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

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
10 November 2016 (10.11.2016)

WIPO | PCT

(10) **International Publication Number**
WO 2016/177887 A1

(51) **International Patent Classification**:
C12N 15/82 (2006.01) ***A01H 5/08*** (2006.01)
C07K 14/415 (2006.01) ***C12Q 1/68*** (2006.01)
A01H 5/00 (2006.01)

(21) **International Application Number**:
PCT/EP2016/060202

(22) **International Filing Date**:
6 May 2016 (06.05.2016)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
15305699.9 7 May 2015 (07.05.2015) EP

(71) **Applicants**: **LIMAGRAIN EUROPE** [FR/FR]; rue Henri Mondor, Biopôle Clermont-Limagne, 63360 Saint-Beauzire (FR). **INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA)** [FR/FR]; 147, Rue de l'Université, 75338 Paris Cedex 07 (FR). **CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE** [FR/FR]; 3, rue Michel-Ange, F-75016 Paris (FR).

(72) **Inventors**: **MARTINANT, Jean-Pierre**; 6 rue Croix de Laire Basse, 63910 Vertaizon (FR). **COMADRAN, Jordi**; 3 rue Gilbert Romme, 63200 Riom (FR). **LAFFAIRE, Jean-Baptiste**; 5 rue des Aigueurs, 63200 Mozac (FR). **ROGOWSKY, Peter**; 45 rue du Rhône, 69007 Lyon (FR). **WIDIEZ, Thomas**; 6 rue de Cuire, 69004 Lyon (FR).

(74) **Agents**: **VEINANTE, Aude** et al.; Lavoix, 62, rue de Bonnel, 69003 Lyon (FR).

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

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: POLYNUCLEOTIDE RESPONSIBLE OF HAPLOID INDUCTION IN MAIZE PLANTS AND RELATED PROCESSES

(57) **Abstract**: The present invention concerns an isolated polynucleotide responsible of haploid induction in maize plants and related processes. Additionally, the invention relates to plants that have been genetically transformed with the polynucleotide of the invention. The invention also relates to a process for screening a mutant plant population for enhanced haploid induction by using said isolated polynucleotide. The invention further relates to molecular markers associated with haploid induction in maize plants and their use in quality control for inducer lines.

WO 2016/177887 A1

Polynucleotide responsible of haploid induction in maize plants and related processes

The present invention concerns an isolated polynucleotide responsible of haploid induction in maize plants and related processes. Additionally, the invention relates to plants that have been genetically transformed with the polynucleotide of the invention. The invention also relates to a process for screening a mutant plant population for enhanced haploid induction. The invention further relates to molecular markers associated with haploid induction in maize plants and their use for allele identification during the production of inducer lines.

The establishment of homozygous lines is a fundamental practice in selection and breeding. One of the major constraints in the establishment of homozygous lines is the long time (usually 8-10 generations) needed for obtaining individuals with a high level of homozygosity.

Doubled haploids represent a major breeding tool (Geiger *et al.*, Doubled haploids in hybrid maize breedings, Maydica, 54(4):485-499*,* 2009 and Röber *et al., In vivo* haploid induction in maize – Performance of new inducers and significance of doubled haploid lines in hybrid breeding, Maydica, 50(3-4): 275-283, 2005). It allows the rapid production of an homozygous line in fewer generations than traditional methods, can be used to benefit of a maximum genetic variance in breeding programs and to accelerate the stacking of genes in a recurrent line.

“Gynogenesis” (development of the non-fertilized ovule into a haploid plant) combined with chromosome doubling is a key step for maize selection. This first step is conducted by crossing the plant with an inducer line. During the past years, works have been done to develop said inducer lines. The inducer line PK6 was developed by the INRA of Clermont-Ferrand (WO 2005/004586) and comes from the publicly available stock 6 (Coe 1959). Few other inducer lines are available but all come from the same origin stock 6 (Weber *et al.*, Today's Use of Haploids in Corn Plant Breeding, Advances In Agronomy, 123: 123-144, 2014).

Improvement of the inducing ability has thus been achieved during the past years but the need to improve this ability still exists. In addition, the availability of molecular markers to follow the induction locus or loci is of interest.

A major QTL (quantitative trait locus) responsible of haploid inducing has been identified on chromosome 1 (bin 1.4) and some minor QTL on others chromosomes

(Deimling *et al.,* Methodology and genetics of *in vivo* haploid induction in maize, 38: 203–224, 1997 and Prigge *et al.,* Doubled Haploids in Tropical Maize: I. Effects of Inducers and Source Germplasm on in vivo Haploid Induction Rates. Crop Science, 51(4): 1498-1506, 2012). Fine mapping of the QTL on chromosome 1 was started by Barret *et al.* (Barret *et al.*, A major locus expressed in the male gametophyte with incomplete penetrance is responsible for *in situ* gynogenesis in maize, Theoretical And Applied Genetics, 117(4): 581-594, 2008) with the PK6 inducer line and the QTL *ggi1* (*gynogenesis inducer 1*) was localized between the markers umc 1917 and bnlg 1811 (Barret *et al.,* 2008, figure 2) in a 11.6 cM interval. The authors of this work concluded that a more precise cloning was not possible in their particular population because of the wild type standard deviation (percentage of haploid production obtained by crossing non inducer lines) and the heterofertilization (when egg and central cell are fertilized by independent pollen grains).

