Title: METHODS AND COMPOSITIONS TO ALTER FLOWERING TIME IN SORGHUM

Abstract: Compositions and methods useful in identifying and selecting sorghum plants with different flowering times are provided herein. The methods and compositions provided herein can be used to select sorghum plants with modulated photoperiod sensitivity. The methods use molecular genetic markers to identify and/or select sorghum plants with different flowering times and/or to identify and select sorghum plants with different photoperiod sensitivities.
METHODS AND COMPOSITIONS TO ALTER FLOWERING TIME IN SORGHUM

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 62/160,110, filed May 12, 2015, the entire contents of which are herein incorporated by reference.

FIELD

This field is related to plant breeding and methods of identifying and selecting sorghum plants.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named 20160428_RTS20253PCT_SequenceListing_ST25.txt created on April 28, 2016 and having a size of 18 kilobytes and is filed concurrently with the specification. The sequence listing contained in this ASCII formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND

Sorghum is the world’s fifth most important cereal crop, after wheat, rice, maize, and barley. Besides being an increasingly important food crop for the world population, sorghum is also an important animal feed used in many countries. Moreover, sorghum is a biofuel crop of increasing importance and is currently the second source of grain-based ethanol in US after maize.


To maximize yield, it is important to tailor the plants’ life cycle to the environments in which they are grown. The transition from vegetative to reproductive growth is a critical developmental switch and a key adaptive trait that
ensures plants set their flowers at an optimum time for pollination, seed
development, and dispersal. For example, temperate environments with long
growing season allow cereal crops to exploit an extended vegetative period for
resource storage. Conversely, early flowering has evolved as an adaptation to short
growing seasons.

The transition from vegetative to reproductive growth is the result of
activation of genes responsible for inflorescence and floral organ formation. The
mechanisms by which environmental factors activate inflorescence and floral organ
production are complex and many genes are known to be involved in the
transduction of environmental signals that regulate flowering. Some of these genes
have been identified in sorghum including PRR37, also referred to as Ma1 (Murphy
et al. 2011. *PNAS*. 108(39):16469-16474; WO2012158630); PHYB, also referred to
as Ma3 (Childs, K.L. 1997. *Plant Physiology* 113(2):611-619; GHD7, also referred to
as Ma6 (Murphy, R.L. et al. *The Plant Genome*. 7(2); US2010024065); Ma5
(US2010024065); and CONSTANS (Yang, S. et al. 2014. *BMC Plant Biology*
14:148; WO201455477).

Selection through the use of molecular markers associated with traits such as
flowering time or photoperiod sensitivity allows selections based solely on the
genetic composition of the progeny. As a result, plant breeding can occur more
rapidly, thereby generating sorghum plants with desirable flowering times and/or
photoperiod insensitivity.

**SUMMARY**

Compositions and methods for identifying and/or selecting sorghum plants
having favorable alleles associated with flowering time are provided herein.

In one embodiment, a method of identifying a sorghum plant having a
favorable allele of a QTL associated with flowering time is presented herein, the
method comprising the steps of: (a) screening a population with at least one marker
to determine if one or more sorghum plants from the population comprises a
favorable allele at the QTL associated with flowering time, wherein the marker is
located within an interval on chromosome 10 comprising and flanked by V02323-001-l001 and V01796-1-l001; and (b) selecting from said population at least one
sorghum plant comprising the favorable allele. The method may further include
selecting a sorghum plant from a breeding program if the QTL allele is detected or counter-selecting a sorghum plant if the QTL allele is not detected.

In one aspect, the favorable QTL allele may be associated with early flowering time and/or photoperiod insensitivity. The QTL allele associated with early flowering time and/or photoperiod insensitivity may comprise at least one, at least two, at least three, at least four, or at least five of the following: a “T” at marker V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001. The QTL allele may comprise: a “T” at marker V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001.

In another aspect, the favorable QTL allele may be associated with late flowering time. The QTL allele associated with late flowering time may comprise at least one, at least two, at least three, at least four or at least five of the following: a “C” at marker V02324-001-I001, a “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001. The QTL allele associated with late flowering time may comprise: a “C” at marker V02324-001-I001, a “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001.

In another embodiment, a method of identifying and/or selecting a sorghum plant that displays late flowering time is provided herein. The method comprises the steps of (a) detecting in a sorghum plant an allele of a marker locus, wherein said marker locus is located within a chromosomal interval comprising and flanked by V02323-001-I001 and V01796-1-I001 and said allele is associated with a haplotype comprising: a “C” at V02324-001-I001, an “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001; and (b) selecting a sorghum plant that has the allele of the marker locus that is associated with the haplotype.

In another embodiment, a method of identifying and/or selecting a sorghum plant that displays late flowering time is provided in which the method comprises detecting in a sorghum plant, at least one, at least two, at least three, at least four, or at least five of the following: a “C” at V02324-001-I001, a “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-001-I001;
and selecting a sorghum plant that has the at least one, at least two, at least three, at least four, or at least five of the following: a “C” at V02324-001-I001, a “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-001-I001.

5 In another embodiment, a method of identifying and/or selecting a sorghum plant that displays early flowering time and/or photoperiod insensitivity is provided herein. The method comprises the steps of (a) detecting in a sorghum plant an allele of a marker locus, wherein said marker locus is located within a chromosomal interval comprising and flanked by V02323-001-I001 and V01796-1-I001 and said allele is associated with a haplotype comprising: a “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001; and (b) selecting a sorghum plant that has the allele of the marker locus that is associated with the haplotype.

10 In another embodiment, a method of identifying and/or selecting a sorghum plant that displays early flowering time and/or photoperiod insensitivity is provided in which the method comprises detecting in a sorghum plant that at least one, at least two, at least three, at least four, or at least five of the following: “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001, and selecting a sorghum plant that has the at least one, at least two, at least three, at least four, or at least five of the following: “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001.

15 In another embodiment, a method of selecting sorghum plants that display late flowering time is provided in which a first sorghum plant is obtained that comprises within its genome: a “C” at V02324-001-I001, an “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001; the sorghum plant is crossed to another sorghum plant; the progeny are evaluated for the presence of a “C” at V02324-001-I001, an “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001, or an allele of a marker locus located within a chromosomal interval defined by and including V02323-001-I001 and V01796-1-I001 wherein said allele is associated with a haplotype comprising a “C” at V02324-001-I001, an “A” at V02131-001-I001, a “G”
at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001; and progeny plants that have the marker allele(s) are selected.

In another embodiment, a method of selecting sorghum plants that display early flowering time and/or photoperiod insensitivity is provided in which a first sorghum plant is obtained that comprises within its genome: a “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001; the sorghum plant is crossed to another sorghum plant; the progeny are evaluated for the presence of a “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001, or an allele of a marker locus located within a chromosomal interval defined by and including V02323-001-001 and V01796-1-I001 wherein said allele is associated with a haplotype comprising a “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001; and progeny plants that have the marker allele(s) are selected.

Sorghum plants with early or late flowering times generated by the methods disclosed herein are also provided.

**BRIEF DESCRIPTION OF THE SEQUENCE LISTING**

The invention can be more fully understood from the following detailed description and the accompanying Sequence Listing which form a part of this application. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821 1.825. The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC IUBMB standards described in *Nucleic Acids Res.* 13:3021 3030 (1985) and in the *Biochemical J.* 219 (2):345 373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NO:1 is the reference sequence for marker V02323-001-I001.

SEQ ID NO:2 is the reference sequence for marker V02324-001-I001.

SEQ ID NO:3 is the reference sequence for marker V02131-001-I001.

SEQ ID NO:4 is the reference sequence for marker V02132-001-I001.
SEQ ID NO:5 is the reference sequence for marker V02330-001-I001.
SEQ ID NO:6 is the reference sequence for marker V01757-1-I001.
SEQ ID NO:7 is the reference sequence for marker V02332-001-I001.
SEQ ID NO:8 is the reference sequence for marker V01796-1-I001.

DETAILED DESCRIPTION

Breeding for flowering time and maturity in sorghum plants can result in plants with varying flowering times and maturity. The current disclosure provides for marker assisted selection of sorghum plants with early or late flowering time, and/or identification of sorghum plants with photoperiod insensitivity. The present disclosure also provides sorghum marker loci that demonstrate statistically significant co-segregation with early or late flowering and/or photoperiod insensitivity in a recombinant inbred population. Detection of these loci or additional linked loci can be used in marker assisted sorghum breeding programs to produce sorghum plants with a variety of flowering times and maturity periods.

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular embodiments, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, terms in the singular and the singular forms “a”, “an” and “the”, for example, include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to “plant”, “the plant” or “a plant” also includes a plurality of plants; also, depending on the context, use of the term “plant” can also include genetically similar or identical progeny of that plant; use of the term “a nucleic acid” optionally includes, as a practical matter, many copies of that nucleic acid molecule; similarly, the term “probe” optionally (and typically) encompasses many similar or identical probe molecules.

Unless otherwise indicated, nucleic acids are written left to right in 5’ to 3’ orientation. Numeric ranges recited within the specification are inclusive of the numbers defining the range and include each integer or any non-integer fraction within the defined range. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although any methods and materials
similar or equivalent to those described herein can be used in the practice for testing of the present invention, the preferred materials and methods are described herein. In describing and claiming the present invention, the following terminology will be used in accordance with the definitions set out below.

