can detect the presence or absence of the target sequence. Direct labeling methods include radioisotope labeling, such as with P³² or S³⁵. Indirect labeling methods include fluorescent tags, biotin complexes which may be bound to avidin or streptavidin, or peptide or protein tags. Visual detection methods include photoluminescents, Texas red,
20 rhodamine and its derivatives, red leuco dye and 3,3',5,5'-tetramethylbenzidine (TMB), fluorescein, and its derivatives, dansyl, umbelliferone and the like or with horse radish peroxidase, alkaline phosphatase and the like.

Hybridization probes include any nucleotide sequence capable of hybridizing to the chromosome where VGT1 resides, and thus defining a genetic marker linked
25 to VGT1, including a restriction fragment length polymorphism, a hypervariable region, repetitive element, or a variable number tandem repeat. Further suitable hybridization probes include exon fragments or portions of cDNAs or genes known to map to the relevant region of the chromosome.

Preferred tandem repeat hybridization probes for use according to the present
30 invention are those that recognize a small number of fragments at a specific locus at high stringency hybridization conditions, or that recognize a larger number of fragments at that locus when the stringency conditions are lowered.

One or more additional restriction enzymes and/or probes and/or primers can be used. Additional enzymes, constructed probes, and primers can be determined by routine

experimentation by those of ordinary skill in the art and are intended to be within the scope of the invention.

Although the methods described herein may be in terms of the use of a single restriction enzyme and a single set of primers, the methods are not so limited. One or more additional restriction enzymes and/or probes and/or primers can be used, if desired. Indeed, in some situations it may be preferable to use combinations of markers giving specific haplotypes. Additional enzymes, constructed probes and primers can be determined through routine experimentation, combined with the teachings provided and incorporated herein.

The sequences surrounding the polymorphism will facilitate the development of alternate PCR tests in which a primer of about 4-30 contiguous bases taken from the sequence immediately adjacent to the polymorphism is used in connection with a polymerase chain reaction to greatly amplify the region before treatment with the desired restriction enzyme. The primers need not be the exact complement; substantially equivalent sequences are acceptable. The design of primers for amplification by PCR is known to those of skill in the art and is discussed in detail in Ausubel (ed.), *Short Protocols in Molecular Biology*, 4th Edition, John Wiley and Sons (1999).

The following is a brief description of primer design.

Primer Design Strategy

Increased use of polymerase chain reaction (PCR) methods has stimulated the development of many programs to aid in the design or selection of oligonucleotides used as primers for PCR. Four examples of such programs that are freely available via the Internet are: PRIMER by Mark Daly and Steve Lincoln of the Whitehead Institute (UNIX, VMS, DOS, and Macintosh), Oligonucleotide Selection Program (OSP) by Phil Green and LaDeana Hiller of Washington University in St. Louis (UNIX, VMS, DOS, and Macintosh), PGEN by Yoshi (DOS only), and Amplify by Bill Engels of the University of Wisconsin (Macintosh only). Generally these programs help in the design of PCR primers by searching for bits of known repeated-sequence elements and then optimizing the T_m by analyzing the length and GC content of a putative primer. Commercial software is also available and primer selection procedures are rapidly being included in most general sequence analysis packages.

Sequencing and PCR Primers

Designing oligonucleotides for use as either sequencing or PCR primers requires selection of an appropriate sequence that specifically recognizes the target, and then testing the sequence to eliminate the possibility that the oligonucleotide will have a stable secondary structure. Inverted repeats in the sequence can be identified using a repeat-identification or RNA-folding program such as those described above. If a possible stem structure is observed, the sequence of the primer can be shifted a few nucleotides in either direction to minimize the predicted secondary structure. The sequence of the oligonucleotide should also be compared with the sequences of both strands of the appropriate vector and insert DNA. Obviously, a sequencing primer should only have a single match to the target DNA. It is also advisable to exclude primers that have only a single mismatch with an undesired target DNA sequence. For PCR primers used to amplify genomic DNA, the primer sequence should be compared to the sequences in the GenBank database to determine if any significant matches occur. If the oligonucleotide sequence is present in any known DNA sequence or, more importantly, in any known repetitive elements, the primer sequence should be changed.

The following examples serve to better illustrate the invention and are not intended to limit the scope of the invention in any way.

- Salvi S., Morgante M., Fengler K., Meeley R., Ananiev E., Svtashev S., Bruggemann E., Niu X., Li B., Tingey SV., Tomes D., Miao G.-H., Phillips RL., Tuberosa R. Progress in the positional cloning of *Vgt1*, a QTL controlling flowering time in maize (2003). Proceedings of 57th Corn and Sorghum and 32nd Soybean Seed Research Conference. Dec. 11-13, 2002, Chicago.
- Phillips R. L., Kim T. S., Kaeppler S. M., Parentoni S. N., Shaver D. L., Stucker R. I. and Openshaw S. J. 1992. Genetic dissection of maturity using RFLPs. Proc. 47th Ann. Corn and Sorghum Res. Conf. 47:135-150.
- Vladutu, C., McLaughlin J. and Phillips R.L. 1999. Fine Mapping and Characterization of Linked Quantitative Trait Loci Involved in the Transition of the Maize Apical Meristem From Vegetative to Generative Structures. Genetics 153: 993-1007.
- Salvi S., Tuberosa R., Chiapparino E., Maccaferri M., Veillet S., van Beuningen L., Isaac P., Edward K.J., Phillips R.L. (2002). Toward positional cloning of *Vgt1*, a QTL controlling the transition from the vegetative to the reproductive phase in maize. Plant Mol Biol 48:601-613.

Examples**Example 1**

RAP2.7 expression level is associated with differences in maturity

5 It was determined that RAP2.7 gene expression level determines the transition to flowering in plants and that *vgt1* is a *cis*-element that regulates RAP2.7 transcription. Two plants with different maturities (N28 and C22-4) were then screened to identify if the RAP2.7 expression levels differ between them.

10 RNA was synthesized from N28 and C22-4 from different tissues and stages of development, and RAP2.7 expression was measured by RT-PCR.

Tissue Types:

Mature leaves - exposed, blades + sheaths

Immature leaves - in whorl, blades + sheaths

15 Shoot apical meristems

Roots - whole, including root apical meristems

Stalks - leftovers

Developmental Stages

20

Sample Number

Genotype	1	2	3	4	5
N28 (late)	veg	veg	veg	trans	rep
C22-4 (early)	veg	trans	rep	rep	rep

25

Gene Expression Assay

First total RNA was prepared from frozen tissue samples. cDNA was then made by reverse transcriptase and the PAR2.7 region was amplified using PCR with agarose gel and ethidium bromide staining. Band fluorescence was quantified and tubulin was used as
30 an internal control, and to normalize expression levels.

Figure 1 shows the levels of RAP2.7 expression at Day: 14 before transition, C22-4 on transition at 20 days, and N28 at transition at 27 days. Last two dates have 3 samples of each. P values are significantly different at the first two sampling dates, but not at the

latest date. There is RAP2.7 signal from meristems, and other tissues. RAP2.7 is a repressor of flowering, and must be below a particular level for flowering to occur.

The results indicate that RAP2.7 is expressed in every tissue type. RAP2.7 expression levels are lower in C22-4 than in N28 before the reproductive transition (mature leaves), and RAP2.7 expression levels decrease during development in both C22-4 and N28. Thus, overexpression of RAP2.7 will delay transition from vegetative state to flowering.

Example 2

Over-expression of maize RAP2.7 under a moderate-strength constitutive promoter

The cDNA sequences of RAP2.7 were obtained from a RT-PCR experiment. Specifically, total RNA from C22-4 leaves was isolated and used as template in RT-PCR with gene-specific primers. The gene-specific primers were designed based on RAP2.7 genomic sequences from B73 genotype. The primer sequences are sense – ATGCAGTTGGATCTGAACGT (SEQ ID NO: 9) and antisense – GCCATCACCATCCCCGCTGA (SEQ ID NO:10).

The RT-PCR amplified fragment of RAP2.7, including the entire 1371-bp sequence including the ATG start codon and the TAG stop codon, was fused to the rice actin promoter and pinII terminator to produce an expression cassette. This expression cassette was then linked to a selectable marker cassette containing a bar gene driven by CaMV 35S promoter and a pinII terminator in Figure 2.

Transgenic maize plants were produced by transforming Immature GS3 maize embryos with this expression cassette, using the Agrobacterium-based transformation method described as below.

For Agrobacterium-mediated transformation of maize with the expression cassette comprising the rice actin promoter operably linked to the maize RAP2.7 gene, the method of Zhao was employed (U.S. Patent No. 5,981,840, and PCT patent publication

WO98/32326.). While the method below is described for the transformation of maize plants with the actin promoter – RAP2.7 expression cassette, those of ordinary skill in the art recognize that this method can be used to produce transformed maize plants with any nucleotide construct or

expression cassette of the invention that comprises a promoter operably linked to maize RAP2.7 gene for expression in a plant.

Briefly, immature embryos were isolated from maize and the embryos contacted with a suspension of *Agrobacterium*, where the bacteria are capable of transferring the
5 ACTIN promoter-RAP2.7 expression cassette (illustrated above) to at least one cell of at least one of the immature embryos (step 1: the infection step). In this step the immature embryos were immersed in an *Agrobacterium* suspension for the initiation of inoculation. The embryos were co-cultured for a time with the *Agrobacterium* (step 2: the co-cultivation step). The immature embryos were cultured on solid medium following the
10 infection step. Following this co-cultivation period an optional "resting" step was included. In this resting step, the embryos were incubated in the presence of at least one antibiotic known to inhibit the growth of *Agrobacterium* without the addition of a selective agent for plant transformants (step 3: resting step). The immature embryos were cultured on solid medium with antibiotic, but without a selecting agent, for elimination of
15 *Agrobacterium* and for a resting phase for the infected cells. Next, inoculated embryos were cultured on medium containing a selective agent and growing transformed callus was recovered (step 4: the selection step). Preferably, the immature embryos were cultured on solid medium with a selective agent resulting in the selective growth of transformed cells. The resulting calli were then regenerated into plants by culturing the calli on solid,
20 selective medium (step 5: the regeneration step).

The transformation produced 15 events in GS3XGaspe flint, an early-flowering genotype. Transformation in GS3XHC69, a normal maturity genotype, was problematic during shoot regeneration process and only produced 5 events. The problem was presumed to be related to the transgene. Two of the 5 events had delays in flowering time
25 for up to 30 days. In GS3XGaspe flint genotype, majority (16 out of 20) of the events had various degrees of delay in flowering time, from 16 days to 36 days. In almost all events with delayed flowering, there was significant change in plant architecture, mostly increase in plant height and internode length.

Ectopic expression of Rap2.7 with the rice actin promoter resulted in over
30 expression of the gene, and significantly delayed flowering as measured by the number of leaves (nodes) produced prior to flowering. According to the invention, vectors for transformation were produced in two corn genetic backgrounds (GS3 X HC69, and GS3 X GF, gaspe flint) using the Rap2.7 structural gene with the rice actin promoter.

A brief summary of the T0 phenotype from over-expressing RAP2.7 in maize is shown in the following tables. As a reference, GS3XGaspe plants coming out of tissue culture get pollinated (exuding silks) within 45 – 60 days, and have 10 or less leaves. Since these are T0 plants, accurate counts for leaf number and days to flowering are not possible. The plants that were late in flowering also had substantial increase in internode elongation, resulting in increase in plant height. These plants also had delayed senescence.

Events	GS3XHC69	GS3XGaspe
Late	2	16
Total	5	20

Days to Pollination	GS3XGaspe Events
< 60 days	4
61 - 70 days	5
71 - 80 days	7
> 81 days	1
no ear	3

Leaf Number (estimated)	GS3XGaspe Events
<10	4
10 - 12	10
13 -14	4
> 15	2

Example 3

Down-regulation of maize RAP2.7 by RNA interference

Two fragments from the cDNA sequences of RAP2.7 from genotype C22-4 were cloned by PCR to create an inverted repeat as illustrated below. Specifically, a 955-bp fragment starting at 9-bp downstream from ATG was cloned by PCR and designated as ZM-RAP2.7 (TR1) as a truncated form. Another fragment, 499-bp in length starting from the same position, was cloned and designated as ZM-RAP2.7 (IR1) for inverted repeat Figure 5.

Figure 5 Illustration of construction of the RAP2.7 gene fragments for RNA interference vector. TR1 fragment was then ligated to IR1, with IR1 in reverse orientation. The ligated 2-piece fragment was then linked to rice actin promoter with actin 5'-UTR and actin intron1 to create a full expression cassette. This expression cassette was then linked to a selectable marker cassette containing a bar gene driven by CaMV 35S promoter and a pinII terminator.

Figure 6 is an illustration of the vector used in transformation. Note only the portion between the right and left T-DNA borders (RB, LB) is shown. Transgenic maize plants were produced by transforming Immature GS3 maize embryos with this expression cassette, using the Agrobacterium-based transformation method described for the RAP2.7 over-expression study. The transformation produced 20 transgenic events from GS3XHC69, a genotype with a normal maturity; and 15 events from GS3XGaspe flint, an early-flowering genotype. Based on preliminary observation on T0 plants, all 15 events from the GS3XGaspe flint background showed no visible change in flowering time. However, 9 out of the 20 GS3XHC69 transgenic events had various degrees of early flowering phenotype. The earliest event flowered approximately 2 week earlier compared to other events with the same construct, and to other unrelated transgenic plants in the same greenhouse room. All plants had normal plant architecture.

The T0 data shows that if you down regulate RAP2.7 earlier flowering is observed. The following table shows the next (T1) generation of plants. We can see from this data that the phenotype is heritable and stable and the initial trend of decreased days to pollination is still observed.

PHP21842 T1 Phenotype

Event-Plant	Transgene	Leaf #	Days to Pollination
1-1	+	15	58
1-2	+	13	55
1-3	+	15	57
2-1	+	14	57
2-2	+	14	54
2-3	+	13	53
3-1	+	14	57
3-2	+	13	57
3-3	+	14	59
Control average (n=20)	+	19.2	66

Example 4

Positional cloning of Vgt1.