A more precise cloning of the chromosome 1 QTL of inducer line UH400 (Prigge *et al.,* New insights into the genetics of *in vivo* induction of maternal haploids, the backbone of doubled haploid technology in maize, Genetics, 190(2): 781-793, 2011), *qhir1*, was then achieved by Dong *et al.* (Dong et *al.*, Fine mapping of qhir1 influencing *in vivo* haploid induction in maize. Theoretical And Applied Genetics, 126(7): 1713-1720, 2013). This QTL is located between markers X291 and X263, with a distance between these markers of 243 Kb.

Another fine mapping of the region *sed1 (segregation distortion 1)* was achieved by Xu *et al.* (Xu *et al.*, Gametophytic and zygotic selection leads to segregation distortion through *in vivo* induction of a maternal haploid in maize. Journal Of Experimental Botany, 64(4): 1083-1096, 2013), and narrowed the QTL to a 450 kb region, between markers X22 and X96.

However, these fine mapping approaches did not identify the gene or precise DNA region responsible for this major QTL on chromosome 1.

In addition, the identification of new informative molecular markers for the inducing capacity of maize lines is still needed to improve the breeding of new inducer lines. Phenotypic markers are currently used to distinguish haploid from diploids plants. They present some drawbacks such as their expression, which depends on their genetic background and which is not usable for all genetics. Moreover, current inducer lines are more appropriate to temperate germplasm and there is still a need to develop inducer lines adapted to other agronomical regions. Besides, multiplication of inducer lines is critical as, due to poor pollen efficiency of these lines, pollination by foreign material is frequent. As a consequence, seed lots for inducer lines or hybrids need to be regularly controlled for purity. As for both applications, breeding of new lines and seed multiplication of available inducer lines, controls need to be done by an induction step and the sowing of thousands of seeds, the use of informative molecular markers should drastically reduce the time and means needed. As a consequence, the currently used methods have the disadvantage of being long and expensive in equipment. The identification of new reliable markers for an easier, faster and cheaper method is needed.

The inventors of the present invention identified a gene unexpectedly homologous to the acyl transferase/acyl hydrolase/lysophospholipase family as being responsible for haploid induction in maize, as well as molecular markers thereof, opening the way to the understanding and improvement of the induction mechanism in maize.

Described herein are polynucleotides and related processes (methods) directed to haploid induction in maize plants. Said polynucleotides are novel polynucleotides that induce haploid maize plants. Specifically, described herein are polynucleotides having the nucleic acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 3, and fragments thereof. SEQ ID NO: 1 corresponds to the gDNA sequence of the candidate gene responsible of haploid induction in maize plants that has been identified by the inventors and being homologous to the acyl transferase/acyl hydrolase/lysophospholipase family. The sequence identified by the inventors has been mutated by the insertion of the 4 bp CGAG in exon 4 of the gene in comparison with the wild type sequence of the gene which is found in maize. SEQ ID NO: 3 corresponds to the cDNA sequence of the candidate gene of SEQ ID N°1.

The present invention thus relates to an isolated polynucleotide responsible of haploid induction in plants comprising or consisting of:

(a) a nucleotide sequence which comprises or consists in SEQ ID N°1 or 3;

(b) a nucleotide sequence at least 80% identical to the nucleotide sequence of (a) and which is responsible of haploid induction in maize plants ; or

(c) a complementary sequence of the nucleotide sequence of (a) or (b), wherein the complementary sequence and the nucleotide sequence consist of the same number of nucleotides and are 100% complementary.

By “responsible of haploid induction in maize plants” is meant the *in situ* formation of haploid offspring at a significantly higher frequency than what is observed in natural populations of maize.

By “haploid” is meant a character related to cells or plants or parts of plants comprising such cells whose chromosomes contained in their nucleus are each in only one copy (n).

Such polynucleotides can easily be obtained by the man skilled in the art. The polynucleotides according to the invention can for example be obtained by cloning the candidate gene or cDNA using suitable primers. The polynucleotide can then be cloned into a vector, preferably into an expression vector.

"Isolated polynucleotide" refers herein to both RNA and DNA, including cDNA, genomic DNA, and synthetic DNA. Polynucleotides can have any three-dimensional structure. A polynucleotide can be double-stranded or single-stranded (*i.e.*, a sense strand or an antisense strand). Non-limiting examples of polynucleotides include genes, gene fragments, exons, introns, messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA, siRNA, micro-RNA, ribozymes, cDNA, recombinant polynucleotides, and branched polynucleotides. A polynucleotide may contain unconventional or modified nucleotides. Isolated polynucleotides according to the invention may be purified or recombinant.

The polynucleotides according to the invention may be of any length, *e.g.* at least, at most and/or about 20, 25, 30, 35, 50, 100, 250, 500 or 1000 nucleotides long.

“gDNA” or genomic deoxyribonucleic acid is chromosomal DNA, in contrast to extrachromosomal DNAs like plasmids, genomic DNA represent the polynucleotide as present in the cell comprising exons and intron sequences.

“cDNA” or complementary DNA is DNA synthesized from a messenger RNA (mRNA) template in a reaction catalyzed by the enzyme reverse transcriptase.