The term "allele" refers to one of two or more different nucleotide sequences that occur at a specific locus.

"Allele frequency" refers to the frequency (proportion or percentage) at which an allele is present at a locus within an individual, within a line, or within a population of lines. For example, for an allele "A", diploid individuals of genotype "AA", "Aa", or "aa" have allele frequencies of 1.0, 0.5, or 0.0, respectively. One can estimate the allele frequency within a line by averaging the allele frequencies of a sample of individuals from that line. Similarly, one can calculate the allele frequency within a population of lines by averaging the allele frequencies of lines that make up the population. For a population with a finite number of individuals or lines, an allele frequency can be expressed as a count of individuals or lines (or any other specified grouping) containing the allele.

An "amplicon" is an amplified nucleic acid, e.g., a nucleic acid that is produced by amplifying a template nucleic acid by any available amplification method (e.g., PCR, LCR, transcription, or the like).

The term "amplifying" in the context of nucleic acid amplification is any process whereby additional copies of a selected nucleic acid (or a transcribed form thereof) are produced. Typical amplification methods include various polymerase based replication methods, including the polymerase chain reaction (PCR), ligase mediated methods such as the ligase chain reaction (LCR) and RNA polymerase based amplification (e.g., by transcription) methods.

An allele is "associated with" a trait when it is part of or linked to a DNA sequence or allele that affects the expression of the trait. The presence of the allele is an indicator of how the trait will be expressed.

"Backcrossing" refers to the process whereby hybrid progeny are repeatedly crossed back to one of the parents. In a backcrossing scheme, the "donor" parent refers to the parental plant with the desired gene/genes, locus/loci, or specific phenotype to be introgressed. The "recipient" parent (used one or more times) or
"recurrent" parent (used two or more times) refers to the parental plant into which the gene or locus is being introgressed. For example, see Ragot, M. et al. (1995) Marker-assisted backcrossing: a practical example, in Techniques et Utilisations des Marqueurs Moléculaires Les Colloques, Vol. 72, pp. 45-56, and Openshaw et al., (1994) Marker-assisted Selection in Backcross Breeding, Analysis of Molecular Marker Data, pp. 41-43. The initial cross gives rise to the F1 generation; the term “BC1” then refers to the second use of the recurrent parent, “BC2” refers to the third use of the recurrent parent, and so on.

A centimorgan (“cM”) is a unit of measure of recombination frequency. One cM is equal to a 1% chance that a marker at one genetic locus will be separated from a marker at a second locus due to crossing over in a single generation.

As used herein, the term “chromosomal interval” designates a contiguous linear span of genomic DNA that resides in planta on a single chromosome. The genetic elements or genes located on a single chromosomal interval are physically linked. The size of a chromosomal interval is not particularly limited. In some aspects, the genetic elements located within a single chromosomal interval are genetically linked, typically with a genetic recombination distance of, for example, less than or equal to 20 cM, or alternatively, less than or equal to 10 cM. That is, two genetic elements within a single chromosomal interval undergo recombination at a frequency of less than or equal to 20% or 10%.

A “chromosome” is a single piece of coiled DNA containing many genes that act and move as a unity during cell division and therefore can be said to be linked. It can also be referred to as a “linkage group”.

The phrase “closely linked”, in the present application, means that recombination between two linked loci occurs with a frequency of equal to or less than about 10% (i.e., are separated on a genetic map by not more than 10 cM). Put another way, the closely linked loci co-segregate at least 90% of the time. Marker loci are especially useful in the present invention when they demonstrate a significant probability of co-segregation (linkage) with a desired trait. Closely linked loci such as a marker locus and a second locus can display an inter-locus recombination frequency of 10% or less, preferably about 9% or less, still more preferably about 8% or less, yet more preferably about 7% or less, still more
preferably about 6% or less, yet more preferably about 5% or less, still more preferably about 4% or less, yet more preferably about 3% or less, and still more preferably about 2% or less. In highly preferred embodiments, the relevant loci display a recombination frequency of about 1% or less, e.g., about 0.75% or less, more preferably about 0.5% or less, or yet more preferably about 0.25% or less. Two loci that are localized to the same chromosome, and at such a distance that recombination between the two loci occurs at a frequency of less than 10% (e.g., about 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25%, or less) are also said to be “proximal to” each other. In some cases, two different markers can have the same genetic map coordinates. In that case, the two markers are in such close proximity to each other that recombination occurs between them with such low frequency that it is undetectable.

The term “complement” refers to a nucleotide sequence that is complementary to a given nucleotide sequence, i.e. the sequences are related by the Watson-Crick base-pairing rules.

When referring to the relationship between two genetic elements, such as a genetic element contributing to a trait of interest and a proximal marker, “coupling” phase linkage indicates the state where the “favorable” allele at the genetic element contributing to flowering time and/or photoperiod sensitivity is physically associated on the same chromosome strand as the “favorable” allele of the respective linked marker locus. In coupling phase, both favorable alleles are inherited together by progeny that inherit that chromosome strand.

The term “crossed” or “cross” refers to a sexual cross and involved the fusion of two haploid gametes via pollination to produce diploid progeny (e.g., cells, seeds or plants). The term encompasses both the pollination of one plant by another and selfing (or self-pollination, e.g., when the pollen and ovule are from the same plant).

A plant referred to herein as “diploid” has two sets (genomes) of chromosomes.

A plant referred to herein as a “doubled haploid” is developed by doubling the haploid set of chromosomes (i.e., half the normal number of chromosomes). A doubled haploid plant has two identical sets of chromosomes, and all loci are considered homozygous.
A “favorable allele” is the allele at a particular locus that confers, or contributes to, an agronomically desirable phenotype, e.g., early of late flowering time in sorghum plant, and that allows the identification of plants with that agronomically desirable phenotype. A favorable allele of a marker is a marker allele that segregates with the favorable phenotype.

“Fragment” is intended to mean a portion of a nucleotide sequence. Fragments can be used as hybridization probes or PCR primers using methods disclosed herein.

A “genetic map” is a description of genetic linkage relationships among loci on one or more chromosomes (or linkage groups) within a given species, generally depicted in a diagrammatic or tabular form. For each genetic map, distances between loci are measured by how frequently their alleles appear together in a population (their recombination frequencies). Alleles can be detected using DNA or protein markers, or observable phenotypes. A genetic map is a product of the mapping population, types of markers used, and the polymorphic potential of each marker between different populations. Genetic distances between loci can differ from one genetic map to another. However, information can be correlated from one map to another using common markers. One of ordinary skill in the art can use common marker positions to identify positions of markers and other loci of interest on each individual genetic map. The order of loci should not change between maps, although frequently there are small changes in marker orders due to e.g. markers detecting alternate duplicate loci in different populations, differences in statistical approaches used to order the markers, novel mutation or laboratory error.

A “genetic map location” is a location on a genetic map relative to surrounding genetic markers on the same linkage group where a specified marker can be found within a given species.

“Genetic mapping” is the process of defining the linkage relationships of loci through the use of genetic markers, populations segregating for the markers, and standard genetic principles of recombination frequency.

“Genetic markers” are nucleic acids that are polymorphic in a population and where the alleles of which can be detected and distinguished by one or more analytic methods, e.g., RFLP, AFLP, isozyme, SNP, SSR, and the like. The term
also refers to nucleic acid sequences complementary to the genomic sequences, such as nucleic acids used as probes. Markers corresponding to genetic polymorphisms between members of a population can be detected by methods well-established in the art. These include, e.g., PCR-based sequence specific amplification methods, detection of restriction fragment length polymorphisms (RFLP), detection of isozyme markers, detection of polynucleotide polymorphisms by allele specific hybridization (ASH), detection of amplified variable sequences of the plant genome, detection of self-sustained sequence replication, detection of simple sequence repeats (SSRs), detection of single nucleotide polymorphisms (SNPs), or detection of amplified fragment length polymorphisms (AFLPs). Well established methods are also known for the detection of expressed sequence tags (ESTs) and SSR markers derived from EST sequences and randomly amplified polymorphic DNA (RAPD).

"Genetic recombination frequency" is the frequency of a crossing over event (recombination) between two genetic loci. Recombination frequency can be observed by following the segregation of markers and/or traits following meiosis.

"Genome" refers to the total DNA, or the entire set of genes, carried by a chromosome or chromosome set.

The term "genotype" is the genetic constitution of an individual (or group of individuals) at one or more genetic loci. Genotype is defined by the allele(s) of one or more known loci that the individual has inherited from its parents. The term genotype can be used to refer to an individual's genetic constitution at a single locus, at multiple loci, or, more generally, the term genotype can be used to refer to an individual's genetic make-up for all the genes in its genome.

"Germplasm" refers to genetic material of or from an individual (e.g., a plant), a group of individuals (e.g., a plant line, variety or family), or a clone derived from a line, variety, species, or culture, or more generally, all individuals within a species or for several species. The germplasm can be part of an organism or cell, or can be separate from the organism or cell. In general, germplasm provides genetic material with a specific molecular makeup that provides a physical foundation for some or all of the hereditary qualities of an organism or cell culture. As used herein, germplasm
includes cells, seed or tissues from which new plants may be grown, or plant parts, such as leafs, stems, pollen, or cells, which can be cultured into a whole plant.