Positional cloning was completed after mapping Vgt1 in a 1.3-cM interval flanked by two AFLP markers (Salvi et al., 2002), in a cross between two nearly isogenic lines, N28 and its early derivative C22-4 (Vladutu et al., 1999). Following BAC library screening and the analysis of the relevant BAC contig, further development of new markers and genetic mapping allowed for the delimiting of Vgt1 within a ca. 2.7-kb region (fig. 7.)

Sequencing of the relevant BAC clone and of the corresponding DNA region from the parental lines involved in the cross showed that Vgt1 lies within an intergenic region, ca. 75 kb upstream of an Ap-2 like gene (Rap2.7) and ca. 10 kb downstream of a RAD51-like gene. The 2.7 kb region found to be completely associated with Vgt1 is essentially a low-copy region with a number of polymorphisms between N28 and C22-4 (one of the polymorphisms is caused by the insertion of a MITE transposable element in C22-4). Surprisingly, this sequence does not code for any known protein. It is hypothesized to either be a RNAi element or a regulatory RNA or DNA element that either directly regulates expression of flowering genes such as Rap2.7 or specifically targets expression of other genes which control flowering genes such as Rap2.7.

Example 5

Vgt1 is associated with maturity shift in inbred lines.

The information gathered with the positional cloning study allowed the testing of the 2.7 kb region as candidate for controlling flowering time through association mapping. Several SNPs and other polymorphisms identified at and around Vgt1 were screened on a panel of ca. 100 lines representative of cultivated maize germplasm (Remington et al., 2001) and a panel of elite proprietary lines.

Linkage disequilibrium at the Vgt1 region was quickly dissipated over distance of ca. 1 kb within the panel of 100 lines. Association analysis showed that among the genes and sequence at and around Vgt1, the only DNA region statistically associated with flowering time was a sub-region of ca. 2 kb within the same 2.7 kb region identified by positional cloning (Fig. 8.).

This element thus can be used as a sequence-based marker to identify inbred and hybrids which have altered maturity.

DEMANDES OU BREVETS VOLUMINEUX

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

CECI EST LE TOME __1__ DE __2__

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__

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

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule that encodes a polypeptide having RAP2.7-
5 like activity, wherein said nucleic acid molecule is:
 - (a) a nucleic acid molecule comprising the sequence set forth in SEQ ID NOS: 1, 3, or 4;
 - (b) a nucleic acid molecule comprising a sequence encoding the amino acid sequence set forth in SEQ ID NO: 2;
 - 10 (c) a nucleic acid molecule comprising a sequence having at least 90% sequence identity to the full length nucleotide sequence set forth in SEQ ID NOS: 1, 3, or 4; or
 - (d) a nucleic acid molecule comprising a sequence of at least 1000 consecutive nucleic acids of any of (a) through (c),
15 wherein RAP2.7-like activity comprises modulation of onset of flowering time in plants.
2. A vector comprising the nucleic acid molecule of claim 1.
- 20 3. A transgenic plant cell having stably incorporated in its genome the isolated nucleic acid molecule of claim 1.
4. The plant cell of claim 3, wherein said plant cell is from a monocot plant.
- 25 5. The plant cell of claim 4, wherein said monocot plant is maize.
6. A method for altering flowering time in a plant comprising:
 - (i) transforming a plant cell with a vector comprising a nucleic acid molecule operably
30 linked to a promoter that regulates transcription of said nucleic acid molecule in a plant cell; wherein said nucleic acid molecule comprises:
 - (a) a nucleotide sequence comprising the sequence set forth in SEQ ID NOS: 1, 3, or 4;
 - (b) a nucleotide sequence encoding the amino acid sequence of SEQ ID NO:
35 2; or

- (c) a nucleotide sequence having at least 90% sequence identity to the full length sequence of SEQ ID NOS: 1, 3, or 4, wherein said nucleotide sequence encodes a protein which regulates flowering in plants and
- (ii) regenerating a stably transformed plant from the plant cell.

5

7. The method of claim 6, wherein said plant is a monocot.

8. The method of claim 7, wherein said monocot is maize.

10 9. The method of claim 6 wherein said vector over- expresses RAP2.7.

10. The method of claim 9 wherein said plant flowers later than a plant which does not have the vector incorporated therein.

15 11. The method of claim 6 wherein said vector inhibits expression of RAP2.7.

12. The method of claim 11 wherein said vector is an RNA interference vector.

13. The method of claim 12 wherein said plant flowers earlier than a plant which
20 does not have the RNA interference vector incorporated therein.

14. A method for altering onset of maturity of a plant, said method comprising:
transforming said plant with a vector comprising a heterologous nucleic acid molecule
operably linked to a promoter that induces transcription of said heterologous nucleic acid
25 molecule in a plant cell; and
regenerating stably transformed plants, wherein said heterologous nucleic acid molecule
comprises:

- (a) a nucleotide sequence comprising the sequence set forth in SEQ ID NOS: 1, 3 or 4;
- 30 (b) a nucleotide sequence comprising at least 900 contiguous nucleotides of the sequence of SEQ ID NOS: 1, 3, or 4, wherein said nucleotide sequence encodes a protein having flowering regulatory activity; or

- (c) a nucleotide sequence having at least 90% sequence identity to the full length sequence of SEQ ID NOS: 1, 3, or 4, wherein said nucleotide sequence encodes a protein having flowering regulatory activity.

- 5 15. The method of claim 14, wherein said promoter is a constitutive promoter.
16. The method of claim 14, wherein said promoter is a tissue-preferred promoter.
17. The method of claim 14, wherein said promoter is an inducible promoter.
- 10 18. The method of claim 14, wherein said plant is a monocot.
19. The method of claim 18, wherein said monocot is maize.
- 15 20. The method of claim 14 wherein said vector over- expresses RAP2.7.
21. The method of claim 20 wherein said plant matures later than a plant which does not have the vector incorporated therein.
- 20 22. The method of claim 14 wherein said vector inhibits expression of RAP2.7.
23. The method of claim 22 wherein said vector is an RNA interference vector.
24. The method of claim 22 wherein said plant matures earlier than a plant which
- 25 does not have the RNA interference vector incorporated therein.
25. An isolated polypeptide having flowering regulatory activity, wherein the polypeptide is:
- 30 (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2;
- (b) a polypeptide encoded by a nucleic acid molecule comprising the sequence set forth in SEQ ID NOS: 1, 3, or 4;

- (c) a polypeptide encoded by a nucleic acid molecule that has at least 80% sequence identity to the full length sequence set forth in SEQ ID NOS: 1, 3, or 4;
- (d) a polypeptide comprising an amino acid sequence having at least 90% sequence identity to the full length sequence set forth in SEQ ID NO:2; or
- (e) a polypeptide comprising an amino acid sequence of at least 200 consecutive amino acids of any of (a) through (d).
- 10 26. A nucleotide construct comprising:
a nucleic acid molecule encoding the polypeptide of parts (a) to (d) of claim 25,
wherein said nucleic acid molecule is operably linked to a promoter that drives expression
in a host cell.
- 15 27. A plant cell having stably incorporated in its genome the nucleotide construct of
claim 26.
28. The plant cell of claim 27, wherein said plant cell is from a monocot plant.
- 20 29. The plant cell of claim 28, wherein said monocot plant is maize.

1/25

FIGURE 1

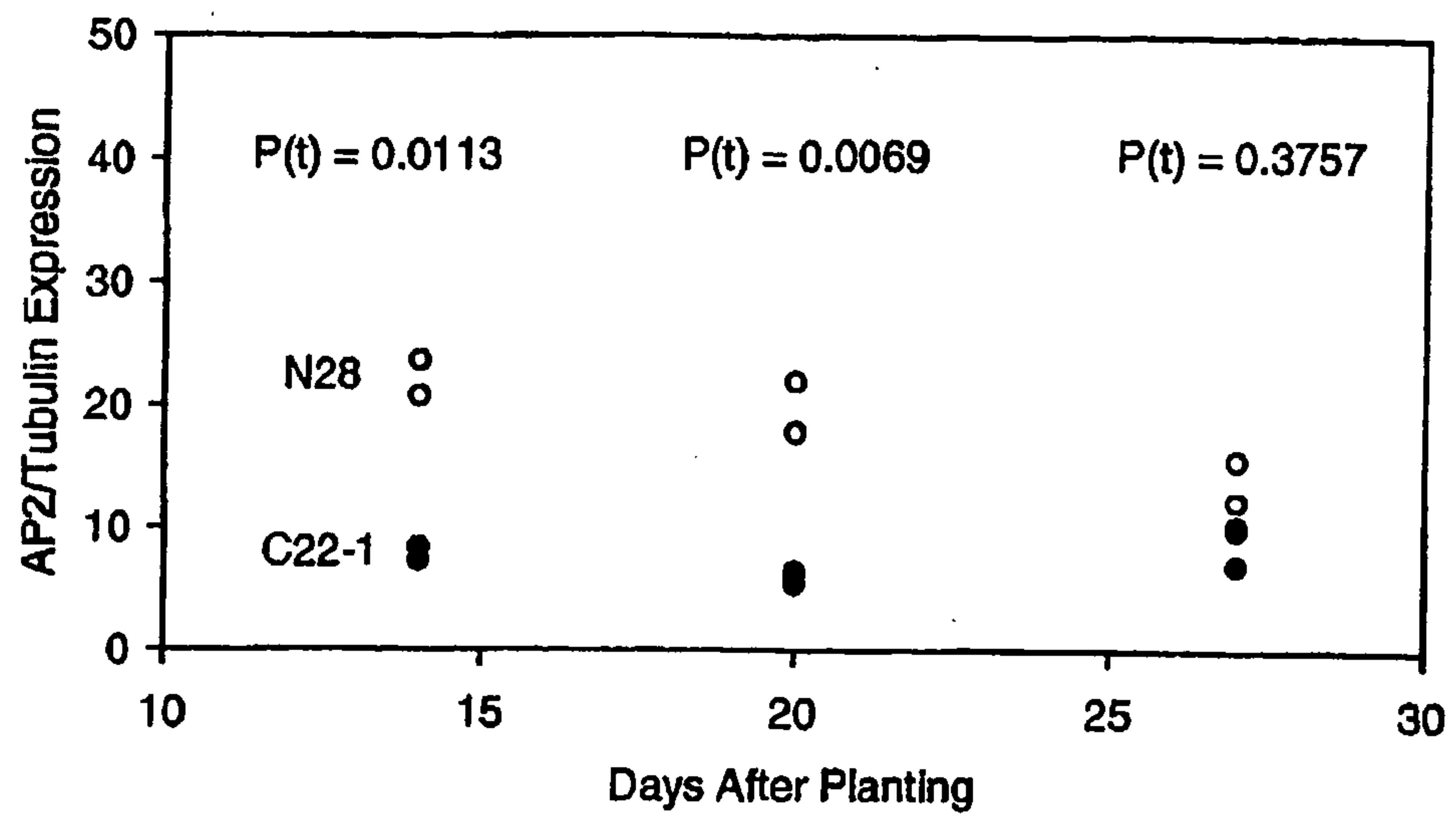
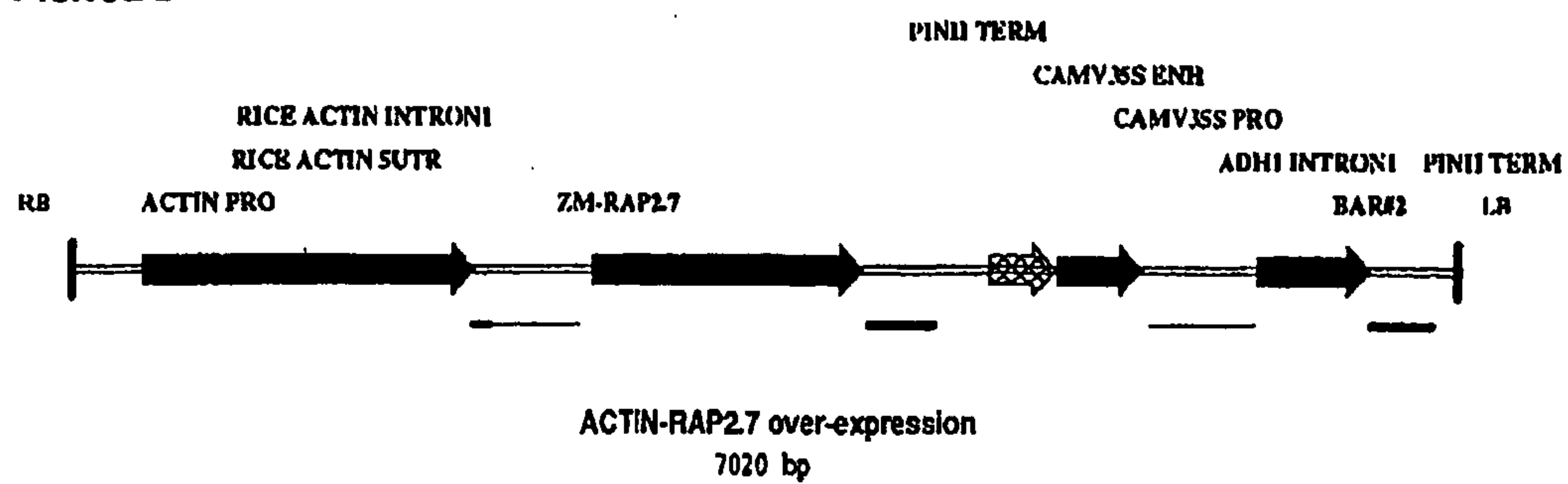