The term “isolated” in reference to a biological component (such as a nucleic acid, a vector or a protein) refers to a biological component that has been substantially separated or purified away from other biological components of the cell of the organism, or the organism itself, in which the component naturally occurs, such as other chromosomal and extra-chromosomal DNA and RNA, proteins, cells, and organelles. “Isolated nucleic acids” or “isolated vectors” include nucleic acid molecules purified by standard purification methods. These terms also encompass nucleic acids and vectors prepared by amplification and/or cloning, as well as chemically synthesized nucleic acids and vectors.

By “a nucleotide sequence at least 80% identical to the nucleotide sequence of (a)” is meant in particular, a nucleotide sequence 81, 82, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99% identical to the nucleotide sequence of (a). For example, a nucleotide sequence 95% "identical" to a query sequence of the present invention, is intended to mean that the sequence of the polynucleotide is identical to the query sequence except that the sequence may include up to five nucleotide alterations per each 100 nucleotides of the query sequence. In other words, to obtain a polynucleotide having a sequence at least 95% identical to a query sequence, up to 5% (5 of 100) of the nucleotides of the sequence may be inserted, deleted, or substituted with another nucleotide. In other terms, the sequences should be compared on their entire length (i.e. by preparing a global alignment). For example, a first polynucleotide of 100 nt (nucleotides) that is comprised within a second polynucleotide of 200 nt is 50% identical to said second polynucleotide. The needle program, which uses the Needleman-Wunsch global alignment algorithm (Needleman and Wunsch, 1970, A general method applicable to the search for similarities in the amino acid sequence of two proteins, J. Mol. Biol. 48:443-453) to find the optimum alignment (including gaps) of two sequences when considering their entire length, may for example be used. Preferably, the percentage of identity in accordance with the invention is calculated using the needle program with a “Gap open” parameter equal to 10.0, a “Gap Extend” parameter equal to 0.5, and a Blosum 62 matrix. The needle program is for example available on the ebi.ac.uk World Wide Web site.

In one embodiment, the nucleotide sequence at least 80% identical to the nucleotide sequence of (a) and which is responsible of haploid induction in maize plants may be a genomic sequence of maize origin comprising some allelic variations. Allelic variations may be nucleotide differences without consequence to the transcribed amino acid sequence or with consequences to the amino acid sequence but with a conservation of the protein function in maize haploid induction.

A “complementary sequence” as used herein refers to a sequence that specifically hybridizes in solution, e.g., according to Watson-Crick base pairing rules.

By “fragment” is meant a number of nucleotides sufficient to provide for a specific hybridization to the nucleic acid comprising or consisting of the complementary sequence of one of the SEQ ID N°s 1 and 3. In particular, the fragment of nucleic acid comprises at least 20 nucleotides, more particularly at least 150 nucleotides.

The present invention also relates to a vector comprising the recombinant DNA construct according to the invention.

The vector is preferably an isolated vector.

A "vector" may be such a construct that includes a replication system and sequences that are capable of transcription and translation of a polypeptide-encoding sequence in a given host cell. A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described in, e.g., Pouwels et al., Cloning Vectors: A Laboratory Manual, 1985, supp. 1987; Weissbach and Weissbach, Methods for Plant Molecular Biology, Academic Press, 1989; and Flevin et al., Plant Molecular Biology Manual, Kluwer Academic Publishers, 1990. Typically, plant expression vectors include, for example, one or more cloned plant sequences under the transcriptional control of 5’ and 3’ regulatory sequences and a dominant selectable marker. 5’ and 3’ regulatory sequences comprise but are not limited to a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally or developmentally-regulated, or cell- or tissue-specific expression), transcription initiation start site, ribosome binding site, RNA processing signal, transcription termination site, and polyadenylation signal.

By “plasmid”, it is herein meant a double-stranded circular DNA. The plasmid may include a marker gene enabling to select the cells comprising said plasmid, an origin of replication to allow the cell to replicate the plasmid and/or a multiple cloning site allowing the insertion of a DNA fragment, in particular the polynucleotide according to the invention.

In particular, said plasmid is chosen from L1457, L1465, L1478, L1482, L1479, L1483, L1542, L1543, L1540, L1541.

In particular, the vectors according to the invention can be chosen from vectors pSB11 (*Ishida et al.,* (1996).), pBIOS898 (Biogemma, for sense construct), pBIOS895 (Biogemma, for RNAi construct), pBb7m34GW, pBb7m42GW7 (Karimi *et al.*, 2012), pDONR221 (Invitrogen).

The vector preferably comprises an expression cassette of the polynucleotide according to the invention, i.e. a nucleic acid corresponding to the polynucleotide according to the invention placed under the control of at least one expression signal allowing its expression.

The expression signal is particularly selected among a promoter, a terminator, an enhancer and their combinations.

Suitable promoters, terminators and enhancers are well-known by the skilled person.

In particular, the promoter may be a constitutive promoter selected in the group consisting of the rice actin promoter (Act1) and intron (McElroy et al., 1990, Plant Cell, 2 :163-171), the ubiquitin promoter of maize (Christensen et al., 1996, Transgenic. Res., 5 :213) or the CsVMV promoter (Verdaguer et al., 1998, Plant Mol Biol. 6:1129-39) and FAD$_2$ intron (patent application WO 2006/003186 from Biogemma). Other examples of constitutive promoters useful for expression include the 35S promoter, the 19S promoter (Kay et al., 1987, Science, 236:1299-1302), or the pCRV promoter (Depigny-This et al., 1992, Plant Molecular Biology, 20 :467-479).