A plant referred to as “haploid” has a single set (genome) of chromosomes.

A “haplotype” is the genotype of an individual at a plurality of genetic loci, i.e. a combination of alleles. Typically, the genetic loci described by a haplotype are physically and genetically linked, i.e., on the same chromosome segment. The term “haplotype” can refer to alleles at a particular locus, or to alleles at multiple loci along a chromosomal segment.

The term “heterogeneity” is used to indicate that individuals within the group differ in genotype at one or more specific loci.

An individual is “heterozygous” if more than one allele type is present at a given locus (e.g., a diploid individual with one copy each of two different alleles).

The term “homogeneity” indicates that members of a group have the same genotype at one or more specific loci.

An individual is “homozygous” if the individual has only one type of allele at a given locus (e.g., a diploid individual has a copy of the same allele at a locus for each of two homologous chromosomes).

The term “hybrid” refers to the progeny obtained between the crossing of at least two genetically dissimilar parents.

“Hybridization” or “nucleic acid hybridization” refers to the pairing of complementary RNA and DNA strands as well as the pairing of complementary DNA single strands.

The term “hybridize” means to form base pairs between complementary regions of nucleic acid strands.

The term “inbred” refers to a line that has been bred for genetic homogeneity.

The term "indel" refers to an insertion or deletion, wherein one line may be referred to as having an inserted nucleotide or piece of DNA relative to a second line, or the second line may be referred to as having a deleted nucleotide or piece of DNA relative to the first line.

The term “introgression” refers to the transmission of a desired allele of a genetic locus from one genetic background to another. For example, introgression of a desired allele at a specified locus can be transmitted to at least one progeny via
a sexual cross between two parents of the same species, where at least one of the parents has the desired allele in its genome. Alternatively, for example, transmission of an allele can occur by recombination between two donor genomes, e.g., in a fused protoplast, where at least one of the donor protoplasts has the desired allele in its genome. The desired allele can be, e.g., detected by a marker that is associated with a phenotype, at a QTL, a transgene, or the like. In any case, offspring comprising the desired allele can be repeatedly backcrossed to a line having a desired genetic background and selected for the desired allele, to result in the allele becoming fixed in a selected genetic background.

The process of “introgressing” is often referred to as “backcrossing” when the process is repeated two or more times.

A “line” or “strain” is a group of individuals of identical parentage that are generally inbred to some degree and that are generally homozygous and homogeneous at most loci (isogenic or near isogenic). A “subline” refers to an inbred subset of descendents that are genetically distinct from other similarly inbred subsets descended from the same progenitor.

As used herein, the term “linkage” is used to describe the degree with which one marker locus is associated with another marker locus or some other locus. The linkage relationship between a molecular marker and a locus affecting a phenotype is given as a “probability” or “adjusted probability”. Linkage can be expressed as a desired limit or range. For example, in some embodiments, any marker is linked (genetically and physically) to any other marker when the markers are separated by less than 50, 40, 30, 25, 20, or 15 map units (or cM) of a single meiosis map (a genetic map based on a population that has undergone one round of meiosis, such as e.g. an F₂). In some aspects, it is advantageous to define a bracketed range of linkage, for example, between 10 and 20 cM, between 10 and 30 cM, or between 10 and 40 cM. The more closely a marker is linked to a second locus, the better an indicator for the second locus that marker becomes. Thus, “closely linked loci” such as a marker locus and a second locus display an inter-locus recombination frequency of 10% or less, preferably about 9% or less, still more preferably about 8% or less, yet more preferably about 7% or less, still more preferably about 6% or less, yet more preferably about 5% or less, still more preferably about 4% or less,
yet more preferably about 3% or less, and still more preferably about 2% or less. In highly preferred embodiments, the relevant loci display a recombination frequency of about 1% or less, e.g., about 0.75% or less, more preferably about 0.5% or less, or yet more preferably about 0.25% or less. Two loci that are localized to the same chromosome, and at such a distance that recombination between the two loci occurs at a frequency of less than 10% (e.g., about 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25%, or less) are also said to be “in proximity to” each other. Since one cM is the distance between two markers that show a 1% recombination frequency, any marker is closely linked (genetically and physically) to any other marker that is in close proximity, e.g., at or less than 10 cM distant. Two closely linked markers on the same chromosome can be positioned 9, 8, 7, 6, 5, 4, 3, 2, 1, 0.75, 0.5 or 0.25 cM or less from each other.

The term “linkage disequilibrium” refers to a non-random segregation of genetic loci or traits (or both). In either case, linkage disequilibrium implies that the relevant loci are within sufficient physical proximity along a length of a chromosome so that they segregate together with greater than random (i.e., non-random) frequency. Markers that show linkage disequilibrium are considered linked. Linked loci co-segregate more than 50% of the time, e.g., from about 51% to about 100% of the time. In other words, two markers that co-segregate have a recombination frequency of less than 50% (and by definition, are separated by less than 50 cM on the same linkage group.) As used herein, linkage can be between two markers, or alternatively between a marker and a locus affecting a phenotype. A marker locus can be “associated with” (linked to) a trait. The degree of linkage of a marker locus and a locus affecting a phenotypic trait is measured, e.g., as a statistical probability of co-segregation of that molecular marker with the phenotype (e.g., an F statistic or LOD score).

Linkage disequilibrium is most commonly assessed using the measure $r^2$, which is calculated using the formula described by Hill, W.G. and Robertson, A; Theor. Appl. Genet. 38:226-231(1968). When $r^2 = 1$, complete linkage disequilibrium exists between the two marker loci, meaning that the markers have not been separated by recombination and have the same allele frequency. The $r^2$ value will be dependent on the population used. Values for $r^2$ above 1/3 indicate sufficiently
strong linkage disequilibrium to be useful for mapping (Ardlie et al., Nature Reviews Genetics 3:299-309 (2002)). Hence, alleles are in linkage disequilibrium when $r^2$ values between pairwise marker loci are greater than or equal to 0.33, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1.0.

As used herein, “linkage equilibrium” describes a situation where two markers independently segregate, i.e., sort among progeny randomly. Markers that show linkage equilibrium are considered unlinked (whether or not they lie on the same chromosome).

A “locus” is a position on a chromosome, e.g. where a nucleotide, gene, sequence, or marker is located.

The “logarithm of odds (LOD) value” or “LOD score” (Risch, Science 255:803-804 (1992)) is used in genetic interval mapping to describe the degree of linkage between two marker loci. A LOD score of three between two markers indicates that linkage is 1000 times more likely than no linkage, while a LOD score of two indicates that linkage is 100 times more likely than no linkage. LOD scores greater than or equal to two may be used to detect linkage. LOD scores can also be used to show the strength of association between marker loci and quantitative traits in “quantitative trait loci” mapping. In this case, the LOD score’s size is dependent on the closeness of the marker locus to the locus affecting the quantitative trait, as well as the size of the quantitative trait effect.

A “marker” is a means of finding a position on a genetic or physical map, or else linkages among markers and trait loci (loci affecting traits). The position that the marker detects may be known via detection of polymorphic alleles and their genetic mapping, or else by hybridization, sequence match or amplification of a sequence that has been physically mapped. A marker can be a DNA marker (detects DNA polymorphisms), a protein (detects variation at an encoded polypeptide), or a simply inherited phenotype (such as the ‘waxy’ phenotype). A DNA marker can be developed from genomic nucleotide sequence or from expressed nucleotide sequences (e.g., from a spliced RNA or a cDNA). Depending on the DNA marker technology, the marker will consist of complementary primers flanking the locus and/or complementary probes that hybridize to polymorphic alleles at the locus. A DNA marker, or a genetic marker, can also be used to describe the gene, DNA
sequence or nucleotide on the chromosome itself (rather than the components used to detect the gene or DNA sequence) and is often used when that DNA marker is associated with a particular trait in human genetics (e.g. a marker for breast cancer). The term marker locus is the locus (gene, sequence or nucleotide) that the marker detects.

Markers that detect genetic polymorphisms between members of a population are well-established in the art. Markers can be defined by the type of polymorphism that they detect and also the marker technology used to detect the polymorphism. Marker types include but are not limited to, e.g., detection of restriction fragment length polymorphisms (RFLP), detection of isozyme markers, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), detection of simple sequence repeats (SSRs), detection of amplified variable sequences of the plant genome, detection of self-sustained sequence replication, or detection of single nucleotide polymorphisms (SNPs).

SNPs can be detected e.g. via DNA sequencing, PCR-based sequence specific amplification methods, detection of polynucleotide polymorphisms by allele specific hybridization (ASH), dynamic allele-specific hybridization (DASH), molecular beacons, microarray hybridization, oligonucleotide ligase assays, Flap endonucleases, 5' endonucleases, primer extension, single strand conformation polymorphism (SSCP) or temperature gradient gel electrophoresis (TGGE). DNA sequencing, such as the pyrosequencing technology has the advantage of being able to detect a series of linked SNP alleles that constitute a haplotype. Haplotypes tend to be more informative (detect a higher level of polymorphism) than SNPs.