FIGURE 2



3/25

FIGURE 3

MQLDLNVAEAPPPVEMEASDSGSSVLNASEAASAGGAPAPAEEGSSSTPAVLEFSILIRS
DSDAAGADEDEDATPSPPPRHRHQHQQLVTRELF PAGAGPPAPTPRHWAELGFFRADLQ
QQQAPGPRIVPHPHAAPPPAKKSRRGPRSRSSQYRGVTFYRRTGRWESHIWDCGKQVYL
GPDTAHAAARAYDRAAIKFRGVDAIDNFNLSDYEDDMKQMGSLSKKEEFVHVLRRQSTGFS
RGSSRYRGVTLHKCGRWEARMGQFLGKKAYDKAAIKCNGREAVTNFEPSTYHGELPTEVA
DVDLNLISQPSPPQDKNSCLGLQLHHGPFEGSELKKTIDDAPELPGRRRLSPVVAE
HPPAWPAQPPHPPFFVFTNHEMSASGDLHRRPAGAVPSWAWQVAAAAPPPAALPSSAAASS
GFSNTATTAATAAPSASSLRYCPPPPPPSSHHHPR

4/25

FIGURE 4A

```

1                                     caagacttgagctcgaa 17
    |||  ||  ||
601 tagaatatataatctagagcaaactagttagtccaaatatttgtgttggg 650

18 aagtagcacagagaggttcacaactcgaacggagctcaaataactaacaca 67
    || |  ||  |  ||  |  ||  |  ||  |  ||  |  ||
651 aattcaaccaccaaattatttataggaaaagggttaaaccctatttcct 700

68 atcgatcaaattgagaggaggcggagtggtgggagtccttagaatgcttagtg 117
    ||  ||  |  ||  |  ||  |  |||  ||  |  ||
701 ttcactaattaattggaagaacttgaggtgtagtcttcttcgtcgtcgtg 750

118 gatgcttagatgtttcctccatgcgcctagaggtecccttttatagcccca 167
    |  ||  |  |  |  |  |  ||  |  |  ||
751 ccgttaatggggtcctagcacagtacttgctctaccgaggttgggtacca 800

168 agacacctaagagccgttggagatcaaca          tggaatgctatcct 210
    ||  |  |  |  |  |  |  ||  ||  |  |  |
801 aggttcttttgttttgcttttgttagacaccccatgtggggaggggtact 850

211 tgccttctgtcgagtggcgccacgggacaggtcctgtagattgttcgggtgc 260
    || |  ||  |||  |||  |  ||  |  |
851 atgtttatcaaactgtagaaacctaacaggcgactttgacctctggagaa 900

261 gcgatctccttccaaatttggcatatccgaccgttgctcctctgggctga 310
    |  |  |  |  |  |  |  |  |  |  |  |  |  |
901 tctttgtaatgctacatagtgaaccttgttgactcaccataggagtg 950

311 ttggcg          caccggacacagtcgggtgcac 338
    ||  |  |||  ||  ||
951 ttaagggtttgatcgacttatggcaaaaagggggtcacgggtcgtgagta 1000

339 accggacagtcgggtgcaccagct          gaccgttggagcagtcacagtg 384
    |  |  ||  ||  |||  |  ||  ||  |  |||  |
1001 aagtgtgaagacctttgcatagggttagaaaactgatatatcagtcagtgc 1050

385 tcgcgcgaagattgc          gtggccgaccgttgctcaggcgaccgttggct 431
    |  |  |  ||  |  |  |  |  |  |  |  ||  |
1051 acaattaagaacggccttgggagctcctttgattagagatactgtagata 1100

432 caccggacagtcgggtgcaccaccggacagtc          cggatgaattatagtc 478
    ||  |  |||  ||  ||  ||  |||  |||  |||
1101 cattcatgatgatgggttgatgatgggtgcctctaattatgatttctagta 1150

479 gtacgcgcgcgtcgaaacccgagagcggcgagttcacagtggaccagcct 528
    |  |  ||  ||  |  |  ||  |  |  |  |  |
1151 ttttctctacgaggaggtactatttgggataataagctagggttttaagat 1200

529 ggcgcaccggacactgtccgggtgcaccattggacattgtccgggtgcacca 578
    |  |  |  |  |  |  |  ||  ||  ||
1201 aaaatttggcttatattaatgattaaaacctgataaagtaaaagcaacct 1250

579 ccggacagtcctcggtgtgcgaga          ccgagcacaag 611
    |  |||  |  |  |||  ||  ||
1251 gctatcagcttaactccacataaagctagtccatttttagccaaacaagat 1300

```

5/25

Figure 4B

```
612 attggctgcacagagccaagcttttcccttttctcccttcttttttagt 661
    ||| ||| | | | | | | | | | | | |
1301 atttgctgagtacgttgatgtgtgcaaaatggagaacttttatcttaaaa 1350

662 cactgtttctagcacttggataaccatgttagtacataaaacaattcacc 711
    ||| | | | | | | | | | | | | | |
1351 caccagggtgtccacactgcaaccactgctcaagcgaggatgaaggcaac 1400

712 aagcttagaaacataccttttgccttgattttcacttctcactttatttg 761
    | | ||| | | | | | | | | | | |
1401 a tgaagaactttcaggagtttctagacttcaaggagtttaaactaga 1448

762 gcacataagaacttaattaaacgtgttgggca c 794
    | | | | | | | | | |
1449 ttagtggtaaaccacagtcagctgcctgtgaaggccttatctttactacg 1498

795 ttaatcacaaaacattatagaaatggcccaaaggcacatttccctttca 844
    || | | | | | | | | | | | | | |
1499 ttccgcgtagcactttgtttacttgttaagttgatggatacatcatgtt 1548

845 gtaataagttaatacctcttttatatcattatttgaacactgtgcaatgat 894
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1549 gtaataagttaatacctcttttatatcattatttgaacactgtgcaatgat 1598

895 gttcatttatgtaatcggtgtgtacgtcagttctaattctagcacgtaca 944
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1599 gttcatttatgtaatcgctgtgtatgtcagttctaattctagcacataca 1648

945 tggttcacatccaatttgtcttct aaaaacgaatgtgacataatgtcat 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1649 tggttcacatccagtttgtcttctaaaaaacgaatgtgacataatgtcat 1698

994 atgtatgtgataatgctttttgttggggtccttcgtctttcaaaggtcct 1043
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1699 atgtatgtgataatgctttttgttggggtccttcgtctttcaaaggtcct 1748

1044 caaaaacacatttaaccattgggtgttagcacatccttaagtgttgcagg 1093
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1749 caaaaacacatttaaccattgggtgttagcacatccttaagtgttgcagg 1798

1094 agctttgggtattgaataccttcggagcaggacatggaggaagacgaagat 1143
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1799 agctttgggtattgaataccttcggagcaggacatggaggaagacgaagat 1848

1144 gttagcttcgtcataacaacaaggaaacgaaggcagaagtggaacaag 1193
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1849 gttagcttcgtcataacaacaaggaaacgaaggcagaagtggaacaag 1898

1194 gccgggatatggtgttttcaagactctgtatccaaagcaaaaaagacaga 1243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1899 gccgggatatggtgttttcaagactctgtaaccaaaagcaaaaaagacata 1948

1244 aagacgatactgcccttacataaatttgtaaactatgtgaacaagttttat 1293
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1949 aagacgatactgtccttacataaatttgtaaactatatgaacaagttttat 1998

1294 ggacatgtttgttaactttacacgaaattgtaccaccacactataaataga 1343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

6/25

Figure 4C

1999 ggacatgtttgtaactttacacgaaactgtaccaccacactatagataga 2048
1344 taaatagtgccttgcctgagggcgctcttggaacaatgaggaacaactc 1393
|||||
2049 taaatagtgccttgcctgagggcgctcttggaacaatgaggaacaactg 2098
1394 tgtataatcctttttcttctaagtaccttcgggttttctcctcatcaaaa 1443
||| |||||
2099 tgtgtaatcctttttcttctaagtaccttcgggttttctcctcatcaaaa 2148
1444 agccgaagggtactattgtaaattcgtttcatataaagaaagaaatcccaa 1493
||| |||||
2149 agcgaagggtactattgtaaatttggttcatataaagaaagaaatcccaa 2198
1494 gttgtttgagataagtaatcttatctagctttgttatagccatgtgtgta 1543
|||||
2199 gttgtttgagataagtaatcttatctagcttcggttatagccctgtgtgta 2248
1544 atctttatctttatcctctgacaatcctatatattatataataaacctt 1593
|||||
2249 atctttatctttatcctctgacaatcctatatattatataataaacctt 2298
1594 cgtactttacttggatgtcccgaaggacaaactctttaagtacgaaggat 1643
|||||
2299 cgtactttacttggatgtctcgaaggacaaactctttaagtacgaaggat 2348
1644 aacatcttttttaataatgtgttgccttg ttttttattgtgtacaacaa 1692
|||||
2349 aacatcttttttaataatgtgttgccttg ttttttattgtgtacaacaa 2398
1693 ttaaaaacgagtgaccaacattttcatgtcggggtatggggaccattgg 1742
|||||
2399 ttaaaaacgagtgaccaacattttcatgtcggggtatgggagaccattgg 2448
1743 agactcgatatccaaatgaggatgagtatatgatgaatcctatacctatg 1792
|||||
2449 agactc ctaaatgaggatgggtatatgatgaatcctatacctatg 2493
1793 atgagtataagtatgagaatcaggatgagtataacttcatcagaataggt 1842
|||||
2494 atgagtataagtatgagaatcaggatgagtataacttcatcagaaaaggt 2543
1843 gcggggggtccttgtgggggtgcctaccgtgcgatcgcaaaaggcctcc 1892
|||||
2544 acggggggtccttgtgggggtgcctacagtgcgatcgcaaaaggcctcc 2593
1893 aaaaccatagggcccaaaaatttataacaatctttatatacaatataaat 1942
|||||
2594 aaaaccatagagcccaaaaatttataacaatctttatatacaatataaat 2643
1943 aaaaaatattattttatataaaaatatttaccacatatagcatagaatc 1991
|||||
2644 aaaaaatattattttatataaaaatatttaccacatatagcatagaatc 2693
1992 gtaaaagcggttgaaatcgatctgttcttattgttattcaaactatttacc 2041
|||||
2694 gtaaaagcggttgaaatcgatctgttcttattgttattcaaactatttacc 2743
2042 tccagcacattgtagtcattagataaaaaagattgagatcttattgtcac 2091

[illegible]

Figure 4E

2791 cgcgggagcaccgaaagacccccgttctttcctgtaaaaaaaaacccgc 2840
|||||
3486 cgcgggagcaccgaaagacccccgttctttcctgtaaaaaaaaacccgc 3535

2841 cgcctttagctagctaaccggtcgtcctcttcaccccctagctttgctag 2890
|||||
3536 cgcctttagctagctaaccggtcgtcctcttcaccccctagctttgctag 3585

2891 ctctagctaggaacgaaagaaattaaaggataactgagattgctgattgg 2940
|||||
3586 ctctagctaggaacgaaagaaattaaaggataactgagattgctgattgg 3635

2941 tggtcgggtacgggtgttcttgagtcgtgaagcgacagtacagtggctag 2990
|||||
3636 tggtcgggtacgggtgttcttgagtcgtgaagcgacagtacagtggctag 3685

2991 ggtcgtgccgcccctgcagtcctccggggtgctgcaggatggctcgtcag 3040
|||||
3686 ggtcgtgccgcccctgcagtcctccggggtgctgcaggatggctcgtcag 3735

3041 ggatcgggagtgaggaggcatcagctctcgcggtcgtggagcctaaatgt 3090
|||||
3736 ggatcaggagtgaggaggcatcagctctcgcggtcgtggagcctaaatgt 3785

3091 accgcaacaacgactcggcactctcctgcttctacctcttcctcctctgg 3140
|||||
3786 accgcaacaacgactcggcactctcctgcttctacctcttcctcctctgg 3835

3141 ttcttcttcttgaagtagacaccaccagttcgccaggtagttagcagccc 3190
|||||
3836 ttcttcttcttgaagtagacaccaccagttcgccaggtagttagcagccc 3885

3191 agttgcgactggggatcgggtggcgggctgccgcttgcgagttgtaagctt 3240
|||||
3886 agttgcgactggggatcgggtggcgggctgccgcttgcgagttgtaagctt 3935

3241 ggaggggaggggagcaggagcaggagatgcagttggatctgaacgtggcc 3290
|||||
3936 ggaggggaggggagcaggagcaggagatgcagctggatctgaacgtggcc 3985

3291 gaggcgccgccgcccgtggagatggaggcgagcgactcgggggtcgtcgg 3340
|||||
3986 gaggcgccgccgcccgtggagatggaggcgagcgactcgggggtcgtcgg 4035

3341 gctgaacgcgtcggaagcggcgctcggcgggcgggcgcccgccggcg 3390
|||||
4036 gctgaacgcgtcggaagcggcgctcggcgggcgggcgcccgccggcg 4085

3391 aggagggatctagctcaacgccggccgtgctggagttcagcatcctcatc 3440
|||||
4086 aggaggggtccagctcaacgccggccgctggagttcagcatcctcatc 4135

3441 cggagcgatagcgacgcggccggcgcgagcaggacgaggacgccacgcc 3490
|||||
4136 cggagcgacagcgacgcggccggcgcgagcaggacgaggacgccacgcc 4185

3491 atcgctcctcctcgccaccgccaccagcaccagcagcagctcgtgaccc 3540
|||||
4186 gtcgctcctcctcgccaccgccaccagcaccagcagcagctcgtgaccc 4235

9/25

Figure 4F

```
3541 gcgagctgttcccggccggcgccgggtccgcccggccccgacgcgcgcggcat 3590
      |||||||
4236 gcgagctgttcccggctggcgccggccccgcccggcgccgcggcat 4285

3591 tgggcccagagctcggcttcttccgcgcgcgacctgcagcagcaacaggcgcc 3640
      |||||||
4286 tgggcccagagctcggcttcttccgcgcgcgacctgcagcagcaacaggcgcc 4335

3641 gggccccaggatcgtgccgcacccacacgcgcgcgcgcgcgcggccaaga 3690
      |||||||
4336 gggccccaggatcgtgccgcacccacacgcgcgcgcgcgcgcggccaaga 4385