In a preferred embodiment, said construct is under the control of a constitutive promoter. In a most preferred embodiment, said construct is an RNAi construct, under the control of a constitutive promoter. Other suitable promoters could be used. It could be a tissue-specific promoter, for example a pollen-specific promoter, a promoter active in both anther and pollen, or an inducible promoter.

For example, the terminator may be selected in the group consisting of the Nos terminator corresponding to the region in the non coding 3' region of the nopaline synthase gene of the Ti-plasmid of *Agrobacterium tumefaciens* nopaline strain (Depicker et al. 1992) or the AtSac66 terminator. Others possible terminators are the polyA 35S terminator of the cauliflower mosaic virus (CaMV), described in the article of Franck *et al.,* (1980) or the histone terminator (EP 0 633 317).

Any other element like introns, enhancers, transit peptides, etc... may be comprised in the expression cassette. Introns and enhancers may be used to improve the expression of the gene according to the present invention.

Among useful introns, the first intron of maize adh1S can be placed between the promoter and the coding sequence. This intron when included in a gene construct increased the expression of the desired protein in maize cells (Callis *et al.,* 1987). One also can use the 1st intron of the shrunken 1 gene of the maize (Maas *et al.,* 1991), the 1st intron of the catalase gene of the bean catalase (CAT-1) (Ohta *et al.,* 1990), the 2nd intron of the ST-LS1 gene of potato (Vancanneyt *et al.* 1990), the DSV intron of the yellow dwarf virus of tobacco (Morris *et al.,* 1992), the actin-1 intron (act-1) of rice (McElroy *et al.,* 1990) and intron 1 of triosephosphate isomerase (TPI) (Snowdon *et al.,* 1996).

The expression cassettes may additionally contain 5' leader sequences. Such leader sequences can act to enhance translation. Such 5' leaders are known in the art and include, but are not limited to, picornavirus leaders, for example, the EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein, Fuerest, and Moss B., 1989) ; potyvirus leaders, for example, the TEV leader (Tobacco etch Virus) (Allison et al., 1986) ; the human immunoglobulin heavy-chain binding protein leader (BiP) (Macejack and Sarnow, 1991) ; the untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4) (Jobling and Gehrke, 1987) ; the tobacco mosaic virus leader (TMV) (Gallie et al., 1989) ; and the maize chlorotic mottle virus leader (MCMV) (Lommel *et al.,* 1991). See also, Della-Cioppa et al. (1987). Other methods known to enhance translation can be utilized, for example introns, and the like.

In one embodiment of the invention, the vector is a pSB11 vector with a Basta resistance cassette comprising an Actin promoter and intron, a Bar gene and a Nos terminator, a GFP cassette comprising a CsVMV promoter and FAD_2 intron, a GFP gene and a Nos terminator and a nucleotide sequence which comprises or consists in SEQ ID N°1 or 3; a nucleotide sequence at least 80% identical to the nucleotide sequence of (a) and which is responsible of haploid induction in maize plants as herein defined; or a complementary sequence of the nucleotide sequence of (a) or (b), wherein the complementary sequence and the nucleotide sequence consist of the same number of nucleotides and are 100% complementary, in particular a nucleotide sequence which comprises or consists in SEQ ID N°1 or 3 as herein defined, more particularly under the control of the rice actin promoter and first intron and followed by the AtSac66 terminator.

Isolated polynucleotides of the invention can be incorporated into recombinant DNA constructs capable of introduction into and replication in a host cell.

The present invention thus relates to a recombinant DNA construct comprising the polynucleotide according to the invention operably linked to at least one regulatory sequence.

The terms "recombinant construct," "expression cassette," "expression construct," "chimeric construct," "construct," "recombinant DNA construct" and "recombinant DNA fragment" are used interchangeably herein and are nucleic acid fragments. A recombinant construct comprises an artificial combination of nucleic acid fragments, including, and not limited to, regulatory and coding sequences that are not found together in nature. For example, a recombinant DNA construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source and arranged in a manner different than that found in nature. Such construct may be used by itself or may be used in conjunction with a vector as herein defined. If a vector is used then the choice of vector is dependent upon the method that will be used to transform host cells as is well known to those skilled in the art. For example, a plasmid vector, as described above, can be used. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells comprising any of the isolated polynucleotide according to the invention.

The term "recombinant DNA construct" refers to a DNA construct assembled from nucleic acid fragments obtained from different sources. The types and origins of the nucleic acid fragments may be very diverse.

The expression cassette will include 5’ and 3’ regulatory sequences operably linked to a polynucleotide according to the invention. "Operably linked" is intended to mean a functional linkage between two or more elements. "Regulatory sequences" refer to nucleotides located upstream (5’ non-coding sequences), within, or down-stream (3’ non-coding sequences) of a coding sequence, and which may influence the transcription, RNA processing, stability, or translation of the associated coding sequence. Regulatory sequences may include, and are not limited to, promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

In particular, said regulatory sequence (a promoter, for example) is responsible for an expression of the polynucleotide according to the invention.