A “marker allele”, alternatively an “allele of a marker locus”, can refer to one of a plurality of polymorphic nucleotide sequences found at a marker locus in a population.

“Marker assisted selection” (or MAS) is a process by which individual plants are selected based on marker genotypes.

A “marker haplotype” refers to a combination of alleles at a marker locus.

A “marker locus” is a specific chromosome location in the genome of a species where a specific marker can be found. A marker locus can be used to track the presence of a second linked locus, e.g., one that affects the expression of a
phenotypic trait. For example, a marker locus can be used to monitor segregation of alleles at a genetically or physically linked locus.

A "marker probe" is a nucleic acid sequence or molecule that can be used to identify the presence of a marker locus, e.g., a nucleic acid probe that is complementary to a marker locus sequence, through nucleic acid hybridization. Marker probes comprising 30 or more contiguous nucleotides of the marker locus ("all or a portion" of the marker locus sequence) may be used for nucleic acid hybridization. Alternatively, in some aspects, a marker probe refers to a probe of any type that is able to distinguish (i.e., genotype) the particular allele that is present at a marker locus.

The term "molecular marker" may be used to refer to a genetic marker, as defined above, or an encoded product thereof (e.g., a protein) used as a point of reference when identifying a linked locus. A marker can be derived from genomic nucleotide sequences or from expressed nucleotide sequences (e.g., from a spliced RNA, a cDNA, etc.), or from an encoded polypeptide. The term also refers to nucleic acid sequences complementary to or flanking the marker sequences, such as nucleic acids used as probes or primer pairs capable of amplifying the marker sequence. A "molecular marker probe" is a nucleic acid sequence or molecule that can be used to identify the presence of a marker locus, e.g., a nucleic acid probe that is complementary to a marker locus sequence. Alternatively, in some aspects, a marker probe refers to a probe of any type that is able to distinguish (i.e., genotype) the particular allele that is present at a marker locus. Nucleic acids are "complementary" when they specifically hybridize in solution, e.g., according to Watson-Crick base pairing rules. Some of the markers described herein are also referred to as hybridization markers when located on an indel region, such as the non-collinear region described herein. This is because the insertion region is, by definition, a polymorphism vis a vis a plant without the insertion. Thus, the marker need only indicate whether the indel region is present or absent. Any suitable marker detection technology may be used to identify such a hybridization marker, e.g. SNP technology is used in the examples provided herein.
An allele “negatively” correlates with a trait when it is linked to it and when presence of the allele is an indicator that a desired trait or trait form will not occur in a plant comprising the allele.

“Nucleotide sequence”, “polynucleotide”, “nucleic acid sequence”, and “nucleic acid fragment” are used interchangeably and refer to a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. A “nucleotide” is a monomeric unit from which DNA or RNA polymers are constructed, and consists of a purine or pyrimidine base, a pentose, and a phosphoric acid group. Nucleotides (usually found in their 5'-monophosphate form) are referred to by their single letter designation as follows: “A” for adenylate or deoxyadenylate (for RNA or DNA, respectively), “C” for cytidylate or deoxyctydylate, “G” for guanylate or deoxyguanylate, “U” for uridylate, “T” for deoxythymidylate, “R” for purines (A or G), “Y” for pyrimidines (C or T), “K” for G or T, “H” for A or C or T, “I” for inosine, and “N” for any nucleotide.

The term “phenotype”, “phenotypic trait”, or “trait” can refer to the observable expression of a gene or series of genes. The phenotype can be observable to the naked eye, or by any other means of evaluation known in the art, e.g., weighing, counting, measuring (length, width, angles, etc.), microscopy, biochemical analysis, or an electromechanical assay. In some cases, a phenotype is directly controlled by a single gene or genetic locus, i.e., a “single gene trait” or a “simply inherited trait”. In the absence of large levels of environmental variation, single gene traits can segregate in a population to give a “qualitative” or “discrete” distribution, i.e. the phenotype falls into discrete classes. In other cases, a phenotype is the result of several genes and can be considered a “multigenic trait” or a “complex trait”.

Multigenic traits segregate in a population to give a “quantitative” or “continuous” distribution, i.e. the phenotype cannot be separated into discrete classes. Both single gene and multigenic traits can be affected by the environment in which they are being expressed, but multigenic traits tend to have a larger environmental component.

A “physical map” of the genome is a map showing the linear order of identifiable landmarks (including genes, markers, etc.) on chromosome DNA. However, in contrast to genetic maps, the distances between landmarks are
absolute (for example, measured in base pairs or isolated and overlapping contiguous genetic fragments) and not based on genetic recombination (that can vary in different populations).

A “plant” can be a whole plant, any part thereof, or a cell or tissue culture derived from a plant. Thus, the term “plant” can refer to any of: whole plants, plant components or organs (e.g., leaves, stems, roots, etc.), plant tissues, seeds, plant cells, and/or progeny of the same. A plant cell is a cell of a plant, taken from a plant, or derived through culture from a cell taken from a plant.

A “polymorphism” is a variation in the DNA between two or more individuals within a population. A polymorphism preferably has a frequency of at least 1% in a population. A useful polymorphism can include a single nucleotide polymorphism (SNP), a simple sequence repeat (SSR), or an insertion/deletion polymorphism, also referred to herein as an “indel”.

An allele “positively” correlates with a trait when it is linked to it and when presence of the allele is an indicator that the desired trait or trait form will occur in a plant comprising the allele.

The “probability value” or “p-value” is the statistical likelihood that the particular combination of a phenotype and the presence or absence of a particular marker allele is random. Thus, the lower the probability score, the greater the likelihood that a locus and a phenotype are associated. The probability score can be affected by the proximity of the first locus (usually a marker locus) and the locus affecting the phenotype, plus the magnitude of the phenotypic effect (the change in phenotype caused by an allele substitution). In some aspects, the probability score is considered “significant” or “nonsignificant”. In some embodiments, a probability score of 0.05 (p=0.05, or a 5% probability) of random assortment is considered a significant indication of association. However, an acceptable probability can be any probability of less than 50% (p=0.5). For example, a significant probability can be less than 0.25, less than 0.20, less than 0.15, less than 0.1, less than 0.05, less than 0.01, or less than 0.001.

A “production marker” or “production SNP marker” is a marker that has been developed for high-throughput purposes. Production SNP markers are developed to
detect specific polymorphisms and are designed for use with a variety of chemistries and platforms. The term "progeny" refers to the offspring generated from a cross.

A "progeny plant" is a plant generated from a cross between two plants.

The term "quantitative trait locus" or "QTL" refers to a region of DNA that is associated with the differential expression of a quantitative phenotypic trait in at least one genetic background, e.g., in at least one breeding population. The region of the QTL encompasses or is closely linked to the gene or genes that affect the trait in question. An "allele of a QTL" can comprise multiple genes or other genetic factors within a contiguous genomic region or linkage group, such as a haplotype.

An allele of a QTL can denote a haplotype within a specified window wherein said window is a contiguous genomic region that can be defined, and tracked, with a set of one or more polymorphic markers. A haplotype can be defined by the unique fingerprint of alleles at each marker within the specified window.

A "reference sequence" or a "consensus sequence" is a defined sequence used as a basis for sequence comparison. The reference sequence for a PHM marker is obtained by sequencing a number of lines at the locus, aligning the nucleotide sequences in a sequence alignment program (e.g. Sequencher), and then obtaining the most common nucleotide sequence of the alignment. Polymorphisms found among the individual sequences are annotated within the consensus sequence. A reference sequence is not usually an exact copy of any individual DNA sequence, but represents an amalgam of available sequences and is useful for designing primers and probes to polymorphisms within the sequence.

In "repulsion" phase linkage, the "favorable" allele at the locus of interest is physically linked with an "unfavorable" allele at the proximal marker locus, and the two "favorable" alleles are not inherited together (i.e., the two loci are "out of phase" with each other).

Sorghum is a genus of numerous species of grasses, some of which are raised for grain and many of which are used as fodder plants either cultivated or as part of pasture. Sorghum is in the subfamily Panicoideae and the tribe Andropogoneae. Sorghum is well adapted to growth in hot, arid or semi-arid areas. The many subspecies are divided into four groups: grain sorghums (such as milo), grass sorghum (for pasture and hay), sweet sorghum (used to produce sorghum
syrups), and broom corn (for brooms and brushes). The name sweet sorghum is used to identify varieties of *Sorghum bicolor* that are sweet and juicy. High biomass sorghum can be used as a source of biofuels.

Sorghum species encompassed in this disclosure include, but are not limited to, *Sorghum alnum*, *Sorghum amplum*, *Sorghum angustum*, *Sorghum arundinaceum*, *Sorghum bicolor* (primary cultivated species), *Sorghum brachypodum*, *Sorghum bulbosum*, *Sorghum burmahicum*, *Sorghum controversum*, *Sorghum drummondii*, *Sorghum ecarinatum*, *Sorghum extans*, *Sorghum grande*, *Sorghum halepense*, *Sorghum interjectum*, *Sorghum intrans*, *Sorghum lax forum*, *Sorghum leiocladum*, *Sorghum macrospernum*, *Sorghum matarankense*, *Sorghum miliaceum*, *Sorghum nigrum*, *Sorghum ntidum*, *Sorghum plumosum*, *Sorghum propinquum*, *Sorghum purpureosericeum*, *Sorghum stipoideum*, *Sorghum timorense*, *Sorghum trichocladum*, *Sorghum versicolor*, *Sorghum virgatum*, and *Sorghum vulgare*.