3691 agagccgcgcgcggccccgcgctcccgcagctcgcagtagccgcggcgctcacc 3740
      |||||||
4386 agagccgcgcgcggccccgcgctcccgcagctcgcagtagccgcggcgctcacc 4435

3741 ttctaccgcgcgcacaggccgctgggagtagccacatctggtagtagtagtag 3790
      |||||||
4436 ttctaccgcgcgcacaggccgctgggagtagccacatctggtagtagtagtag 4485

3791 actgtctacaactagccacaccacaccgattgcttccgactctcattaat 3840
      |||||||
4486 actgtctacaactagccacaccacaccgattgcttccgactctcattaat 4535

3841 ttctgacacaaactctccgtcttccctcctcttctcccgcgcgcagggat 3890
      |||||||
4536 ttctgacacaaactctccgtcttccctcctcttctcccgcgcgcagggat 4585

3891 tgcggcaagcaggtgtacctaggtgagcaagagcagatctcttttgcgtt 3940
      |||||||
4586 tgcggcaagcaggtgtacctaggtgagcaagagcagatctcttttgcgtt 4635

3941 cccaaagatttttccccttttagttccttatcccatcccatctcgaatgg 3990
      |||||||
4636 cccaaaga ttttccccttttagctcct catcccatctcgaatgg 4679

3991 cctagctaaccgattcagtggtgggtccggctgctggccgatatacgcagg 4040
      |||||||
4680 cctagctaaccgattcactgggtgggtccggctgctggccgatatacgcagg 4729

4041 tggattcgacaccgctcacgccgctgcaaggcacgcactggactggacgc 4090
      |||||||
4730 tggattcgacaccgctcacgccgctgcaaggcacgcactggactggacgc 4779

4091 ccagaattcttcgtcatgtgagtctctgaccg aattgattga 4132
      |||||||
4780 ccagaattcttcgtcatgtgagtctctgaccggattggttgattgattga 4829

4133 ttaacgagtctctggctcctggaactcgcagggcgtagcaccgggcggcg 4182
      |||||||
4830 ttaacgagtctctggctcctggaactcgcagggcgtagcaccgggcggcg 4879

4183 atcaagttccgcggcgctcgacgccgacatcaacttcaacctcagcgacta 4232
      |||||||
4880 atcaagttccgcggcgctggacgccgacatcaacttcaacctcagcgacta 4929

4233 cgaggacgacatgaagcagatggggagcctgtccaaggaggagttcgtgc 4282
      |||||||
```

10/25

Figure 4G

4930 cgaggacgacatgaagcagatggggagcctgtccaaggaggagttcgtgc 4979
4283 acgtcctgcgccgtcagagcaccggcttctcgagaggcagctccaggtac 4332
|||||
4980 acgtcctgcgccgtcagagcaccggcttctcgagaggcagctccaggtac 5029
4333 agaggcgtcaccctgcacaagtgcggccgctgggaggcgcgcatggggca 4382
|||||
5030 agaggcgtcaccctgcacaagtgcggccgctgggaggcgcgcatggggca 5079
4383 gtccctcggcaagaagtaagaaccaaccaacgcttcttttctttttcttt 4432
|||||
5080 gtccctcggcaagaagtaagaaccaaccaacgcttct ttt 5119
4433 tttttatagcatgcag tgatgattcaaccttagttgtgcctttcctcct 4481
|||||
5120 tttttatagcatgcagatgatgattcacacttagttgtgcctcctcctcct 5169
4482 aatcctatatatgtaggatttagtactgggttgacta tataagta 4525
|||||
5170 aatcc tatgtaggatttagtattgggttgactacatatctattgtta 5215
4526 tatatgtattgttcagtaaaagtatacatagggttagctgcatgtttatgt 4575
|||||
5216 tatatgtattgttcagtaaaagtatacatagggttagctgcatgtttatgt 5265
4576 atgtagctgggtgtttcaatcagaagataaaaaaaagggaagtagtggc 4625
|||||
5266 atgtagctgggtgtttcaatcagaagataaaaaaaagggaagtagtggc 5315
4626 tagggaattcctccaatcctcaccgggtgggaacgccgtgcttgggtgcag 4675
|||||
5316 tagggaattcctccaatcctcaccgggtgggaacgccgtgcttgggtgcag 5365
4676 gtacatatatccttgggctattcgacagcgaagtagaggctgcaaggttct 4725
|||||
5366 gtacatatatccttgggctattcgacagcgaagtagaggctgcaaggttct 5415
4726 tcattcttggattctgccgttcataatgcataatcatgtcttttaatttc 4775
|||||
5416 tcattcttggattctgccgttcataatgcataaaccatgtcttttcatttc 5465
4776 caaagggttgagtaccgactcgattcctcttcgtgtcttttttctttctt 4825
|||||
5466 caaagggttgagtaccgactcgattcctct ttttttttctttctt 5510
4826 tcttcgaaatccagagcctacgacaaggccgccatcaaagcaatggcag 4875
|||||
5511 tcttcgaaatccagagcctacgacaaggccgccatcaaagcaatggcag 5560
4876 agaggccgtgacgaacttcgagccgagcacgtatcacggggagctgccga 4925
|||||
5561 agaggccgtgacgaacttcgagccgagcacgtatcacggggagctgccga 5610
4926 ctgaagggtacgtatttttcttctgcataatat atcttcagggtattatt 4974
|||||
5611 ctgaagggtacgta tttcttctgcataatatataatcttcagggtattatt 5659
4975 ggctattaaactgcttggatcttactgcttcttctgcagttgctgatgtc 5024

11/25

Figure 4H

```
|||||
5660 ggctattaaactgcttggattttactgcttcttctgcagttgctgatgtc 5709
|||||
5025 gatctgaacctgagcatatctcagccgagcccccaaagagacaagaacag 5074
|||||
5710 gatctgaacctgagcatatctcagccgagcccccaaagagacaagaacag 5759
|||||
5075 ctgcctaggtctgcagctccaccacggaccattcgagggctccgaactga 5124
|||||
5760 ctgcctaggtctgcagctccaccacggaccattcgagggctccgaactga 5809
|||||
5125 agaaaaccaaggcaagcgctaacgatagatataccttgacaagctagtat 5174
|||||
5810 agaaaaccaaggcaagcgctaacgatagatataccttgacaagctagtat 5859
|||||
5175 caaacaaaaccagtaaaaaaagtttactttcttgtcgaatttcattgcct 5224
|||||
5860 caaacaaaaccagt attttttttactttcttgtcgaatttcattgcct 5907
|||||
5225 acctgatgtacgtacttgtgcttctgcacaaaataacgaaatccttttgc 5274
|||||
5908 acctgatgtacgtacttgtgcttctgcacagaataacgaaatccttttgc 5957
|||||
5275 cctctgatgatgatgcagatcgacgatgctccctctgagctaccggggccg 5324
|||||
5958 cctctgatgatgatgcagatcgacgatgctccctctgacctcccggggccg 6007
|||||
5325 ccctcgtcagctgtctcctctcgtggctgagcatccgccggcctggcctg 5374
|||||
6008 ccctcgtcggctgtctcctctcgtggctgagcatccgccggcctggcctg 6057
|||||
5375 cgcagccgcctcacccttcttctcgtcttcacaaacctagaggttaggtga 5424
|||||
6058 cgcagccgcctcacccttcttctcgtcttcacaaacctagaggttaggtga 6107
|||||
5425 cagctactgatcgagatgcagcagcagttcaaa 5457
|||||
6108 cagctactgatcgagatgcagcagcagttcaaaacttagatccggtcctt 6157

5458 cctgtctgttccaaggaccttttaggccggattaccaaatacatcggtc 5504
|||||
6308 tgtcctgtctgttccaaggaccttttaggccggattaccaaatacatcggtc 6357
|||||
5505 aactgtcctgtctgttatatatatttatgtgttaatttataatacaagtgtg 5554
|||||
6358 aactgtcctgtctgttatatatatttatgtg tttataatacaagtgtg 6403
|||||
5555 actatttttcaaacccttcttcaaaatgcatgaaaagag tttttttta 5602
|||||
6404 actatttttcaaacccttcttcaaaatgcatgaaaagagttttttttta 6453
|||||
5603 acgaaaggcgaaaagaaaatatgatac ttgggacaggagcaagccttga 5651
|||||
6454 acgaaaggcgaaaagaaaatatgatactttgggacaggagcaagccttga 6503
|||||
5652 tcatacagaaagtattattaattaggatcactgagctgttcattttgttct 5701
```

12/25

Figure 4I

```
|||||
6504 tcatcagaaagtattattaattaggatcactgagctgttcattttgttct 6553

5702 tgagtcaatcctaataatcgactatgtcagtgaatgaacttgtgttgacca 5751
||| |||||
6554 tgaatcaatcctaataatcgactatgtcagtgaatgaacttgtgttgacca 6603

5752 atgcagatgagtgcacaggagatctccacaggaggcctgcaggggctgt 5801
|||||
6604 atgcagatgagtgcacaggagatctccacaggaggcctgcaggggctgt 6653

5802 tcccagctgggcatggcaggtggcagcagcagctcctcctcctgccgccc 5851
|||||
6654 tcccagctgggcatggcaggtggcagcagcagctcctcctcctgccgccc 6703

5852 tgccgtcgtccgctgcagcatcatcaggattctccaacaccgccacgaca 5901
|||||
6704 tgccgtcgtccgctgcagcatcatcaggattctccaacaccgccacgaca 6753

5902 gctgccaccaccgccccatcggcctcctccctccggtactgc ccgcc 5948
||||| |||||
6754 gctgccaccgcccgccccatcggcctcctccctccggtactgccgccgcc 6803

5949 gccgccgccgccgctcgagccatcaccatccccgctgagagaatcaagaag 5998
|||||
6804 gccgccgccgccgctcgagccatcaccatcgccgctgagagaatcaagaag 6853

5999 ccgcactgtaaatctgccgg gaagctageattttccccccggcccc 6044
||||| |||||
6854 ccgcactgtaaatctgccgggaatgaagctagcattttccccccggccc 6902

6045 tccccctctccgggcttgccgactttttcagttttgcgccgccggccggg 6094
|||||
6903 tccccctctccgggcttgccgactttttcagttttgcgccgccagccggg 6951

6095 gtggtggtttcttctgtagccgatcgattggattcctcgtattactgctgct 6144
||||| |||||
6952 gtggtggtttcttctgtaaccgatcggttgattcctcgtatta ctgct 6998

6145 tacactcccaattaagtgaacacacgctcctctactctttacactac 6194
|||||
6999 tacactcccaattaagt ggacacacgctcctctactctttacactag 7047

6195 acacactgttagctgatcgattggacgtacttgctagctgctggtgctgc 6244
|||||
7048 acacactgttagctgatcgattggacgtacttgctagctgctggtgctgc 7097

6245 tgctagcttgagattgactaacttcagcacttggttgatctatatctat 6294
|||||
7098 tgctagctagagattgactaactgaagcacttggttgatctatatctat 7147

6295 atgactatatagacgacacattgtgtacgtgtagataaatatttcttcttt 6344
|||||
7148 atgactatata gacacattgtgtacgtgtagataaatatttcttctatt 7194

6345 tcctgaccgccataaaaactgtttactctggccattttgaactaaaggcta 6394
|||||
7195 tcctgaccgccataaaaactgtttactctagccattttgaactaaaggcta 7244
```

13/25

Figure 4J

```

6395 gctacaaatgagtgtccttctcggccttctacatgttctggtcatggaca 6444
      |||
7245 gctacaaatgagtgtccgtctcggccttctacatgttctggtcatggaca 7294

6445 tcgagagatcaaacttctctgtcctgcttactagatacgtactagattta 6494
      |||
7295 tcgagagatcaaacttctctgtcctgcttactagatacgtactagattta 7344

6495 cttagcctagatagattccgttccaaactcgaggccaggcgcatcgagat 6544
      |||
7345 cttagcctagatagatttcgttccaaactcgaggccaggcgcatcgagat 7394

6545 ccgagaacttcatccactcgtcgtcatcatgctgcatgcatgatggtct 6594
      |||
7395 ccgagcacttcatccactcgtcgtcatcatgctgcatgcatgatggtct 7444

6595 caactctgaggcatgcaaacgcagtgag acgaactgggaggaatttata 6643
      |||
7445 caactctgaggcatgcaaacgcagtgagaacgaactgggaggaatttata 7494

6644 tagagtatatattgtccggcctgttggtgataaagatagaatgcatgcac 6693
      |||
7495 tagagtatatattgtccggcctgttggtgataaagatagaatgcatgcac 7544

6694 gctaactgccaacatgcatgggtgctgcatcgaatttttggtat      g 6738
      |||
7545 gctaactgccaacatgcatgggtgctgcatcgaatttttggtatggtgcg 7594

6739 gtgcatgcataccgtgcattggtgctctgctagtactaggaccaatctcc 6788
      |||
7595 gtgcatgcataccgtgcattggtg      gg 7620

6789 atggctccattagatctcttgtttactcgtctccatgtgcctctcaaagt 6838
      | ||| | || | | | | | | | | | | |
7621 gagaatccatgaaatagcacttggttgagacggcgt      tcacgaaat 7665

6839 gtgtactagctagtgtgcggcacacaagttggcagttgtttgttttcag 6888
      | | | | | | | | | | | | | | | |
7666 agcatccgatttcaagtaattcatagaatagcacttgttttaccaaatta 7715

6889 cggggaagaaggaggtcaccgttgctcatcgtcaggggcgaagctaggatc 6938
      | | | | | | | | | | | | | | | |
7716 attcagaaaataacactctatctatattttgcatctcttttattgcttacc 7765

6939 agaagacagagggggcaggttagcctccaagcgaccaaaccagtaagaa 6988
      | | | | | | | | | | | | | | | |
7766 tcacatacaaattggaccaaattacccctatttatcacaaccttctttttt 7815

6989 cacaatatataaatggcaagagaaccaaccataatatatatattgatata 7038
      | | | | | | | | | | | | | | | |
7816 tatctcttatgtccataagaacaacataaataaaatatatgaccta 7865