The expression cassette may additionally contain selectable marker genes.

The expression cassette is further described as previously mentioned above.

The regulatory regions (i.e., promoters, transcriptional regulatory regions, and translational termination regions) and/or the polynucleotide according to the invention may be native/analogous to the host cell or to each other. Alternatively, the regulatory regions and/or the polynucleotide according to the invention may be heterologous to the host cell or to each other. As used herein, "heterologous" in reference to a sequence is a sequence that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous polynucleotide is from a species different from the species from which the polynucleotide was derived, or, if from the same/analogous species, one or both are substantially modified from their original form and/or genomic locus, or the promoter is not the native promoter for the operably linked polynucleotide.

A number of promoters can be used, as mentioned above. In particular, as previously mentioned, the promoter may be a constitutive promoter selected in the group consisting of the rice actin promoter (Act1) and intron, the maize ubiquitin promoter or the CsVMV promoter, the 35S promoter, the 19S promoter, the pCRV promoter and FAD_2 intron.

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

In particular, the DNA construct according to the invention comprises a polynucleotide which comprises or consists in a nucleotide sequence of SEQ ID N°1 or 3 operably linked to a promoter, in particular to the Actin promoter and intron.

As previously mentioned, the gene identified by the inventors as being responsible for haploid induction in maize is a mutant form of the wild type gene. The consequence of said mutation is a frame shift leading to 20 non conserved amino acids followed by a premature STOP codon downstream of the mutation. This would mean that the modification of function of the wild type gene of SEQ ID N°2 or 4 found in maize would induce haploid maize plants.

As a consequence, in the scope of the present invention, is also a recombinant DNA construct, comprising:

(a) a fragment of the polynucleotide according to the invention or of a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4 ; and

(b) the complementary sequence thereof;

said fragment and complementary sequence thereof being transcribed in an hairpin RNA to induce RNA interference.

By “fragment” is meant a number of nucleotides sufficient to provide for a specific hybridization to the nucleic acid comprising or consisting of the complementary sequence of one of the SEQ ID N°s 1 to 4. In particular, the fragment of nucleic acid comprises at least 20 nucleotides, more particularly at least 150 nucleotides.

"Stringent conditions" can be easily defined by the man skilled in the art using common knowledge. If necessary, guidance for defining such conditions can be found in numerous textbooks, such as Molecular Cloning: A Laboratory Manual by Sambrook et al., 2000. In particular, stringent conditions according to the invention can be constituted using a hybridization reaction requiring an optimized combination of hybridization buffer and hybridization temperature, depending on the ingredients of the hybridization buffer. Such determination of hybridization conditions falls within the routine work of the person skilled in the art.

As mentioned above, SEQ ID N°2 represents the wild type gene sequence (gDNA sequence) found in maize corresponding to the mutated gene of SEQ ID N°1 identified by the inventors as being responsible for haploid induction in maize. SEQ ID N°4 corresponds to the cDNA sequence of said wild type gene.

By “hairpin RNA” is meant a sequence of RNA that makes a tight hairpin turn that can be used to silence target gene expression, in the context of the invention the wild type gene of SEQ ID N°2, via RNA interference (RNAi). Expression of hairpin RNA in cells is typically accomplished by delivery of plasmids or through viral or bacterial vectors as herein mentioned.

By “RNA interference” is meant the biological process in which RNA molecules inhibit gene expression, typically endogenous gene expression by causing the destruction of specific mRNA molecules.

The polynucleotide according to the invention can be expressed as a transgene in order to make haploid inducer maize plants by overexpression of the SEQ ID N°1 or 3 or by overexpression of sequence SEQ ID N°2 or 4 by a process known as co -suppression . The DNA construct able to induce RNA interference can also be expressed in order to inhibit the expression of the endogenous wild type gene and lead to haploid inducer maize plants.

The present invention thus also relates to a process for transforming a host cell, comprising transforming the host cell with the recombinant DNA construct as defined above.

It also relates to a haploid inducer maize plant or seed comprising the recombinant DNA construct as defined above.

As used herein, the term "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 (including but not limited to embryos, pollen, ovules, seeds, leaves, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, and the like), plant tissues, plant cells, plant protoplasts, plant cell tissue cultures from which maize plant can be regenerated, plant callus, plant clumps, and plant seeds. A plant cell is a cell of a plant, either taken directly from a seed or plant, or derived through culture from a cell taken from a plant.

Said process and said maize plant transformation involve, and are not limited to, introducing a DNA construct into a plant or host cell. "Introducing" is intended to mean presenting to the plant the DNA construct. Processes for introducing polynucleotides into plants are known in the art including, and not limited to, stable transformation methods and transient transformation methods.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms.

"Host cell" refers the cell into which transformation of the DNA construct takes place and may include a yeast cell, a bacterial cell, and a plant cell. Examples of methods of plant transformation include Agrobacterium-mediated transformation (De Blaere et al., 1987, Meth. Enzymol. 143:277), particle-accelerated or "gene gun" transformation technology (Klein et al., 1987, Nature (London) 327:70-73; U.S. Patent No. 4,945,050) and protoplast transformation by electroporation or the use of chemicals such as PEG, among others.