The term "sorghum plant" includes whole sorghum plants, sorghum plant cells, sorghum plant protoplast, sorghum plant cell or sorghum tissue culture from which sorghum plants can be regenerated, sorghum plant calli, sorghum plant clumps and sorghum plant cells that are intact in sorghum plants or parts of sorghum plants, such as sorghum seeds, sorghum flowers, sorghum cotyledons, sorghum leaves, sorghum stems, sorghum buds, sorghum roots, sorghum root tips and the like.

A "topcross test" is a test performed by crossing each individual (e.g. a selection, inbred line, clone or progeny individual) with the same pollen parent or "tester", usually a homozygous line.

The phrase "under stringent conditions" refers to conditions under which a probe or polynucleotide will hybridize to a specific nucleic acid sequence, typically in a complex mixture of nucleic acids, but to essentially no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at
which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is at 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. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions are often: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 0.1% SDS at 65°C. For PCR, a temperature of about 36°C is typical for low stringency amplification, although annealing temperatures may vary between about 32°C and 48°C, depending on primer length. Additional guidelines for determining hybridization parameters are provided in numerous references.

An "unfavorable allele" of a marker is a marker allele that segregates with the unfavorable plant phenotype, therefore providing the benefit of identifying plants that can be removed from a breeding program or planting.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Sambrook").

Day length is the most important environmental factor involved in flower induction. The phenomenon whereby daylength regulates flowering is referred to as photoperiodism. An understanding of the effect of daylength on reproductive development has agronomic importance because the ability to alter flowering time allows the cultivation of a species in environments that differ greatly from the one in which it originally evolved.

Photoperiod sensitivity refers to the sensitivity of flowering time of plants to daylength. Some plants do not flower until they are exposed to daylengths that are less than a critical photoperiod (short day plants) or greater than a critical
photoperiod (long day plants). Long day (LD) and short day (SD) plant designations refer to the daylength required to induce flowering. Most plants including sorghum must pass through a juvenile stage (lasting ~14-21 days for sorghum) before they become sensitive to photoperiod.

Sorghum is a facultative short day plant where long days inhibit flowering and short days accelerate flowering. The degree of photoperiod sensitivity in sorghum refers to the length of short days that are required to induce flowering. A highly photoperiod sensitive sorghum will require photoperiods less than ~12 hours before flowering occurs whereas plants with low photoperiod sensitivity (or "photoperiod insensitivity”) only require day lengths less than ~14 hours to induce flowering.

Different sorghum cultivars vary in their degree of photoperiod sensitivity. Sorghum inbreds have been identified with photoperiod sensitivity ranging from ~10.5 to ~14 hours and others are insensitive to photoperiod.

For example “early flowering sorghum” as described herein may be a plant that flowers in 45 to 75 days; a “late flowering sorghum” as described herein may be a plant that flowers 80 to 120 days or more after planting or does not flower under these conditions. The number of days to flowering depends on the planting date and photoperiod because these factors determine when the plants are exposed to days that are sufficiently short to induce flowering.

In general, “late flowering photoperiod sensitive sorghum plants” described herein will not flower until day lengths are less than 11-12 hours; whereas the “less photoperiod sensitive” or “photoperiod insensitive” or “early flowering plants” described herein will flower when exposed to daylengths (photoperiods) of approximately 10-14 hrs or longer, depending on the genotype.

Photoperiod sensitivity and late flowering is mediated in sorghum and rice by genes that repress activation of genes in the FT gene family (flowering locus T) and AP1, which control the transition of the apex from vegetative growth to formation of reproductive structures. The repressors of flowering in sorghum act in a dominant or additive fashion and are inactivated or less active under short photoperiods (or thermal periods).

Numerous QTLs that modulate flowering time in sorghum have been identified in various studies (e.g. Lin et al (1995) Genetics 141:391-411; Paterson et

Embodiments:

Genetic mapping

It has been recognized for quite some time that specific genetic loci correlating with particular traits can be mapped in an organism's genome. The plant breeder can advantageously use molecular markers to identify desired individuals by detecting marker alleles that show a statistically significant probability of co-segregation with a desired phenotype, manifested as linkage disequilibrium. By identifying a molecular marker or clusters of molecular markers that co-segregate with a trait of interest, the breeder is able to rapidly select a desired phenotype by selecting for the proper molecular marker allele (a process called marker-assisted selection).

A variety of methods well known in the art are available for detecting molecular markers or clusters of molecular markers that co-segregate with a trait of interest, such as flowering time in sorghum. The basic idea underlying these methods is the detection of markers, for which alternative genotypes (or alleles) have significantly different average phenotypes. Thus, one makes a comparison among marker loci of the magnitude of difference among alternative genotypes (or alleles) or the level of significance of that difference. Trait genes are inferred to be located nearest the marker(s) that have the greatest associated genotypic difference. Two such methods used to detect trait loci of interest are: 1) Population-based association analysis and 2) Traditional linkage analysis.

In a population-based association analysis, lines are obtained from pre-existing populations with multiple founders, e.g. elite breeding lines. Population-based association analyses rely on linkage disequilibrium (LD) and the idea that in an unstructured population, only correlations between genes controlling a trait of
interest and markers closely linked to those genes will remain after so many generations of random mating. In reality, most pre-existing populations have population substructure. Thus, the use of a structured association approach helps to control population structure by allocating individuals to populations using data obtained from markers randomly distributed across the genome, thereby minimizing disequilibrium due to population structure within the individual populations (also called subpopulations). The phenotypic values are compared to the genotypes (alleles) at each marker locus for each line in the subpopulation. A significant marker-trait association indicates the close proximity between the marker locus and one or more genetic loci that are involved in the expression of that trait.

The same principles underlie traditional linkage analysis; however, linkage disequilibrium is generated by creating a population from a small number of founders. The founders are selected to maximize the level of polymorphism within the constructed population, and polymorphic sites are assessed for their level of cosegregation with a given phenotype. A number of statistical methods have been used to identify significant marker-trait associations. One such method is an interval mapping approach (Lander and Botstein, Genetics 121:185-199 (1989), in which each of many positions along a genetic map (say at 1 cM intervals) is tested for the likelihood that a gene controlling a trait of interest is located at that position. The genotype/phenotype data are used to calculate for each test position a LOD score (log of likelihood ratio). When the LOD score exceeds a threshold value, there is significant evidence for the location of a gene controlling the trait of interest at that position on the genetic map (which will fall between two particular marker loci).

The present invention provides sorghum marker loci that demonstrate statistically significant co-segregation with flowering time, photoperiod sensitivity, as determined by association mapping and traditional linkage analysis. Detection of these loci or additional linked loci can be used in marker assisted sorghum breeding programs to produce plants with early flowering time and/or photoperiod insensitivity, or to produce plants with late flowering time.
QTL locations

A QTL on chromosome 10 was identified as being associated with flowering time and photoperiod sensitivity in sorghum plants (Example 1). The QTL is located at about 47.9-65.4 cM on a single meiosis based genetic map.

Chromosomal intervals

Chromosomal intervals that correlate with flowering time and/or photoperiod sensitivity in sorghum are provided. A variety of methods well known in the art are available for identifying chromosomal intervals. The boundaries of such chromosomal intervals are drawn to encompass markers that will be linked to the gene(s) controlling the trait of interest. In other words, the chromosomal interval is drawn such that any marker that lies within that interval (including the terminal markers that define the boundaries of the interval) can be used as a marker for flowering time and/or photoperiod insensitivity. Tables 1 and 2 show markers within the chromosome 10 QTL region which were shown herein to associate with flowering time and/or photoperiod sensitivity.

Each interval comprises at least one QTL, and furthermore, may indeed comprise more than one QTL. Close proximity of multiple QTL in the same interval may obfuscate the correlation of a particular marker with a particular QTL, as one marker may demonstrate linkage to more than one QTL. Conversely, e.g., if two markers in close proximity show co-segregation with the desired phenotypic trait, it is sometimes unclear if each of those markers identifies the same QTL or two different QTL. Regardless, knowledge of how many QTL are in a particular interval is not necessary to make or practice the invention.

The intervals described below encompass markers that co-segregate with flowering time and/or photoperiod sensitivity (Table 1 and 2) in a recombinant inbred population. The clustering of markers that co-segregate with a trait within a localized region occurs in relatively small domains on the chromosomes, indicating the presence of one or more QTL in those chromosome regions. The interval was drawn to encompass markers that co-segregate with flowering time (as well as the other related traits). The intervals are defined by the markers on their termini, where the interval encompasses markers that map within the interval as well as the markers that define the termini. An interval described by the terminal markers that
define the endpoints of the interval will include the terminal markers and any marker localizing within that chromosomal domain, whether those markers are currently known or unknown. Chromosome 10 intervals described herein include:

A) the interval defined by and including V02323-001-I001 and V01796-1-I001, by 17.5 cM on a single meiosis based genetic map;

B) a subinterval of A) defined by and including V02324-001-I001 and V01757-1-I001, which are separated by 5.3 cM on a single meiosis based genetic map.