7039 taatcttcattataaaaaacagtataatgaaacaacatctattttgtcaaac 7088
      || | | | | | | | | | | | | | | |
7866 tatttttatagtgttaaacatacaaataagacacatgttttatgaacgaag 7915

7089 aaaataaaattaaatctcagttattttgaatttagctctacgtgtattag 7138
      || | | | | | | | | | | | | | | |
7916 aagtctatatattaa      ctattaaattagaactaggtatcttcg 7957

```

14/25

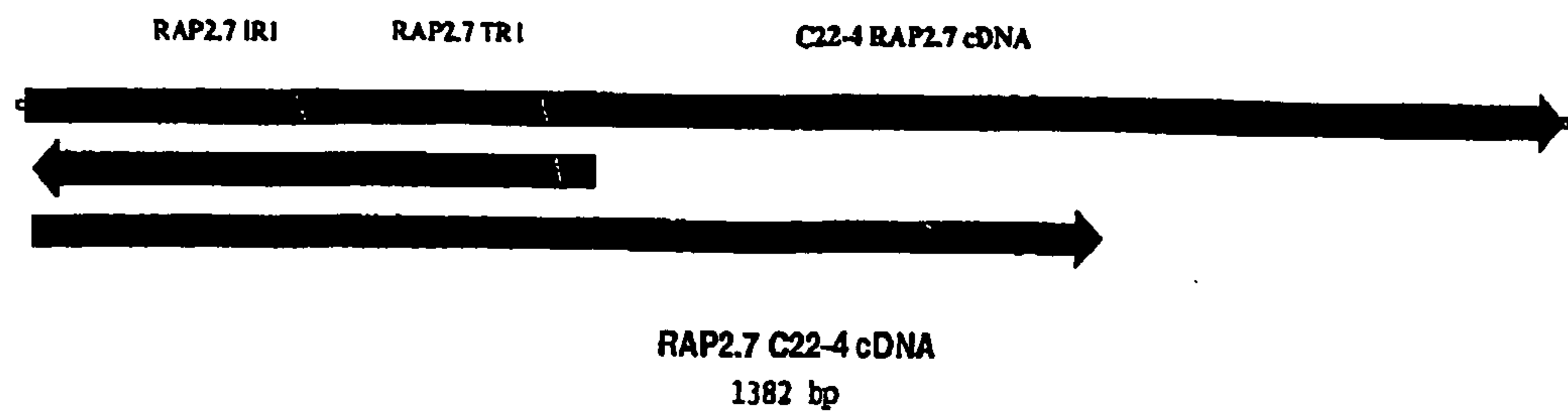
Figure 4K

```
7139 ctagatcataggtgaaagt cgcctagagggggggtgaataggcgaaa 7186
      | |   |   | ||| |   | ||   ||   | || ||| | |
7958 gttgtcgaacttttaaaattagacatatgttgggtccattaaaagggtcttag 8007