"Stable transformation" is intended to mean that the polynucleotide introduced into a plant integrates into the genome of the plant and is capable of being inherited by the progeny thereof. "Transient transformation" or "transient expression" is intended to mean that the polynucleotide is introduced into the plant and does not integrate into the genome of the plant.

Suitable methods of introducing polynucleotides into plant cells are known by the man skilled in the art and include *Agrobacterium*- mediated transformation (U.S. Patent Nos. 5,563,055 and 5,981,840, and Ishida *et al.*, (2007). Agrobacterium-mediated transformation of maize. Nat Protoc *2*, 1614–1621), particle bombardment (Gordon-Kamm *et al.,* 1990, Plant Cell 2:603-618) and electroporation of protoplasts (Rhodes *et al.,* 1988, Science 240:204-207).

Methods are known in the art for the targeted insertion of a polynucleotide at a specific location in the plant genome. In one embodiment, the insertion of the polynucleotide at a desired genomic location is achieved using a site-specific recombination system. See, for example, WO99/25821, WO99/25854, WO99/25840, WO99/25855, and WO99/25853. Briefly, the polynucleotide according to the invention can be contained in a transfer cassette flanked by two non-identical recombination sites. An appropriate recombinase is needed for the integration of the transfer cassette fragment at the target site. Others methods without recombinase have also been described in WO 90/11354 but also in WO 96/14408. Said methods start from the observation that Double Strand Break (DSB) enhances the probability of homologous recombination at a given position. Said methods have been improved by the development of Meganucleases, Zn finger nuclease and others tools like TALENs and CRISPR/Cas9 system that have been developed to obtain targeted DSB.

These tools can be used for “gene editing”. The terms “gene editing” cover targeted mutations, theses mutations can be random mutation or directed mutations, by targeted is intend that the localization of the mutation is chosen. In particular, by the previously described homologous recombination system, a nucleotide exchange can be done to induce targeted mutations. Others methods for “gene editing” include the use of a DSB and a repair template to induce a specific nucleotide exchange during DNA repair. The CRISPR/Cas9 system is one of the specific methods of “gene editing” whereas the Cas9 protein and an ARN are used for obtaining a targeted DSB. Alternatively a simple DSB (double strand break) without repair template can be done on a targeted sequence to induce random mutations at this site. These mutations should be short insertions or deletions based on NHEJ (near Homologous End Joining) or MMEJ (microhomology mediated end joining).

The cells that have been transformed may be grown into plants in accordance with conventional ways. See, for example, McCormick et al. (1986) Plant Cell Reports 5:81-84.

The present invention thus also relates to process for producing haploid inducer maize plant or enhancing haploid induction ability of a maize plant comprising:

(a) transforming a maize plant cell with the recombinant DNA construct as described herein;

(b) regenerating a maize plant from the transformed maize cell; and

(c) growing the transformed maize plant under conditions that are suitable for expression of the recombinant DNA construct, wherein expression of the recombinant DNA construct results in haploid induction or enhancement of haploid induction by pollen of the transformed maize plant.

Methods that combine transformation and regeneration of stably transformed plants are well known. For example, to regenerate a maize plant, transformed cells are successively cultured on shooting and rooting media prior to transfer to soil (Ishida *et al.*, 2007, Agrobacterium-mediated transformation of maize. Nat Protoc *2*, 1614–1621).

In one embodiment, the present invention also relates to a process for inducing haploid maize plant lines comprising:

(a) growing transformed plants by the process for producing haploid inducer maize plant or enhancing haploid induction ability of a maize plant described above;

(b) using said plants as pollinators during the crossing with a female plant; and

(c) screening the progeny of the cross for haploid plants.

By “pollinators” is meant the plant that is a source of pollen for the pollination process.

By “screening the progeny of the cross”, it is meant determining the percentage of haploid and diploids plants in the progeny. Genetic markers can be used for phenotyping such as for example a dominant color marker expressed in the embryo and endosperm like R1-nj (navajo) and which should be homozygous in the inducing line. When this inducer line is used with r1/r1 female plants, haploid kernels will have purple endosperm crowns and colorless embryos, diploid kernels will have purple endosperm crowns and purple embryos (Weber 2014). The screening can also be done in particular by using the molecular markers according to the invention, which are described below. “Screening the progeny of the cross”, in the context of the invention also means selecting by screening of the haploid seeds, these haploid seeds being possibly further grown to produce homozygous fertile lines after their chromosome set is doubled spontaneously or by using specific protocols.

The inventors of the present invention also show, the spatial expression pattern of the gene GRMZM2G471240 by fusion of the promoters of this gene from line PK6 (SEQ

ID N°1) or line B73 (SEQ N°23) with a GUS reporter gene. Based on the blue staining indicative of Gus activity, the gene has a strict gametophytic expression in pollen and in pollen tube and is not expressed in the sporophytic tissues of the anther.