Any marker located within these intervals can find use as a marker for flowering time, photoperiod sensitivity and can be used in the context of the methods presented herein to identify and/or select sorghum plants that have early or late flowering phenotype; and/or photoperiod insensitivity.

The chromosome 10 interval may encompass any of the markers identified herein as being associated with flowering time and/or photoperiod sensitivity/insensitivity especially V02324-001-I001, V02131-001-I001, V02132-001-I001, V02330-001-I001, and V01757-1-I001.

Chromosomal intervals can also be defined by markers that are linked to (show linkage disequilibrium with) a QTL marker and $r^2$ is a common measure of linkage disequilibrium in the context of association studies. If the $r^2$ value of linkage disequilibrium between a chromosome 10 marker locus located at or near the QTL associated with flowering time or photoperiod sensitivity, for example, and another chromosome 10 marker locus in close proximity is greater than 1/3 (Ardlie et al., Nature Reviews Genetics 3:299-309 (2002)), the loci are considered to be in linkage disequilibrium with one another.

Markers and linkage relationships

A common measure of linkage is the frequency with which traits cosegregate. This can be expressed as a percentage of cosegregation (recombination frequency) or in centiMorgans (cM). The cM is a unit of measure of genetic recombination frequency. One cM is equal to a 1% chance that a trait at one genetic locus will be separated from a trait at another locus due to crossing over in a single generation (meaning the traits segregate together 99% of the time). Because chromosomal distance is approximately proportional to the frequency of crossing over events
between traits, there is an approximate physical distance that correlates with recombination frequency.

Marker loci are themselves traits and can be assessed according to standard linkage analysis by tracking the marker loci during segregation. Thus, one cM is equal to a 1% chance that a marker locus will be separated from another locus, due to crossing over in a single generation.

The closer a marker is to a gene controlling a trait of interest, the more effective and advantageous that marker is as an indicator for the desired trait. Closely linked loci display an inter-locus cross-over frequency of about 10% or less, preferably about 9% or less, still more preferably about 8% or less, yet more preferably about 7% or less, still more preferably about 6% or less, yet more preferably about 5% or less, still more preferably about 4% or less, yet more preferably about 3% or less, and still more preferably about 2% or less. In highly preferred embodiments, the relevant loci (e.g., a marker locus and a target locus) display a recombination frequency of about 1% or less, e.g., about 0.75% or less, more preferably about 0.5% or less, or yet more preferably about 0.25% or less. Thus, the loci are about 10 cM, 9 cM, 8 cM, 7 cM, 6 cM, 5 cM, 4 cM, 3 cM, 2 cM, 1 cM, 0.75 cM, 0.5 cM or 0.25 cM or less apart. Put another way, two loci that are localized to the same chromosome, and at such a distance that recombination between the two loci occurs at a frequency of less than 10% (e.g., about 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25%, or less) are said to be "proximal to" each other.

Although particular marker alleles can co-segregate with early or late flowering time, it is important to note that the marker locus is not necessarily responsible for the expression of the flowering time phenotype. For example, it is not a requirement that the marker polynucleotide sequence be part of a gene that is responsible for the phenotype (for example, is part of the gene open reading frame). The association between a specific marker allele and a trait is due to the original "coupling" linkage phase between the marker allele and the allele in the ancestral sorghum line from which the allele originated. Eventually, with repeated recombination, crossing over events between the marker and genetic locus can change this orientation. For this reason, the favorable marker allele may change
depending on the linkage phase that exists within the parent having the favorable trait that is used to create segregating populations. This does not change the fact that the marker can be used to monitor segregation of the phenotype. It only changes which marker allele is considered favorable in a given segregating population.

Methods presented herein include detecting the presence of one or more marker alleles associated with an early or late flowering time phenotype (and/or photoperiod insensitivity) in a sorghum plant and then identifying and/or selecting sorghum plants that have favorable alleles at those marker loci or detecting the presence of a marker allele associated with the other state of the trait and then identifying and/or counterselecting sorghum plants that have unfavorable alleles.

Markers listed in Tables 1 and 2 have been identified herein as being associated with flowering time in sorghum and hence can be used to identify and select sorghum plants having the phenotype of early or late flowering time and/or decreased or increased photoperiod sensitivity (or a related trait). Any marker within 20 cM, 15 cM, 10 cM, 9 cM, 8 cM, 7 cM, 6 cM, 5 cM, 4 cM, 3 cM, 2 cM, 1 cM, 0.9 cM, 0.8 cM, 0.7 cM, 0.6 cM, 0.5 cM, 0.4 cM, 0.3 cM, 0.2 cM, 0.1 cM or less (based on a single meiosis map) of any of the markers in Tables 1 and 2 could also be used to identify and select sorghum plants with early or late flowering time and/or decreased or increased photoperiod sensitivity (or a related trait). Any marker allele linked to and associated with the favorable alleles of the markers listed in Table 2 can be used for detection purposes in the identification and/or selection of plants with early or late flowering time.

Marker assisted selection

Molecular markers can be used in a variety of plant breeding applications (e.g. see Staub et al. (1996) Hortscience 31: 729-741; Tanksley (1983) Plant Molecular Biology Reporter. 1: 3-8). One of the main areas of interest is to increase the efficiency of backcrossing and introgressing genes using marker-assisted selection. A molecular marker that demonstrates linkage with a locus affecting a desired phenotypic trait provides a useful tool for the selection of the trait in a plant population. This is particularly true where the phenotype is hard to assay. Since DNA marker assays are less laborious and take up less physical space than field
phenotyping, much larger populations can be assayed, increasing the chances of finding a recombinant with the target segment from the donor line moved to the recipient line. The closer the linkage, the more useful the marker, as recombination is less likely to occur between the marker and the gene causing the trait, which can result in false positives. Having flanking markers decreases the chances that false positive selection will occur as a double recombination event would be needed. The ideal situation is to have a marker in the gene itself, so that recombination cannot occur between the marker and the gene. Such a marker is called a 'perfect marker'.

When a gene is introgressed by marker assisted selection, it is not only the gene that is introduced but also the flanking regions (Gepts. (2002). Crop Sci; 42: 1780-1790). This is referred to as "linkage drag." In the case where the donor plant is highly unrelated to the recipient plant, these flanking regions carry additional genes that may code for agronomically undesirable traits. This "linkage drag" may also result in reduced yield or other negative agronomic characteristics even after multiple cycles of backcrossing into the elite sorghum line. This is also sometimes referred to as "yield drag." The size of the flanking region can be decreased by additional backcrossing, although this is not always successful, as breeders do not have control over the size of the region or the recombination breakpoints (Young et al. (1998) Genetics 120:579-585). In classical breeding it is usually only by chance that recombinations are selected that contribute to a reduction in the size of the donor segment ( Tanksley et al. (1989). Biotechnology 7: 257-264). Even after 20 backcrosses in backcrosses of this type, one may expect to find a sizeable piece of the donor chromosome still linked to the gene being selected. With markers however, it is possible to select those rare individuals that have experienced recombination near the gene of interest. In 150 backcross plants, there is a 95% chance that at least one plant will have experienced a crossover within 1 cM of the gene, based on a single meiosis map distance. Markers will allow unequivocal identification of those individuals. With one additional backcross of 300 plants, there would be a 95% chance of a crossover within 1 cM single meiosis map distance of the other side of the gene, generating a segment around the target gene of less than 2 cM based on a single meiosis map distance. This can be accomplished in two generations with markers, while it would have required on average 100
generations without markers (See Tanksley et al., supra). When the exact location of a gene is known, flanking markers surrounding the gene can be utilized to select for recombinations in different population sizes. For example, in smaller population sizes, recombinations may be expected further away from the gene, so more distal flanking markers would be required to detect the recombination.

The key components to the implementation of marker assisted selection are: (i) Defining the population within which the marker-trait association will be determined, which can be a segregating population, or a random or structured population; (ii) monitoring the segregation or association of polymorphic markers relative to the trait, and determining linkage or association using statistical methods; (iii) defining a set of desirable markers based on the results of the statistical analysis, and (iv) the use and/or extrapolation of this information to the current set of breeding germplasm to enable marker-based selection decisions to be made. The markers described in this disclosure, as well as other marker types such as SSRs and FLPS, can be used in marker assisted selection protocols.


Various types of SSR markers can be generated, and SSR profiles can be obtained by gel electrophoresis of the amplification products. Scoring of marker genotype is based on the size of the amplified fragment. Various types of FLP markers can also be generated. Most commonly, amplification primers are used to generate fragment length polymorphisms. Such FLP markers are in many ways similar to SSR markers, except that the region amplified by the primers is not
typically a highly repetitive region. Still, the amplified region, or amplicon, will have sufficient variability among germplasm, often due to insertions or deletions, such that the fragments generated by the amplification primers can be distinguished among polymorphic individuals, and such indels are known to occur frequently in sorghum (Evans et al. PLoS One(2013). 8 (11): e79192).