7187 ctgaaatttaciaaatataaacacaaactacaagccgggttagcggtataag 7236
      |   | ||| | |   |   | | | | |   | | ||
8008 attgtacttcctatgttcttttttatttgatacggttggttttttttcaa 8056
```

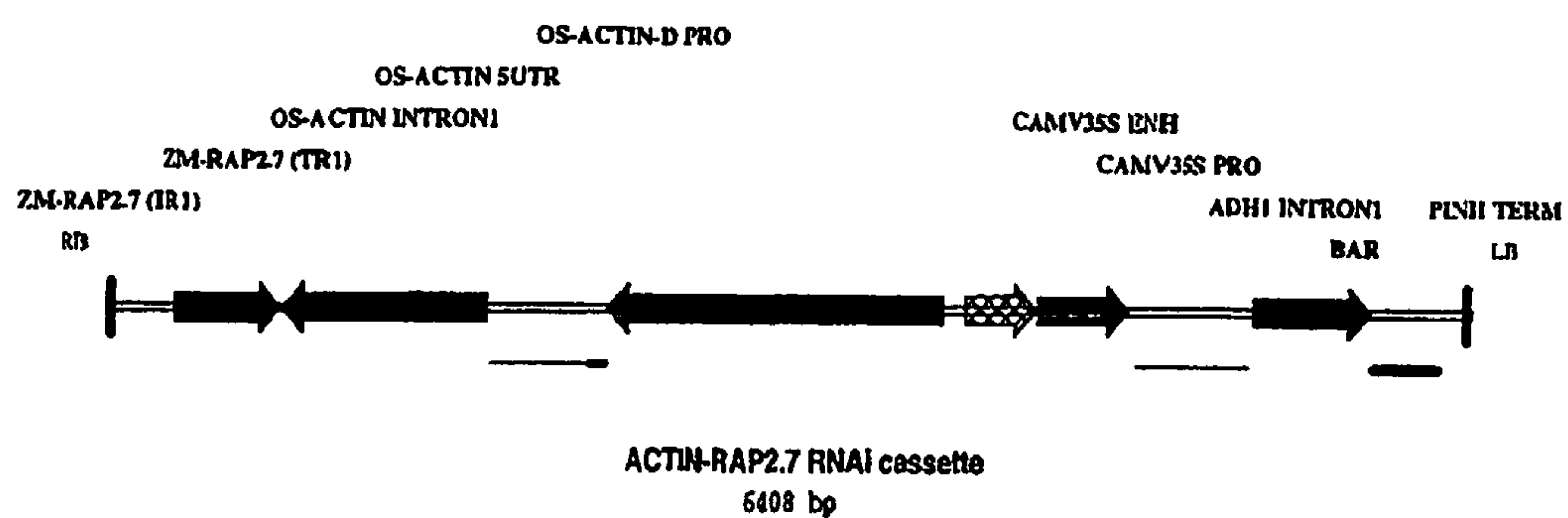
15/25

FIGURE 5



16/25

FIGURE 6



17/25

Figure 7

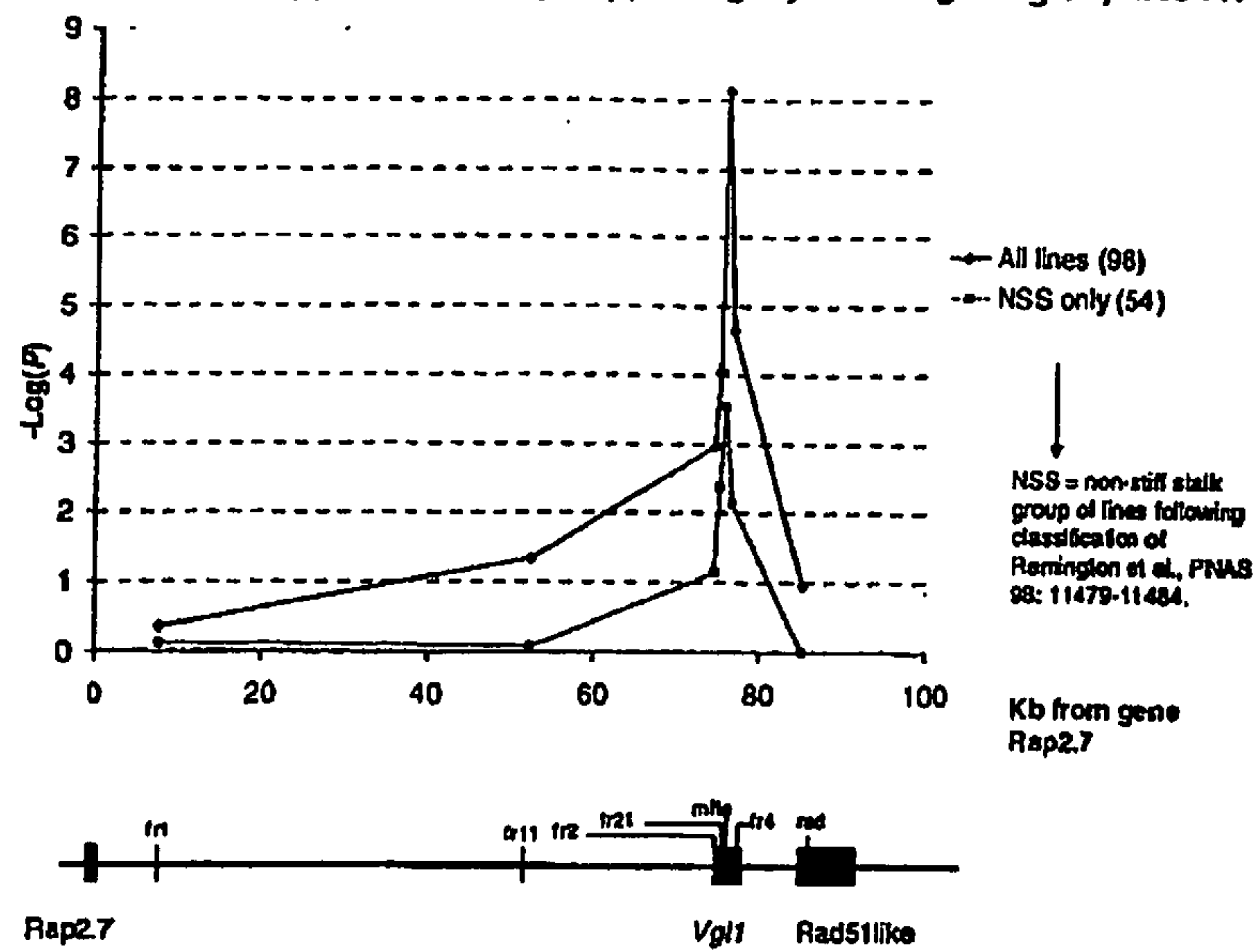
Association mapping for DPS (3 rep, Bologna) at the *Vgt1* region, ANOVA

Figure 8

[illegible]

19/25

FIGURE 9

	1	50
VGT1_C22	(1)	-----
VGT1_M017	(1)	-----
VGT1_B73	(1)	ATAAATTTATTTTGAAGATATAAATATTTTAAACGGTATTTTATTATAATA
Consensus	(1)	-----
	51	100
VGT1_C22	(1)	-----
VGT1_M017	(1)	-----
VGT1_B73	(51)	AAATGATTTATATGGATTATACAGATAGAGTCTCTCTACTCTCTAGTC
Consensus	(51)	-----
	101	150
VGT1_C22	(1)	-----
VGT1_M017	(1)	-----
VGT1_B73	(101)	CTGTGTACGTGCAGGTTTGGCAAGGTATGAACAGGACGTCAAACGGCTAC
Consensus	(101)	-----
	151	200
VGT1_C22	(1)	-----CTCCACCTCTCTCGTACGCTTTGATTGGCCGGCTC
VGT1_M017	(1)	-----CTCCACCTCTCTCGTACGCTTTGATTGGCCGGCTC
VGT1_B73	(151)	GTGTACTCCACACACACTCCACCTCTCTCGTACGCTTTGATTGGCCGGCTC
Consensus	(151)	CTCCACCTCTCTCGTACGCTTTGATTGGCCGGCTC
	201	250
VGT1_C22	(35)	GGAGGCATTATCGGCAAGATCGGGCTTGCCCATCACTGGGATCATCTG
VGT1_M017	(35)	GCACGCATTATCGGCAAGATCGGGCTTGCCCATCACTGGGATCATCTG
VGT1_B73	(201)	GGAGGCATTATCGGCAAGATCGGGCTTGCCCATCACTGGGATCATCTG
Consensus	(201)	GCACGCATTATCGGCAAGATCGGGCTTGCCCATCACTGGGATCATCTG
	251	300
VGT1_C22	(85)	GGTGGCTGGCTGGCCGGCCACATATCAAACATGG-----ATTTCGTA
VGT1_M017	(85)	GGTGGCTGGCTGGCCGGCCACATATCAAACATGG-----ATTTCGTA
VGT1_B73	(251)	GGTGGCTGGCTGGCCGGCCACATATCAAACATGGATTTCGTCATTTCGTA
Consensus	(251)	GCTGGCTGGCTGGCCGGCCACATATCAAACATGGATTTCGTA
	301	350
VGT1_C22	(127)	TATTTTAAATGCGCTGAATATGTTATACTATGTTTATTAGATAACGATCT
VGT1_M017	(127)	TATTTTAAATGCGCTGAATATGTTATACTATGTTTATTAGATAACGATCT
VGT1_B73	(301)	TATTTTAAATGCGCTGAATATGTTATACTATGTTTATTAGATAACGATCT
Consensus	(301)	TATTTTAAATGCGCTGAATATGTTATACTATGTTTATTAGATAACGATCT
	351	400
VGT1_C22	(177)	TTACAGTCGATTATGAATGTGTAGATTAAAGATTGAGCTTCTCGATGAA
VGT1_M017	(177)	TTACAGTCGATTATGAATGTGTAGATTAAAGATTGAGCTTCTCGATGAA
VGT1_B73	(351)	TTACAGTCGATTATGAATGTGTAGATTAAAGATTGAGCTTCTCGATGAA
Consensus	(351)	TTACAGTCGATTATGAATGTGTAGATTAAAGATTGAGCTTCTCGATGAA
	401	450
VGT1_C22	(227)	AAATTCAGGCAAGAGGCTCTTTTGGTTGGAGACCTGGCATGGCACTTGCC
VGT1_M017	(227)	AAATTCAGGCAAGAGGCTCTTTTGGTTGGAGACCTGGCATGGCACTTGCC
VGT1_B73	(401)	AAATTCAGGCAAGAGGCTCTTTTGGTTGGAGACCTGGCATGGCACTTGCC
Consensus	(401)	AAATTCAGGCAAGAGGCTCTTTTGGTTGGAGACCTGGCATGGCACTTGCC
	451	500
VGT1_C22	(277)	TGCGAGTCAAGCAAACCTATGCCACTGTTTGGACCAGGCCACTTTCTTA
VGT1_M017	(277)	TGCGAGTCAAGCAAACCTATGCCACTGTTTGGACCAGGCCACTTTCTTA
VGT1_B73	(451)	TGCGAGTCAAGCAAACCTATGCCACTGTTTGGACCAGGCCACTTTCTTA
Consensus	(451)	TGCGAGTCAAGCAAACCTATGCCACTGTTTGGACCAGGCCACTTTCTTA
	501	550
VGT1_C22	(327)	GAGAGCCGAATCAGACATCACTATCGGAATTGGAATGAGTGCTCAGCACT
VGT1_M017	(327)	GAGAGCCGAATCAGACATCACTATCGGAATTGGAATGAGTGCTCAGCACT
VGT1_B73	(501)	GAGAGCCGAATCAGACATCACTATCGGAATTGGAATGAGTGCTCAGCACT
Consensus	(501)	GAGAGCCGAATCAGACATCACTATCGGAATTGGAATGAGTGCTCAGCACT
	551	600
VGT1_C22	(377)	TTACCGAGTCAATTAATCGGACACTCGGCAAAGCAATCTTTGTGAGTG
VGT1_M017	(377)	TTACCGAGTCAATTAATCGGACACTCGGCAAAGCAATCTTTGTGAGTG
VGT1_B73	(551)	TTACCGAGTCAATTAATCGGACACTCGGCAAAGCAATCTTTGTGAGTG
Consensus	(551)	TTACCGAGTCAATTAATCGGACACTCGGCAAAGCAATCTTTGTGAGTG
	601	650
VGT1_C22	(427)	CCACTCTCGGCGAATTAATAAGGCTCTCGGACAGATCTCGTATGTCGAG
VGT1_M017	(427)	CCACTCTCGGCGAATTAATAAGGCTCTCGGACAGATCTCGTATGTCGAG
VGT1_B73	(601)	TCACTCTCGGCGAATTAATAAGGCTCTCGGACAGATCTCGTATGTCGAG
Consensus	(601)	CCACTCTCGGCGAATTAATAAGGCTCTCGGACAGATCTCGTATGTCGAG
	651	700
VGT1_C22	(477)	AGCGGAACACTCGACATAGAAAACACTCGGCAAAGAGGTTTGTGCGAA
VGT1_M017	(477)	AGCGGAACACTCGACATAGAAAACACTCGGCAAAGAGGTTTGTGCGAA
VGT1_B73	(651)	AGCGGAACACTCGACATAGAAAACACTCGGCAAAGAGGTTTGTGCGAA

Consensus	(651)	AGCGGAACACTCGACATAGAAAAACACTCGGC AAAAGAGGTTTTGCGGAA	750
VGT1_C22	(527)	TGACAAGCTCTCGGCAAAATGTGACACTCGATAAGGTCGTCAAGAGTCG	
VGT1_M017	(527)	TGCCAAGCTCTCGGCAAAATGCGACACTCGGTAAGGGTCGTCAAGAGTCG	
VGT1_B73	(701)	TGCCAAGCTCTCGGCAAAATGCGACACTCGGTAAGGGTCGTCAAGAGTCG	
Consensus	(701)	TGCCAAGCTCTCGGCAAAATGCGACACTCGGTAAGGGTCGTCAAGAGTCG	800
VGT1_C22	(577)	TCTATTGTTTGACGGCCATTAACCTTCCCGAGTGTCAAACGTTGACACTTG	
VGT1_M017	(577)	TCTATTGTTTGACGGCCATTAACCTTCCCGAGTGTCAAACGTTGACACTTG	
VGT1_B73	(751)	TCTATTGTTTGACGGCCATTAACCTTCCCGAGTGTCAAACGTTGACACTTG	
Consensus	(751)	TCTATTGTTTGACGGCCATTAACCTTCCCGAGTGTCAAACGTTGACACTTG	850
VGT1_C22	(627)	GGAAATPATCTTCTTTTTTGTGCGATTGTAACTGGCAAAACCTCGGCAAAAG	
VGT1_M017	(627)	GGAAATPATCTTCTTTTTTGTGCGATTGTAACTGGCAAAACCTCGGCAAAAG	
VGT1_B73	(801)	GGAAATPATCTTCTTTTTTGTGCGATTGTAACTGGCAAAACCTCGGCAAAAG	
Consensus	(801)	GGAAATPATCTTCTTTTTTGTGCGATTGTAACTGGCAAAACCTCGGCAAAAG	900
VGT1_C22	(677)	TATACCTTTGCGGAGTGTCTTCCCTCAACACTCGGATAAAGAATATTTGTTT	
VGT1_M017	(677)	TATACCTTTGCGGAGTGTCTTCCCTCGACACTCGAATAAAGAATATTTGTTT	
VGT1_B73	(851)	TATACCTTTGCGGAGTGTCTTCCCTCAACACTCGGATAAAGAATATTTGTTT	
Consensus	(851)	TATACCTTTGCGGAGTGTCTTCCCTCAACACTCGGATAAAGAATATTTGTTT	950
VGT1_C22	(727)	CTTTTCTTTTTCCTATACCAAAAGCTCTTGTGACGTTTTCCTGAGTAT	
VGT1_M017	(727)	CTTTTCTTTTTCCTATACCAAAAGCTCTTGTGACGTTTTCCTGAGTAT	
VGT1_B73	(901)	CTTTTCTTTTTCCTATACCAAAAGCTCTTGTGACGTTTTCCTGAGTAT	
Consensus	(901)	CTTTTCTTTTTCCTATACCAAAAGCTCTTGTGACGTTTTCCTGAGTAT	1000
VGT1_C22	(777)	ATAGACATACATATTCAAATTTACACAATTTATCAAAGTGTTCGTATAAT	
VGT1_M017	(777)	ATAGACATACATATTCAAATTTACACAATTTATCAAAGTGTTCGTATAAT	
VGT1_B73	(951)	ATAGACATACATATTCAAATTTACACAATTTATCAAAGTGTTCGTATAAT	
Consensus	(951)	ATAGACATACATATTCAAATTTACACAATTTATCAAAGTGTTCGTATAAT	1050
VGT1_C22	(827)	TAATAGATTTAGTTTGTTTAATTGAATTTCTGAGAACCAACCAATAGAC	
VGT1_M017	(827)	TAATAGATTTAGTTTGTTTAATTGAATTTCTGAGAACCAACCAATAGAC	
VGT1_B73	(1001)	TAATAGATTTAGTTTGTTTAATTGAATTTCTGAGAACCAACCAATAGAC	
Consensus	(1001)	TAATAGATTTAGTTTGTTTAATTGAATTTCTGAGAACCAACCAATAGAC	1100
VGT1_C22	(877)	CGGAATTAGACATATCTAGACATTTAAAGTAAGATACTTAAAAATATAT	
VGT1_M017	(877)	CGGAATTAGACATATCTAGACATTTAAAGTAAGATACTTAAAAATATAT	
VGT1_B73	(1051)	CGGAATTAGACATATCTAGACATTTAAAGTAAGATACTTAAAAATATAT	
Consensus	(1051)	CGGAATTAGACATATCTAGACATTTAAAGTAAGATACTTAAAAATATAT	1150
VGT1_C22	(927)	AGTGTTTACCTTCAACCGGTACTAAATATCATTCTCTGTAGTAGACAGGAT	
VGT1_M017	(927)	AGTGTTTACCTTCAACCGGTACTAAATATCATTCTCTGTAGTAGACAGGAT	
VGT1_B73	(1101)	AGTGTTTACCTTCAACCGGTACTAAATATCATTCTCTGTAGTAGACAGGAT	
Consensus	(1101)	AGTGTTTACCTTCAACCGGTACTAAATATCATTCTCTGTAGTAGACAGGAT	1200
VGT1_C22	(977)	AACGAAAGCTAGTCTATTTAGATCATCAGTTCCAGTTTCAGATTTTAAAT	
VGT1_M017	(977)	AACGAAAGCTAGTCTATTTAGATCATCAGTTCCAGTTTCAGATTTTAAAT	
VGT1_B73	(1151)	AACGAAAGCTAGTCTATTTAGATCATCAGTTCCAGTTTCAGATTTTAAAT	
Consensus	(1151)	AACGAAAGCTAGTCTATTTAGATCATCAGTTCCAGTTTCAGATTTTAAAT	1250
VGT1_C22	(1027)	GCTAGTCCCTCCATTTCGAATTTACAATTTATTTAAATTTTTTTAGTGTGAT	
VGT1_M017	(1027)	GCTAGTCCCTCCATTTCGAATTTACAATTTATTTAAATTTTTTTAGTGTGAT	
VGT1_B73	(1201)	GCTAGTCCCTCCATTTCGAATTTACAATTTATTTAAATTTTTTTAGTGTGAT	
Consensus	(1201)	GCTAGTCCCTCCATTTCGAATTTACAATTTATTTAAATTTTTTTAGTGTGAT	1300
VGT1_C22	(1077)	ACCGTTTAGCGTA-----TGAGCTTTGA-TTTTATAT	
VGT1_M017	(1077)	ACCGTTTAGCGTA-----TGAGCTTTGATTTTTTTATAT	
VGT1_B73	(1251)	ACCGTTTAGCATATAATATACTTTAAGTGTAGCTTTGA-TTTTATAT	
Consensus	(1251)	ACCGTTTAGCGTA-----TGAGCTTTGA TTTTTTATAT	1350
VGT1_C22	(1112)	ATTTTTGCAAAATTTGAATAAGACAAGTGTGTCAAATTTGGTGTAAAAA	
VGT1_M017	(1113)	ATTTTTGCAAAATTTGAATAAGACAAGTGTGTCAAATTTGGTGTAAAAA	
VGT1_B73	(1300)	ATTTTTGCAAAATTTGAATAAGACAAGTGTGTCAAATTTGGTGTAAAAA	
Consensus	(1301)	ATTTTTGCAAAATTTGAATAAGACAAGTGTGTCAAATTTGGTGTAAAAA	1400
VGT1_C22	(1162)	TTAAACGAAATTTATAAATTTGAGCGGAGGAGTAGAAATCTACGATTTTTT	
VGT1_M017	(1163)	TTAAACGAAATTTATAAATTTGAGCGGAGGAGTAGAAATCTACGATTTTTT	
VGT1_B73	(1350)	TTAAACGAAATTTATAAATTTGAGCGGAGGAGTAGAAATCTACGATTTTTT	
Consensus	(1351)	TTAAACGAAATTTATAAATTTGAGCGGAGGAGTAGAAATCTACGATTTTTT	1450
VGT1_C22	(1212)	CTAGCTGAGGCAGACTGTGCGCAATCCCATCCATCGAGTCCACTTGCCCT	
VGT1_M017	(1213)	CTAGCTGAGGCAGACTGTGCGCAATCCCATCCATCGAGTCCACTTGCCCT	

21/25

VGT1_B73	(1400)	CTAGCTGAGGCAGACTGTGCGCATTC	1400
Consensus	(1401)	CTAGCTGAGGCAGACTGTGCGCATTC	1500
		1451	1500
VGT1_C22	(1262)	TCTCCTCGACATGAATACGAATGTA	1262
VGT1_M017	(1263)	TCTCCTCGACATGAATACGAATGTA	1263