The invention also concerns a nucleotide sequence which comprises or consists in:

(a) SEQ ID No. 50 (2657 bp promoter region of GRMZM2G471240 from line B73, SEQ ID N°23) or SEQ ID N°51 (2534 bp promoter region of the GRMZM2G471240 from line PK6, SEQ ID N°1),

(b) a nucleotide sequence at least 75% identical to the nucleotide sequence SEQ ID No. 50 or 51, or

(c) a fragment of (a) or (b) which is able to induce the specific expression of associated sequence in the pollen and/or pollen tube . By associated sequence is intend, sequence operably linked in a recombinant construct wherein the above nucleotide sequence is used as a promoter regulatory sequence and associated sequence comprises but is not limited to, the coding sequence, sequence complementary to the coding sequence, construct for RNA interference.

Pollen specific expression of a cytotoxic enzyme like barnase for example can lead to male sterile plants and are of interest for hybrid seed production.

The present invention further relates to a process for screening a mutant plant population or mutant library, for identification of a mutant plant for enhanced haploid induction, mutant termed “haploid inducer plant". Said process comprises the use of a polynucleotide according to the invention or of a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4. In particular, these sequences should be used for identifying an orthologous sequence.

In one embodiment, the invention thus relates to a process for identification of a mutant haploid inducer plant, said process comprising the step of identifying an orthologous sequence of the sequence consisting in:

(a) a nucleotide sequence which comprises or consists in SEQ ID N°1 or 3 or a nucleotide sequence at least 80% identical thereof and which is responsible of haploid induction in maize plants; a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4; and/or

(b) a polypeptide encoded by a polynucleotide of (a).

In particular, said process further comprises the steps of:

(2a) Screening a mutant library to identify a mutant of the orthologous sequence identified previous step (1); or

(2b) generating a mutant of the orthologous sequence identified in step (1) by using a gene editing methods.

More particularly, said step (2a) comprises the use of a sequence consisting in:

(i) the orthologous sequence identified in step (1); or

(ii) a fragment thereof.

Mutant of the orthologous sequence meant mutant of polynucleotidic sequence if orthologous sequence is a polynucleotidic sequence and polynucleotidic sequence coding for the polypeptidic sequence if the orthologous sequence is a polypeptidic sequence.

In another embodiment, the invention relates to a process for identification of a mutant haploid inducer plant, said process comprising :

(a) screening a mutant library to identify a mutant of the sequence consisting in:

(i) a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 ; and/or

(b) generating a mutant of the sequence of step (i) by using a gene editing methods.

In particular, process the step of screening comprises the use of a sequence consisting in:

(i) a polynucleotide according to the invention (i.e. (a) a nucleotide sequence which comprises or consists in SEQ ID N°1 or 3; a nucleotide sequence at least 80% identical to this nucleotide sequence and which is responsible of haploid induction in maize plants ; or a complementary sequence thereof, wherein the complementary sequence and the nucleotide sequence consist of the same number of nucleotides and are 100% complementary); or a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4; or

(ii) a fragment thereof.

In particular, this mutant plant population or mutant library is a mutant maize plant population, sorghum plant population, or plant mutant population from others species.

Still particularly, the isolated polynucleotide responsible of haploid induction in maize plants according to the invention and described herein is used to find orthologous sequences and these orthologous sequences are used for screening other plant mutant populations.

By “orthologous sequence” is intended a gene having the same or a similar function in the same or different species. In particular, in the case or different species is meant homologous sequences that are inferred to be descended from the same ancestral sequence separated by a speciation event: when a species diverges into two separate species, the copies of a single gene in the two resulting species are said to be orthologous. In that case, orthologous genes are genes in different species that originated by vertical descent from a single gene of the last common ancestor.

When the sequence is contemplated in the same species, an orthologous sequence also meant homologous sequences that are inferred to be descended from the same ancestral sequence.

Orthologous sequences may be found by bioinformatics methods known by the man skilled in the art and the starting from the DNA sequence of the isolated polynucleotide responsible of haploid induction in maize plants according to the invention or of a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N° 2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N° 2 or 4 or from a polypeptide encoded by the above polynucleotides sequences (e.g., SEQ ID N°24 to SEQ ID N°25) or finally by a polypeptide sequence at least 80% identical to said polypeptidic sequences (e.g., at least 80% identical to SEQ ID N°24 or SEQ ID N°25), use of a phylogenetic tree is one of these methods,.

Examples of such orthologous sequences may be, but are not limited to, polypeptides sequences of *Hordeum vulgare* (SEQ ID N°28 or SEQ ID N°35), polypeptides sequences of *Brachypodium distachyon* (SEQ ID N°29 or SEQ ID N°36), polypeptides sequences of *Sorghum bicolor* (SEQ ID N°30 or SEQ ID N°37), polypeptides sequences of *Panicum virgatum* (SEQ ID N°31, SEQ ID N°32, SEQ ID N°38 or SEQ ID N°39), polypeptides sequences of *Setaria italica* (SEQ ID N°33 or SEQ ID N°40) or polypeptides sequences of *Oryza sativa* (SEQ ID N°34 or SEQ ID N°41).

In one embodiment, the invention thus relates to a process for screening a mutant plant population for enhanced haploid induction, said process comprising the use of a polypeptide comprising or consisting in the polypeptide sequence of SEQ ID N°28, SEQ ID N°35, SEQ ID N°29, SEQ ID N°36, SEQ ID N°30, SEQ ID N°37, SEQ ID N°31, SEQ ID N°32, SEQ ID N°38, SEQ ID N°39, SEQ ID N°33, SEQ ID N°40, SEQ ID N°34 or SEQ ID N°41 or polynucleotides coding for these sequences.