SNP markers detect single base pair nucleotide substitutions. Of all the molecular marker types, SNPs are the most abundant, thus having the potential to provide the highest genetic map resolution (PLoS One(2013). 8 (11): e79192). SNPs can be assayed at an even higher level of throughput than SSRs, in a so-called `ultra-high-throughput` fashion, as they do not require large amounts of DNA and automation of the assay may be straight-forward. SNPs also have the promise of being relatively low-cost systems. These three factors together make SNPs highly attractive for use in marker assisted selection. Several methods are available for SNP genotyping, including but not limited to, hybridization, primer extension, oligonucleotide ligation, nuclease cleavage, minisequencing and coded spheres. Such methods have been reviewed in: Gut (2001) Hum Mutat 17 pp. 475-492; Shi (2001) Clin Chem 47, pp. 164-172; Kwok (2000) Pharmacogenomics 1, pp. 95-100; and Bhattaramaki and Rafalski (2001) Discovery and application of single nucleotide polymorphism markers in plants. In: R. J. Henry, Ed, Plant Genotyping: The DNA Fingerprinting of Plants, CABI Publishing, Wallingford. A wide range of commercially available technologies utilize these and other methods to interrogate SNPs including Masscode.TM. (Qiagen), INVADER®. (Third Wave Technologies) and Invader PLUS®, SNAPSHOT®. (Applied Biosystems), TAQMAN®. (Applied Biosystems) and BEADARRAYS®. (Illumina).

A number of SNPs together within a sequence, or across linked sequences, can be used to describe a haplotype for any particular genotype (Ching et al. (2002), BMC Genet. 3:19 pp Gupta et al. 2001, Rafalski (2002b), Plant Science 162:329-333). Haplotypes can be more informative than single SNPs and can be more descriptive of any particular genotype. For example, a single SNP may be allele `T` for a specific line or variety with early maturity, but the allele `T` might also occur in the sorghum breeding population being utilized for recurrent parents. In this case, a haplotype, e.g. a combination of alleles at linked SNP markers, may be more
informative. Once a unique haplotype has been assigned to a donor chromosomal region, that haplotype can be used in that population or any subset thereof to determine whether an individual has a particular gene. See, for example, WO2003054229. Using automated high throughput marker detection platforms known to those of ordinary skill in the art makes this process highly efficient and effective.

In addition to SSR’s, FLPs and SNPs, as described above, other types of molecular markers are also widely used, including but not limited to expressed sequence tags (ESTs), SSR markers derived from EST sequences, randomly amplified polymorphic DNA (RAPD), and other nucleic acid based markers.

Isozyme profiles and linked morphological characteristics can, in some cases, also be indirectly used as markers. Even though they do not directly detect DNA differences, they are often influenced by specific genetic differences. However, markers that detect DNA variation are far more numerous and polymorphic than isozyme or morphological markers ( Tanksley (1983) Plant Molecular Biology Reporter 1:3-8).

Sequence alignments or contigs may also be used to find sequences upstream or downstream of the specific markers listed herein. These new sequences, close to the markers described herein, are then used to discover and develop functionally equivalent markers. For example, different physical and/or genetic maps are aligned to locate equivalent markers not described within this disclosure but that are within similar regions. These maps may be within the sorghum species, or even across other species that have been genetically or physically aligned with sorghum, such as maize, rice, wheat, or barley.

In general, marker assisted selection uses polymorphic markers that have been identified as having a significant likelihood of co-segregation with a phenotype, such as flowering time and/or photoperiod sensitivity in sorghum. Such markers are presumed to map near a gene or genes that regulate the flowering time and/or photoperiod sensitivity in sorghum plant, and are considered indicators for the desired trait, or markers. Plants are tested for the presence of a desired allele in the marker, and plants containing a desired genotype at one or more loci are expected to transfer the desired genotype, along with a desired phenotype, to their progeny.
Thus, sorghum plants with early (or delayed) flowering time and/or increased or decreased photoperiod sensitivity can be selected for by detecting one or more marker alleles, and in addition, progeny plants derived from those plants can also be selected. Hence, a plant containing a desired genotype in a given chromosomal region is obtained and then crossed to another plant. The progeny of such a cross would then be evaluated genotypically using one or more markers and the progeny plants with the same genotype in a given chromosomal region would then be selected as exhibiting early flowering time and/or photoperiod insensitivity; or late flowering time.

Markers were identified from both linkage mapping and association analysis as being associated with flowering time and/or photoperiod sensitivity. Reference sequences for each of the markers are represented by SEQ ID NOs: 1-8. Moreover, the SNPs identified in Table 1 or Table 2 could be used alone or in combination (i.e. a SNP haplotype) to select for plants having a favorable QTL allele (i.e. associated with early or late flowering phenotype).

The skilled artisan would expect that there might be additional polymorphic sites at marker loci in and around the chromosome 10 markers identified herein, wherein one or more polymorphic sites is in linkage disequilibrium with an allele at one or more of the polymorphic sites in the haplotype and thus could be used in a marker assisted selection program to introgress a QTL allele of interest. Two particular alleles at different polymorphic sites are said to be in linkage disequilibrium if the presence of the allele at one of the sites tends to predict the presence of the allele at the other site on the same chromosome (Stevens, *Mol. Diag.* 4:309-17 (1999)). The marker alleles that can be used for marker associated selection of sorghum plants with early flowering phenotype and/or photoperiod insensitivity can be linked to and associated with any or all of the following: a “T” at marker V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001. The marker alleles that can be used for marker associated selection of sorghum plants with late flowering phenotype can be linked to and associated with any or all of the following: a “C” at marker V02324-001-I001, a “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001. The markers may be linked by
20 cM, 15 cM, 10 cM, 9 cM, 8 cM, 7 cM, 6 cM, 5 cM, 4 cM, 3 cM, 2 cM, 1 cM, 0.9 cM, 0.8 cM, 0.7 cM, 0.6 cM, 0.5 cM, 0.4 cM, 0.3 cM, 0.2 cM, 0.1 cM or less (on a single meiosis based genetic map).

The skilled artisan would understand that allelic frequency (and hence, haplotype frequency) can differ from one germplasm pool to another. Germplasm pools vary due to maturity differences, heterotic groupings, geographical distribution, etc. As a result, SNPs and other polymorphisms may not be informative in some germplasm pools.

Plant compositions

Sorghum plants identified and/or selected by any of the methods described above are also of interest.

Applications

Methods and compositions are provided herein for modulating photoperiod sensitivity and flowering time in sorghum. The methods disclosed herein may be used for various purposes, examples include, but are not limited to, selecting for late flowering plants that can be used for biomass or bioenergy production, and to provide for high and/or staggered biomass production.

Another example of the way the QTL for flowering time disclosed herein can be used is to design sweet sorghum hybrids that grow for a specific number of days prior to flowering from inbreds suitable for hybrid seed production.

In certain aspects of this disclosure, the methods and compositions provided herein can be used to produce inbreds which, when crossed, generate high biomass bioenergy sorghum hybrid seed that can be planted at any time of the year suitable for substantially continual production with reliable flowering times, maturity and harvest for optimum biomass and moisture levels. Such hybrid plants may have long growth duration (i.e. late flowering or even non-flowering in some conditions). In another aspect, this system can be used to produce sweet sorghum hybrids that grow for a specified number of days prior to flowering at different latitudes, enabling the best harvest timing for optimum sugar yields.

EXAMPLES

The following examples are offered to illustrate, but not to limit, the claimed invention. It is understood that the examples and embodiments described herein
are for illustrative purposes only, and persons skilled in the art will recognize various reagents or parameters that can be altered without departing from the spirit of the invention or the scope of the appended claims.

**EXAMPLE 1**

**Mapping of a QTL for Photoperiod Sensitivity**

A recombinant inbred line (RIL) population, herein referred to as RIL PS1, that was segregating for flowering time and photoperiod sensitivity was used in mapping studies to identify regions controlling photoperiod sensitivity. The populations were produced by making a cross and selfing the progeny through the F₄ stage. After selfing the F₄ individuals, the F₅ seed was bulked to have enough seed to phenotype in several environments.

*Phenotypic Analysis*

For each population, flowering time data (collected as mid-bloom data, or number of days from planting until half of the plants in the row were flowering) was obtained at two locations: Texas and Puerto Rico. Data was obtained from Puerto Rico in Year 1 and Year 2, while data was only obtained from Texas in Year 1. Data from Puerto Rico was taken on F₃ plants and the average of the head row was used for the subsequent F₄ generation, while data from Texas was taken on F₄ plants and the average of the head row was used for the subsequent F₅ bulks. Photoperiod sensitivity was calculated for Year 1 by subtracting the Year 1 Puerto Rico mid-bloom data from the Year 1 Texas mid-bloom data. Photoperiod sensitivity data was calculated for Year 2 by subtracting the Year 2 Puerto Rico mid-bloom data from the Year 1 Texas mid-bloom data. Plant height data was also obtained at each of the locations for each year of testing to investigate correlations between plant height and photoperiod sensitivity.

*Genotypic Analysis*

Genotyping was performed using an Illumina 768 array. There were 109 polymorphic markers. QTL mapping was performed using both interval mapping (QTL Cartographer) and biparental association mapping (GWAS, or genome wide association studies).
QTL Discovery Analysis

Phenotype and genotype data were loaded into Windows QTL Cartographer v 2.5 and QGene v4.3. Composite interval mapping (CIM) showed a significant QTL associated with midbloom on chromosome 10 using the data from Year 1 Puerto Rico (LOD score = 13) and from Year 1 Texas (LOD score = 14). The QTL on chromosome 10 was also associated with Year 1 photoperiod sensitivity (LOD score = 7). The same data was also analyzed using proprietary association mapping software with the results confirming those found with interval mapping. In addition, the whole genome biparental association mapping showed that the QTL on chromosome 10 was also associated with plant height using data from Year 1. For Year 2 data, the chromosome 10 QTL also was observed for photoperiod sensitivity.