VGT1_B73	(1450)	TCTCCTCGACATGAATACGAATGTA	1450
Consensus	(1451)	TCTCCTCGACATGAATACGAATGTA	1550
		1501	1550
VGT1_C22	(1312)	AGCAATTAACTGCAAAATCCATCAT	1312
VGT1_M017	(1313)	AGCAATTAACTGCAAAATCCATCAT	1313
VGT1_B73	(1500)	AGCAATTAACTGCAAAATCCATCAT	1500
Consensus	(1501)	AGCAATTAACTGCAAAATCCATCAT	1600
		1551	1600
VGT1_C22	(1362)	GCGTTTAATTAGTTTGTGTAGACGA	1362
VGT1_M017	(1363)	GCGTTTAATTAGTTTGTGTAGACGA	1363
VGT1_B73	(1550)	GCGTTTAATTAGTTTGTGTAGACGA	1550
Consensus	(1551)	GCGTTTAATTAGTTTGTGTAGACGA	1650
		1601	1650
VGT1_C22	(1412)	TGTCCTAAATTAATTAATTAATTA	1412
VGT1_M017	(1413)	TGTCCTAAATTAATTAATTAATTA	1413
VGT1_B73	(1600)	TGTCCTAAATTAATTAATTAATTA	1600
Consensus	(1601)	TGTCCTAAATTAATTAATTAATTA	1700
		1651	1700
VGT1_C22	(1462)	GTAACATCGCTCTTAAATTTGTG	1462
VGT1_M017	(1463)	GTAACATCGCTCTTAAATTTGTG	1463
VGT1_B73	(1650)	GTAACATCGCTCTTAAATTTGTG	1650
Consensus	(1651)	GTAACATCGCTCTTAAATTTGTG	1750
		1701	1750
VGT1_C22	(1512)	CAAGACTATATTGTTAAAGCAAAT	1512
VGT1_M017	(1513)	CAAGACTATATTGTTAAAGCAAAT	1513
VGT1_B73	(1700)	CAAGACTATATTGTTAAAGCAAAT	1700
Consensus	(1701)	CAAGACTATATTGTTAAAGCAAAT	1800
		1751	1800
VGT1_C22	(1562)	GACCACCTAACTAGTGGGAATGCC	1562
VGT1_M017	(1563)	GACCACCTAACTAGTGGGAATGCC	1563
VGT1_B73	(1750)	GACCACCTAACTAGTGGGAATGCC	1750
Consensus	(1751)	GACCACCTAACTAGTGGGAATGCC	1850
		1801	1850
VGT1_C22	(1612)	TTTGGCAACATCATGGAAGAAATA	1612
VGT1_M017	(1613)	TTTGGCAACATCATGGAAGAAATA	1613
VGT1_B73	(1800)	TTTGGCAACATCATGGAAGAAATA	1800
Consensus	(1801)	TTTGGCAACATCATGGAAGAAATA	1900
		1851	1900
VGT1_C22	(1662)	AGAGAGATGTCAGTGGTCTTGG	1662
VGT1_M017	(1663)	AGAGAGATGTCAGTGGTCTTGG	1663
VGT1_B73	(1850)	AGAGAGATGTCAGTGGTCTTGG	1850
Consensus	(1851)	AGAGAGATGTCAGTGGTCTTGG	1950
		1901	1950
VGT1_C22	(1712)	GAGTTTGGACCTTTGAATCTGGT	1712
VGT1_M017	(1713)	GAGTTTGGACCTTTGAATCTGGT	1713
VGT1_B73	(1900)	GAGTTTGGACCTTTGAATCTGGT	1900
Consensus	(1901)	GAGTTTGGACCTTTGAATCTGGT	2000
		1951	2000
VGT1_C22	(1762)	TGACACATGGCATCATTTGTGAAG	1762
VGT1_M017	(1763)	TGACACATGGCATCATTTGTGAAG	1763
VGT1_B73	(1950)	TGACACATGGCATCATTTGTGAAG	1950
Consensus	(1951)	TGACACATGGCATCATTTGTGAAG	2050
		2001	2050
VGT1_C22	(1812)	GCAAAGTAGCGAAGGTACACTGTT	1812
VGT1_M017	(1813)	GCAAAGTAGCGAAGGTACACTGTT	1813
VGT1_B73	(2000)	GCAAAGTAGCGAAGGTACACTGTT	2000
Consensus	(2001)	GCAAAGTAGCGAAGGTACACTGTT	2100
		2051	2100
VGT1_C22	(1862)	ATGCATGTAAGACATGCATTGTAC	1862
VGT1_M017	(1863)	ATGCATGTAAGACATGCATTGTAC	1863
VGT1_B73	(2050)	ATGCATGTAAGACATGCATTGTAC	2050
Consensus	(2051)	ATGCATGTAAGACATGCATTGTAC	2150
		2101	2150
VGT1_C22	(1912)	TCTATGTAGATCACTCCTACGACC	1912
VGT1_M017	(1913)	TCTATGTAGATCACTCCTACGACC	1913
VGT1_B73	(2100)	TCTATGTAGATCACTCCTACGACC	2100
Consensus	(2101)	TCTATGTAGATCACTCCTACGACC	2200
		2151	2200
VGT1_C22	(1960)	GATGCATCGATCGAATCGATGTA	1960

VGT1_M017	(1961)	GATGCATGGATGGAATGGAATGTAAGTGGAAGTGAGAGCACTATCGGA	2201	2250
VGT1_B73	(2150)	GATACATGGATGGAATGGAATGTAAGTGGAAGTGAGAGCACTATCGGA		
Consensus	(2151)	GATGCATGGATGGAATGGAATGTAAGTGGAAGTGAGAGCACTATCGGA		
VGT1_C22	(2010)	ATCAGTTTCTTCGCGGAGTGTCGCTTTCCATGAATAAACATGTGAACCTA		
VGT1_M017	(2011)	ATCAGTTTCTTCGCGGAGTGTCGCTTTCCATGAATAAACATGTGAACCTA		
VGT1_B73	(2200)	ATCAGTTTCTTCGCGGAGTGTCGCTTTCCATGAATAAACATGTGAACCTA		
Consensus	(2201)	ATCAGTTTCTTCGCGGAGTGTCGCTTTCCATGAATAAACATGTGAACCTA	2251	2300
VGT1_C22	(2060)	GGCCTAGGCACCTTCCACTTCAGGGCTTGTTCGGTTAGCTCTCAATCCAT		
VGT1_M017	(2061)	GGCCTAGGCACCTTCCACTTCAGGGCTTGTTCGGTTAGCTCTCAATCCAT		
VGT1_B73	(2250)	GGCCTAGGCACCTTCCACTTCAGG-----		
Consensus	(2251)	GGCCTAGGCACCTTCCACTTCAGGGCTTGTTCGGTTAGCTCTCAATCCAT	2301	2350
VGT1_C22	(2110)	GTGGATTGAGCGGGATTGGATGGGTTTGAATCCCAAACAAGTCAAACCTC		
VGT1_M017	(2111)	GTGGATTGAGCGGGATTGGATGGGTTTGAATCCCAAACAAGTCAAACCTC		
VGT1_B73	(2274)	-----		
Consensus	(2301)	GTGGATTGAGCGGGATTGGATGGGTTTGAATCCCAAACAAGTCAAACCTC	2351	2400
VGT1_C22	(2160)	TTTCAATTTTTTCCAAATCCCATCCAAATCCATGTGTATTGGGAATAACCG		
VGT1_M017	(2161)	TTTCAATTTTTTCCAAATCCCATCCAAATCCATGTGTATTGGGAATAACCG		
VGT1_B73	(2274)	-----		
Consensus	(2351)	TTTCAATTTTTTCCAAATCCCATCCAAATCCATGTGTATTGGGAATAACCG	2401	2450
VGT1_C22	(2210)	AACAAGCCCTCAGGCATCATCTTCTACAAATGTAGAAATGTGTGAAGGTAG		
VGT1_M017	(2211)	AACAAGCCCTCAGG---CATCTTCTACAAATGTAGAAATGTGTGAAGGTAG		
VGT1_B73	(2274)	-----CATCTTCTACAAATGTAGAAATGTGTGAAGGTAG		
Consensus	(2401)	AACAAGCCCTCAGG CATCTTCTACAAATGTAGAAATGTGTGAAGGTAG	2451	2500
VGT1_C22	(2260)	GCAAACGTAAC-GGAAAGGCAGGAAGCTCATCGCCAAAGCATCTCCTCTC		
VGT1_M017	(2258)	GCAAACGTAAC-GGAAAGGCAGGAAGCTCATCGCCAAAGCATCTCCTCTC		
VGT1_B73	(2307)	GCAAACGTAACGGGAAGGCAGGAAGCTCATCGCCAAAGCATCTCCTCTC		
Consensus	(2451)	GCAAACGTAAC GGAAAGGCAGGAAGCTCATCGCCAAAGCATCTCCTCTC	2501	2550
VGT1_C22	(2309)	CTGCTCCTTTTGACGACCTTTCATACCTGCAACCCGCTTTTCTCGAAAGG		
VGT1_M017	(2307)	CTGCTCCTTTTGACGACCTTTCATACCTGCAACCCGCTTTTCTCGAAAGG		
VGT1_B73	(2357)	CTGCTCCTTTTGACGACCTTTCATACCTGCAACCCGCTTTTCTCGAAAGG		
Consensus	(2501)	CTGCTCCTTTTGACGACCTTTCATACCTGCAACCCGCTTTTCTCGAAAGG	2551	2600
VGT1_C22	(2359)	GCATCAAGATT---TATATATATATATATATATATATATATATATATATAT		
VGT1_M017	(2357)	GCATCAAGATTTATATATATATATATATATATATATATATATATATATATAT		
VGT1_B73	(2407)	GCATCAAGATTTATATATATATATATATATATATATATATATATATATATAT		
Consensus	(2551)	GCATCAAGATTTATATATATATATATATATATATATATATATATATATATAT	2601	2650
VGT1_C22	(2405)	TATTATTAGCTTTGTTTACAACAAGTATATTATTATTATTATCCGGGCTGTG		
VGT1_M017	(2407)	TATTATTAGCTTTGTTTACAACAAGTATATTATTATTATTATTATCCGGGCTGTG		
VGT1_B73	(2441)	TATTATTAGCTTTGTTTACAACAAGTATATTATTATTATTATTATCCGGGCTGTG		
Consensus	(2601)	TATTATTAGCTTTGTTTACAACAAGTATATTATTATTATTATTATCCGGGCTGTG	2651	2700
VGT1_C22	(2455)	AGCAGAGGAGCTGGTACAACTTGTCCCTGAGACGGCCAGAGGCCAACA		
VGT1_M017	(2457)	AGCAGAGGAGCTGGTACAACTTGTCCCTGAGACGGCCAGAGGCCAACA		
VGT1_B73	(2491)	AGCAGAGGAGCTGGTACAACTTGTCCCTGAGACGGCCAGAGGCCAACA		
Consensus	(2651)	AGCAGAGGAGCTGGTACAACTTGTCCCTGAGACGGCCAGAGGCCAACA	2701	2750
VGT1_C22	(2505)	GTGTTGTGGGCTTGGCGGCATGACGCGACGTTGCTAGCTGCCGTGTCCAA		
VGT1_M017	(2507)	GTGTTGTGGGCTTGGCGGCATGACGCGACGTTGCTAGCTGCCGTGTCCAA		
VGT1_B73	(2541)	GTGTTGTGGGCTTGGCGGCATGA--CGACGTTGCTAGCTGCCGTGTCCAA		
Consensus	(2701)	GTGTTGTGGGCTTGGCGGCATGACGCGACGTTGCTAGCTGCCGTGTCCAA	2751	2800
VGT1_C22	(2555)	CAGGAAGAGAAGACGGCGACGTGGTGGCACTGTACGTTTTCGCCCAACA		
VGT1_M017	(2557)	CAGGAAGAGAAGACGGCGACGTGGTGGCACTGTACGTTTTCGCCCAACA		
VGT1_B73	(2589)	CAGGAAGAGAAGACGGCGACGTGGTGGCACTGTACGTTTTCGCCCAACA		
Consensus	(2751)	CAGGAAGAGAAGACGGCGACGTGGTGGCACTGTACGTTTTCGCCCAACA	2801	2850
VGT1_C22	(2605)	CAAAACGGGCGGGGGGGCGGTGGTATACGTTATGGTGGCCACTGGCCAGC		
VGT1_M017	(2607)	CAAAACGGGCGGGGGGGCGGTGGTATACGTTATGGTGGCCACTGGCCAGC		
VGT1_B73	(2639)	CAAAACGGGCGGGGGGGCGGTGGTATACGTTATGGTGGCCACTGGCCAGC		
Consensus	(2801)	CAAAACGGGCGGGGGGGCGGTGGTATACGTTATGGTGGCCACTGGCCAGC	2851	2900
VGT1_C22	(2655)	CGCCGTGCCGTGTCAGGACAGCAGCCCAACAGGACCAACGCCGCCCAATG		
VGT1_M017	(2657)	CGCCGTGCCGTGTCAGGACAGCAGCCCAACAGGACCAACGCCGCCCAATG		
VGT1_B73	(2689)	CGCCGTGCCGTGTCAGGACAGCAGCCCAACAGGACCAACGCCGCCCAATG		
Consensus	(2851)	CGCCGTGCCGTGTCAGGACAGCAGCCCAACAGGACCAACGCCGCCCAATG	2901	2950

VGT1_C22	(2705)	GATCGGACGGCCCTCTGCTACTGCTAGAAATGGAAAGCACGCAGGTACGTTG	2951	3000
VGT1_M017	(2707)	GATCGGACGGCCCTCTGCTACTGCTAGAAATGGAAAGCACGCAGGTACGTTG		
VGT1_B73	(2739)	GATCGGACGGCCCTCTGCTACTGCTAGAAATGGAAAGCACGCAGGTACGTTG		
Consensus	(2901)	GATCGGACGGCCCTCTGCTACTGCTAGAAATGGAAAGCACGCAGGTACGTTG		
VGT1_C22	(2755)	GGGCCCCCTCCCTTTTCCCGCGCAAGTGCAGTGCCAGTGCCGGCAGTGCGTG		
VGT1_M017	(2757)	GGGCCCCCTCCCTTTTCCCGCGCAAGTGCAGTGCCAGTGCCGGCAGTGCGTG		
VGT1_B73	(2789)	GGGCCCCCTCCCTTTTCCCGCGCAAGTGCAGTGCCAGTGCCGGCAGTGCGTG		
Consensus	(2951)	GGGCCCCCTCCCTTTTCCCGCGCAAGTGCAGTGCCAGTGCCGGCAGTGCGTG		
VGT1_C22	(2805)	TGTCATTATTCTGTGTCGGACCGGTAGGTAGTAGTATCAGATGTACTACCA		
VGT1_M017	(2807)	TGTCATTATTCTGTGTCGGACCGGTAGGTAGTAGTATCAGATGTACTACCA		
VGT1_B73	(2839)	TGTCATTATTCTGTGTCGGACCGGTAGGTAGTAGTATCAGATGTACTACCA		
Consensus	(3001)	TGTCATTATTCTGTGTCGGACCGGTAGGTAGTAGTATCAGATGTACTACCA		
VGT1_C22	(2855)	GTCAAACGACAGTGCTTCCCGCGCGGCCAAAGGTACAGTGACACTTTTGC		
VGT1_M017	(2857)	GTCAAACGACAGTGCTTCCCGCGCGGCCAAAGGTACAGTGACACTTTTGC		
VGT1_B73	(2889)	GTCAAACGACAGTGCTTCCCGCGCGGCCAAAGGTACAGTGACACTTTTGC		
Consensus	(3051)	GTCAAACGACAGTGCTTCCCGCGCGGCCAAAGGTACAGTGACACTTTTGC		
VGT1_C22	(2905)	CAAAACCAAAAAAACAACAGCAATAAAGAAAGGAAACCGCGCGGGAATA		
VGT1_M017	(2907)	CAAAACCAAAAAAACAACAGCAATAAAGAAAGGAAACCGCGCGGGAATA		
VGT1_B73	(2939)	CAAAACCAAAAAAACAACAGCAATAAAGAAAGGAAACCGCGCGGGAATA		
Consensus	(3101)	CAAAACCAAAAAAACAACAGCAATAAAGAAAGGAAACCGCGCGGGAATA		
VGT1_C22	(2955)	TATCGATCTCATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTCTTCTACAGAGAT		
VGT1_M017	(2957)	TATCGATCTCATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTCTTCTACAGAGAT		
VGT1_B73	(2989)	TATCGATCTCATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTCTTCTACAGAGAT		
Consensus	(3151)	TATCGATCTCATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTCTTCTACAGAGAT		
VGT1_C22	(3005)	GCTAAGG-AATAAATAAATAAAGGTCCTAATAAAGACCGGATTCTTTAT		
VGT1_M017	(3007)	GCTAAGG-AATAAATAAATAAAGGTCCTAATAAAGACCGGATTCTTTAT		
VGT1_B73	(3039)	GCTAAGG-AATAAATAAATAAAGGTCCTAATAAAGACCGGATTCTTTAT		
Consensus	(3201)	GCTAAGG AATAAATAAATAAAGGTCCTAATAAAGACCGGATTCTTTAT		
VGT1_C22	(3054)	TTCTTTCCAAATCCAGAAAGGAATTATCTTCCCGGAATCTATTTTTGA		
VGT1_M017	(3056)	TTCTTTCCAAATCCAGAAAGGAATTATCTTCCCGGAATCTATTTTTGA		
VGT1_B73	(3089)	TTCTTTCCAAATCCAGAAAGGAATTATCTTCCCGGAATCTATTTTTGA		
Consensus	(3251)	TTCTTTCCAAATCCAGAAAGGAATTATCTTCCCGGAATCTATTTTTGA		
VGT1_C22	(3104)	GCA---AATAAATAAATAAATATATGATTGTTTATTATTTCATTGGTTCT		
VGT1_M017	(3106)	GCAAAATAAATAAATAAATATATGATTGTTTATTATTTCATTGGTTCT		
VGT1_B73	(3139)	GCA---AATAAATAAATAAATATATGATTGTTTATTATTTCATTGGTTCT		
Consensus	(3301)	GCA AATAAATAAATAAATATATGATTGTTTATTATTTCATTGGTTCT		
VGT1_C22	(3151)	CTGGTTAATCATTTTGGACGTCATCTACTAATAACATAGGTCCTCCACT		
VGT1_M017	(3156)	CTGGTTAATCATTTTGGACGTCATCTACTAATAACATAGGTCCTCCACT		
VGT1_B73	(3186)	CTGGTTAATCATTTTGGACGTCATCTACTAATAACATAGGTCCTCCACT		
Consensus	(3351)	CTGGTTAATCATTTTGGACGTCATCTACTAATAACATAGGTCCTCCACT		
VGT1_C22	(3201)	ATGTGGAGCGCACGGCCCTTACCTTAAGACACAAATTGTTGACTTCCAGGA		
VGT1_M017	(3206)	ATGTGGAGCGCACGGCCCTTACCTTAAGACACAAATTGTTGACTTCCAGGA		
VGT1_B73	(3236)	ATGTGGAGCGCACGGCCCTTACCTTAAGACACAAATTGTTGACTTCCAGGA		
Consensus	(3401)	ATGTGGAGCGCACGGCCCTTACCTTAAGACACAAATTGTTGACTTCCAGGA		
VGT1_C22	(3251)	TTATATAATCCACCTTATAGATTATATAATCATATAATGATATCTAGTTA		
VGT1_M017	(3256)	TTATATAATCCACCTTATAGATTATATAATCATATAATGATATCTAGTTA		
VGT1_B73	(3286)	TTATATAATCCACCTTATAGATTATATAATCATATAATGATATCTAGTTA		
Consensus	(3451)	TTATATAATCCACCTTATAGATTATATAATCATATAATGATATCTAGTTA		
VGT1_C22	(3301)	TCAAGATTATATAAATAATCCACCTAATAAATTGTTGTTGTTGTTTGGCTC		
VGT1_M017	(3306)	TCAAGATTATATAAATAATCCACCTAATAAATTGTTGTTGTTGTTTGGCTC		
VGT1_B73	(3336)	TCAAGATTATATAAATAATCCACCTAATAAATTGTTGTTGTTGTTTGGCTC		
Consensus	(3501)	TCAAGATTATATAAATAATCCACCTAATAAATTGTTGTTGTTGTTTGGCTC		
VGT1_C22	(3351)	TTGATATAGTAGGACTATGTAGCCTACTGACATGATCAATTTACTTCTCT		
VGT1_M017	(3356)	TTGATATAGTAGGACTATGTAGCCTACTGACATGATCAATTTACTTCTCT		
VGT1_B73	(3386)	TTGATATAGTAGGACTATGTAGCCTACTGACATGATCAATTTACTTCTCT		
Consensus	(3551)	TTGATATAGTAGGACTATGTAGCCTACTGACATGATCAATTTACTTCTCT		
VGT1_C22	(3401)	AATCACCXXGCAAAATGAAAAATCCATTGTTTACCTTGCCCTTGAGTTGTAT		
VGT1_M017	(3406)	AATCACCXXGCAAAATGAAAAATCCATTGTTTACCTTGCCCTTGAGTTGTAT		
VGT1_B73	(3436)	AATCACCXXGCAAAATGAAAAATCCATTGTTTACCTTGCCCTTGAGTTGTAT		
Consensus	(3601)	AATCACCXXGCAAAATGAAAAATCCATTGTTTACCTTGCCCTTGAGTTGTAT		

24/25

		3651		3700
VGT1_C22	(3451)	TAACGGTTTCCACCAAAGTCGTTCAAGACTCTAAGTAGCTCACATGTATT		
VGT1_M017	(3456)	TAACGGTTTCCACCAAAGTCGTTCAAGACTCTAAGTAGCTCACATGTATT		
VGT1_B73	(3486)	TAACGGTTTCCACCAAAGTCGTTCAAGACTCTAAGTAGCTCACATGTATT		
Consensus	(3651)	TAACGGTTTCCACCAAAGTCGTTCAAGACTCTAAGTAGCTCACATGTATT		
		3701		3750
VGT1_C22	(3501)	CCATCCAGTCTTCAAATACTTTAAGGGGTCTTCACTTTGGATGTAAATAC		
VGT1_M017	(3506)	CCATCCAGTCTTCAAATACTTTAAGGGGTCTTCACTTTGGATGTAAATAC		
VGT1_B73	(3536)	CCATCCAGTCTTCAAATACTTTAAGGGGTCTTCACTTTGGATGTAAATAC		
Consensus	(3701)	CCATCCAGTCTTCAAATACTTTAAGGGGTCTTCACTTTGGATGTAAATAC		
		3751		3800
VGT1_C22	(3551)	TTTATGTTTAAGCATTGAAGCTCATTTTAGACTAGGATACCCCTTGATCT		
VGT1_M017	(3556)	TTTATGTTTAAGCATTGAAGCTCATTTTAGACTAGGATACCCCTTGATCT		
VGT1_B73	(3586)	TTTATGTTTAAGCATTGAAGCTCATTTTAGACTAGGATACCCCTTGATCT		
Consensus	(3751)	TTTATGTTTAAGCATTGAAGCTCATTTTAGACTAGGATACCCCTTGATCT		
		3801		3850
VGT1_C22	(3601)	AGGTGGTCCACATATTCTCCTAGGCATGACAAATTACTCAGAGACACCGT		
VGT1_M017	(3606)	AGGTGGTCCACATATTCTCCTAGGCATGACAAATTACTCAGAGACACCGT		
VGT1_B73	(3636)	AGGTGGTCCACATATTCTCCTAGGCATGACAAATTACTCAGAGACACCGT		
Consensus	(3801)	AGGTGGTCCACATATTCTCCTAGGCATGACAAATTACTCAGAGACACCGT		
		3851		3900
VGT1_C22	(3651)	TCCTCTTCCCACTACATGTGTGCATGCATTGAAACCGTGAAAAACCTAC		
VGT1_M017	(3656)	TCCTCTTCCCACTACATGTGTGCATGCATTGAAACCGTGAAAAACCTAC		
VGT1_B73	(3686)	TCCTCTTCCCACTACATGTGTGCATGCATTGAAACCGTGAAAAACCTAC		
Consensus	(3851)	TCCTCTTCCCACTACATGTGTGCATGCATTGAAACCGTGAAAAACCTAC		
		3901		3950
VGT1_C22	(3701)	GAGTGCAATCGATCAGGTGGGCAAAAAAGGCACACCTAGAGGCTAGAGC		
VGT1_M017	(3706)	GAGTGCAATCGATCAGGTGGGCAAAAAAGGCACACCTAGAGGCTAGAGC		
VGT1_B73	(3736)	GAGTGCAATCGATCAGGTGGGCAAAAAAGGCACACCTAGAGGCTAGAGC		
Consensus	(3901)	GAGTGCAATCGATCAGGTGGGCAAAAAAGGCACACCTAGAGGCTAGAGC		
		3951		4000
VGT1_C22	(3751)	TCAGTCAGTCCCGAAGGATTGC-AAAAAAGAGCACTAGAGCTCGACC		
VGT1_M017	(3756)	TCAGTCAGTCCCGAAGGATTGC-AAAAAAGAGCACTAGAGCTCGACC		
VGT1_B73	(3786)	TCAGTCAGTCCCGAAGGATTGC-AAAAAAGAGCACTAGAGCTCGACC		
Consensus	(3951)	TCAGTCAGTCCCGAAGGATTGC-AAAAAAGAGCACTAGAGCTCGACC		
		4001		4050
VGT1_C22	(3800)	TCAAAACCCAGTGTGTCAATTTTACAGCCAGCGTTGCAACCATCTAAGC		
VGT1_M017	(3806)	TCAAAACCCAGTGTGTCAATTTTACAGCCAGCGTTGCAACCATCTAAGC		
VGT1_B73	(3836)	TCAAAACCCAGTGTGTCAATTTTACAGCCAGCGTTGCAACCATCTAAGC		
Consensus	(4001)	TCAAAACCCAGTGTGTCAATTTTACAGCCAGCGTTGCAACCATCTAAGC		
		4051		4100
VGT1_C22	(3850)	TAACATCTGTATTAAATATTATAGAAACAACAACCTATTAAATAGTT		
VGT1_M017	(3856)	TAACATCTGTATTAAATATTATAGAAACAACAACCTATTAAATAGTT		
VGT1_B73	(3886)	TAACATCTGTATTAAATATTATAGAAACAACAACCTATTAAATAGTT		
Consensus	(4051)	TAACATCTGTATTAAATATTATAGAAACAACAACCTATTAAATAGTT		
		4101		4150
VGT1_C22	(3900)	CCTTGGTATATATACAAGATCATGTCCAAAACATCAAGCATTTGAAGACT		
VGT1_M017	(3906)	CCTTGGTATATATACAAGATCATGTCCAAAACATCAAGCATTTGAAGACT		
VGT1_B73	(3936)	CCTTGGTATATATACAAGATCATGTCCAAAACATCAAGCATTTGAAGACT		
Consensus	(4101)	CCTTGGTATATATACAAGATCATGTCCAAAACATCAAGCATTTGAAGACT		
		4151		4200
VGT1_C22	(3950)	AGATGTAGTACATGGACCAACCTGCAACCTTACTTTATCAGAGAAATCGT		
VGT1_M017	(3956)	AGATGTAGTACATGGACCAACCTGCAACCTTACTTTATCAGAGAAATCGT		
VGT1_B73	(3986)	AGATGTAGTACATGGACCAACCTGCAACCTTACTTTATCAGAGAAATCGT		
Consensus	(4151)	AGATGTAGTACATGGACCAACCTGCAACCTTACTTTATCAGAGAAATCGT		
		4201		4250
VGT1_C22	(4000)	CAGTCCACACATATTTGTCAATTAATTACCAAGGCACCTACTGAGAGGTTAG		
VGT1_M017	(4006)	CAGTCCACACATATTTGTCAATTAATTACCAAGGCACCTACTGAGAGGTTAG		
VGT1_B73	(4036)	CAGTCCACACATATTTGTCAATTAATTACCAAGGCACCTACTGAGAGGTTAG		
Consensus	(4201)	CAGTCCACACATATTTGTCAATTAATTACCAAGGCACCTACTGAGAGGTTAG		
		4251		4300
VGT1_C22	(4050)	ATGCCCCACCGTTTCCACCAATGATTGGTTAAAAAGCACGGACTCCTGT		
VGT1_M017	(4056)	ATGCCCCACCGTTTCCACCAATGATTGGTTAAAAAGCACGGACTCCTGT		
VGT1_B73	(4086)	ATGCCCCACCGTTTCCACCAATGATTGGTTAAAAAGCACGGACTCCTGT		
Consensus	(4251)	ATGCCCCACCGTTTCCACCAATGATTGGTTAAAAAGCACGGACTCCTGT		
		4301		4350
VGT1_C22	(4100)	GGAGCTGCTACTCTGCAGAGGAGCGGAGTCAAGAACCATTTTAAAGAGAG		
VGT1_M017	(4106)	GGAGCTGCTACTCTGCAGAGGAGCGGAGTCAAGAACCATTTTAAAGAGAG		
VGT1_B73	(4136)	GGAGCTGCTACTCTGCAGAGGAGCGGAGTCAAGAACCATTTTAAAGAGAG		
Consensus	(4301)	GGAGCTGCTACTCTGCAGAGGAGCGGAGTCAAGAACCATTTTAAAGAGAG		
		4351		4400
VGT1_C22	(4150)	CCAAAGTCTTACCAACATACTCTTACTCTCTCTTTTAAAGGTACTACC		
VGT1_M017	(4156)	CCAAAGTCTTACCAACATACTCTTACTCTCTCTTTTAAAGGTACTACC		
VGT1_B73	(4186)	CCAAAGTCTTACCAACATACTCTTACTCTCTCTTTTAAAGGTACTACC		

25/25

```
Consensus (4351) CCAAAGTCCTACCAACATACTCTTACTCTCTCTTTTAAAGGTACTACC
4401 4450
VGT1_C22 (4200) TCTTTTGAATTTAAATGTATTACTCCCTCTT-----
VGT1_M017 (4206) TCTTTTGAATTTAAATGTATTACTCCCTCTT-----
VGT1_B73 (4236) TCTTTTGAATTTAAATGTATTACTCCCTCTTTAAAGAGATAATAGAGAAG
Consensus (4401) TCTTTTGAATTTAAATGTATTACTCCCTCTT
4451 4479
VGT1_C22 (4231) -----
VGT1_M017 (4237) -----
VGT1_B73 (4286) ACGAACAAGAAAGGGAGAGGGAGAAACGA
Consensus (4451)
```