QTL and haplotype characterization

The QTL found on C10 in the RIL PS1 population has a large effect (about 7 days), so further analysis was done in this region. Table 1 shows a listing of markers, chromosomal position (cM), p-values and effect sizes ($R^2$) in the C10 QTL region. The markers with the highest significance/effects are in bold. The QTL interval was identified between markers V02323-001-I001 and V01796-1-I001, which are located at 47.9 cM and 65.4 cM, respectively, on a single meiosis based genetic map.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Reference SEQ ID NO:</th>
<th>cM</th>
<th>p-value</th>
<th>R-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>V02323-001-I001</td>
<td>1</td>
<td>47.9</td>
<td>9.10E-15</td>
<td>0.2449</td>
</tr>
<tr>
<td>V02324-001-I001</td>
<td>2</td>
<td>48.9</td>
<td>1.88E-21</td>
<td>0.2822</td>
</tr>
<tr>
<td>V02131-001-I001</td>
<td>3</td>
<td>52.9</td>
<td>2.22E-16</td>
<td>0.2755</td>
</tr>
<tr>
<td>V02132-001-I001</td>
<td>4</td>
<td>53</td>
<td>4.00E-15</td>
<td>0.2594</td>
</tr>
<tr>
<td>V02330-001-I001</td>
<td>5</td>
<td>53.4</td>
<td>4.00E-15</td>
<td>0.2594</td>
</tr>
<tr>
<td>V01757-001-I001</td>
<td>6</td>
<td>54.2</td>
<td>4.89E-15</td>
<td>0.2772</td>
</tr>
<tr>
<td>V02332-001-I001</td>
<td>7</td>
<td>56.3</td>
<td>4.66E-15</td>
<td>0.2319</td>
</tr>
<tr>
<td>V01796-001-I001</td>
<td>8</td>
<td>65.4</td>
<td>2.83E-12</td>
<td>0.1914</td>
</tr>
</tbody>
</table>

Haplotypes comprising alleles at V02324-001-I001, V02131-001-I001, V02132-001-I001, V02330-001-I001, and V01757-1-I001 were identified (Table 2) as being associated with the early or late flowering phenotypes.
Table 2: Haplotypes associated with early or late flowering phenotypes

<table>
<thead>
<tr>
<th>Marker</th>
<th>SNP Position in Reference Sequence</th>
<th>Early</th>
<th>Late</th>
</tr>
</thead>
<tbody>
<tr>
<td>V02324-001-I001</td>
<td>201</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>V02131-001-I001</td>
<td>201</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>V02132-001-I001</td>
<td>201</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>V02330-001-I001</td>
<td>201</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>V01757-001-I001</td>
<td>501</td>
<td>C</td>
<td>G</td>
</tr>
</tbody>
</table>
CLAIMS

What is claimed is:

1. A method of identifying a sorghum plant comprising a favorable allele of a QTL associated with flowering time, said method comprising:
   a. screening a population with at least one marker to determine if one or more sorghum plants from the population comprises a favorable allele at the QTL associated with flowering time, wherein the marker is located within an interval on chromosome 10 comprising and flanked by V02323-001-I001 and V01796-1-I001; and
   b. selecting from said population at least one sorghum plant comprising the favorable allele.

2. The method of claim 1, wherein said favorable allele is a QTL allele associated with early flowering time and/or photoperiod insensitivity.

3. The method of claim 1, wherein said favorable allele is a QTL allele associated with late flowering time.

4. The method of claim 2, wherein said QTL allele associated with early flowering time and/or photoperiod insensitivity comprises at least one of the following: a “T” at marker V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001.

5. The method of claim 2, wherein said QTL allele associated with early flowering time and/or photoperiod insensitivity comprises: a “T” at marker V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001.

6. The method of claim 3, wherein said QTL allele comprises at least one of the following: a “C” at marker V02324-001-I001, a “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001.

7. The method of claim 3, wherein said QTL allele comprises: a “C” at marker V02324-001-I001, a “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001.
8. A method of identifying and/or selecting a sorghum plant that displays late flowering time, said method comprising:
   a. detecting in a sorghum plant an allele of a marker locus, wherein said marker locus is located within a chromosomal interval comprising and flanked by V02323-001-I001 and V01796-1-I001 and said allele is associated with a haplotype comprising: a “C” at V02324-001-I001, an “A” at V02131-001-I001, a “G” at V02132-001-I001, a “T” at V02330-001-I001, and a “G” at V01757-1-I001; and
   b. selecting a sorghum plant that has the allele of the marker locus that is associated with the haplotype.

9. A method of identifying and/or selecting a sorghum plant that displays late flowering time, said method comprising:
   a. detecting in a sorghum plant one or more of the following:
      i. a “C” at V02324-001-I001,
      ii. a “A” at V02131-001-I001,
      iii. a “G” at V02132-001-I001,
      iv. a “T” at V02330-001-I001, and
      v. a “G” at V01757-1-I001;
   b. selecting said sorghum plant that has one or more of (i)- (v).

10. A method of identifying and/or selecting a sorghum plant that displays early flowering time and/or photoperiod insensitivity, said method comprising:
    a. detecting in a sorghum plant an allele of a marker locus, wherein said marker locus is located within a chromosomal interval comprising and flanked by V02323-001-I001 and V01796-1-I001 and said allele is associated with a haplotype comprising: a “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001; and
    b. selecting a sorghum plant that has the allele of the marker locus that is associated with the haplotype.
11. A method of identifying and/or selecting a sorghum plant that displays early flowering time and/or photoperiod insensitivity, said method comprising:
   a. detecting in a sorghum plant one or more of the following:
      i. a "T" at V02324-001-I001,
      ii. a "C" at V02131-001-I001,
      iii. a "T" at V02132-001-I001,
      iv. a "C" at V02330-001-I001, and
      v. a "C" at V01757-1-I001;
   b. selecting said sorghum plant that has one or more of (i)-(v).

12. A method of selecting a sorghum plant that displays late flowering time, the method comprising:
   a. obtaining a first sorghum plant that comprises within its genome: a "C" at V02324-001-I001, an "A" at V02131-001-I001, a "G" at V02132-001-I001, a "T" at V02330-001-I001, and a "G" at V01757-1-I001;
   b. crossing said first sorghum plant to a second sorghum plant;
   c. evaluating progeny plants for the presence of:
      i. a "C" at V02324-001-I001, an "A" at V02131-001-I001, a "G" at V02132-001-I001, a "T" at V02330-001-I001, and a "G" at V01757-1-I001; or
      ii. an allele of a marker locus located within a chromosomal interval defined by and including V02323-001-001 and V01796-1-I001 wherein said allele is associated with a haplotype comprising a "C" at V02324-001-I001, an "A" at V02131-001-I001, a "G" at V02132-001-I001, a "T" at V02330-001-I001, and a "G" at V01757-1-I001; and
   d. selecting progeny plants that have i) or ii).

13. A method of selecting a sorghum plant that displays early flowering time and/or photoperiod insensitivity, the method comprising:
   a. obtaining a first sorghum plant that comprises within its genome: a "T" at V02324-001-I001, a "C" at V02131-001-I001, a "T" at V02132-001-I001, a "C" at V02330-001-I001, and a "C" at V01757-1-I001;
b. crossing said first sorghum plant to a second sorghum plant;

c. evaluating progeny plants for the presence of:

i. a “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001; or

ii. an allele of a marker locus located within a chromosomal interval defined by and including V02323-001-001 and V01796-1-I001 wherein said allele is associated with a haplotype comprising a “T” at V02324-001-I001, a “C” at V02131-001-I001, a “T” at V02132-001-I001, a “C” at V02330-001-I001, and a “C” at V01757-1-I001; and

d. selecting progeny plants that have i) or ii).
**INTERNATIONAL SEARCH REPORT**

A. CLASSIFICATION OF SUBJECT MATTER

**INV. A01H1/04 A01H5/10 C12Q1/68**

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

A01H C12Q

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>
<pre><code>      | -/-                                | 1-3                   |
</code></pre>

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

**See patent family annex.**

---

Date of the actual completion of the international search: **26 July 2016**

Date of mailing of the international search report: **08/08/2016**

Name and mailing address of the ISA/
European Patent Office, P.B. 5018 Patentlaan 2
NL-2280 HV RIJWIJK
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer: Kania, Thomas
## INTERNATIONAL SEARCH REPORT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2012158630 A1</td>
</tr>
<tr>
<td>WO 2010011680 A2</td>
<td>28-01-2010</td>
<td>AU 2009274112 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 102131925 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2010024065 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2013139276 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2013237429 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2010011680 A2</td>
</tr>
<tr>
<td>US 2013298274 A1</td>
<td>07-11-2013</td>
<td>NONE</td>
</tr>
</tbody>
</table>