The invention is also about the use of recombinant DNA comprising the orthologous sequences described herein, for transforming a host cell and obtaining a transgenic plant. In a particular construction of the recombinant DNA may comprise (a) a fragment of the orthologous sequence described herein or a nucleotide sequence at least 80% identical thereof and (b) the complementary sequence thereof, said fragment and complementary sequence thereof being transcribed in an hairpin RNA to induce RNA interference. This recombinant DNA should be used for transforming a host cell and obtaining a transgenic plant. In this specific context, orthologous sequences can also be maize or sorghum sequences with homology to the sequence of the invention.

Screening of mutants within these sequences may allow to identify plants with a modified phenotype for haploid induction.

In particular, when said screening process is done on maize or sorghum, the polynucleotide according to the invention or the polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4 is used to screen for a mutation within said sequence. Mutations can be carried out by EMS, TILLING or eco-Tilling populations but also by T-DNA or transposon techniques. Still particularly, mutations can be obtained by editing methods in particular obtained by the use of meganucleases, Zinc finger nucleases, TALEN or CRISP/Cas9 methods and lead to a new interesting inducer line.

“Mutant plant population” or “mutant library” have here the same meaning and relate, in the context of the invention, to maize plant populations and other species of plant populations such as wheat, sunflower , The screening of these libraries can be done by the use of a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2, but also by the use of an orthologous sequence as described herein. The goal is, to found mutation(s) in SEQ ID N°2 or in a sequence having homology with SEQ ID N°2 in order to identify for mutants in maize population or mutation(s) in an orthologous sequence from others species such as wheat or sunflower.

By “homology” is meant the level of common nucleic acid or amino acid between two nucleic or amino acid sequence.

These screenings can be carried out by using full orthologous sequence has described herein, when the screening is based on sequencing methods, or with a fragment thereof, when the screening is based on PCR methods.

These methods are known by the man skill of the art.

“Fragment”, in the context of screening methods, relates to a number of nucleotides sufficient to provide a specific hybridization on the complementary sequence and allowing amplification by PCR methods. Said fragment may be at least of 8, 10, or 20 nucleotides long and can include the complementary strand of the fragment.

Mutant plant can also be obtained by “gene editing methods”, gene editing being proceeded on the SEQ ID N°2, or on nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or on orthologous sequence as previously obtained. If the “editing method” is based on CRISP/cas9 system, the short RNA molecule used should be homologous to the above targeted sequence.

As previously mentioned, the sequence identified by the inventors has been mutated by the insertion of the 4 bp CGAG in exon 4 of the gene in comparison with the wild type sequence of the gene which is found in maize and the consequence of said mutation is a frame shift leading to 20 non conserved amino acids followed by a premature STOP codon.

Also, in one embodiment of the above mentioned process for screening a mutant plant population for enhanced haploid induction, the polynucleotide according to the invention or the polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4 is used to screen for mutation within said exon 4 of said polynucleotide according to the invention, of said nucleotide sequence of SEQ ID N°2 or 4 or said nucleotide sequence at least 80% identical thereof, and more particularly for frame shift mutation within said polynucleotide according to the invention, of said nucleotide sequence of SEQ ID N°2 or 4 or said nucleotide sequence at least 80% identical thereof.

In another embodiment of the above mentioned process for screening a mutant plant population or mutant library for enhanced haploid induction, the polynucleotide according to the invention or the polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4 is used to screen for a mutation inducing modified expression of said polynucleotide according to the invention, of said polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of said nucleotide sequence at least 80% identical thereof. Indeed, a mutation on the control of the expression, in particular a mutation inducing inhibition of the transcription of said above mentioned sequences can also lead to a new interesting inducer line. For example, this mutation could be in the promoter sequence associated with the SEQ ID N°2 or in the promoter sequence of the orthologous sequence thereof.

The GRMZM2G471240 gene is predicted to code for a phospholipase (PL) and is also named *ZmPL*. Analysis of the deduced amino acid sequences of the wildtype $\mathrm{ZmPL_{HD99}}$ protein of SEQ ID N°25 revealed the presence of two S-palmitoylation sites at position 10 and position 423 as well as an S-farnesylation site at position 423. The second site (423) is missing in the truncated $\mathrm{ZmPL_{PK6}}$ protein of SEQ ID N°24. Together with the fact that the truncated $\mathrm{ZmPL_{PK6}}$ protein of SEQ ID N°24, is no longer localized in the cytoplasmic membrane, this suggests that the presence of a lipid anchor domain at the C-terminus of ZmPL $_{\text{HD99}}$ $_{\text{of}}$ SEQ ID N°25, may be essential for the correct subcellular localization of the protein.

A survey of the deduced amino acids from cereal orthologues showed that the proteins from *Brachypodium dystachion* (SEQ ID 29), *Sorghum bicolor* (SEQ ID 30), *Panicum virgatum* (SEQ ID 31 and SEQ ID 32), *Setaria italica* (SEQ ID 33) and *Oryza sativa* (SEQ ID 34) are all predicted to have lipid anchors at their N-terminus and C-terminus, whereas for *Hordeum Vulgare* (SEQ ID 28) only a N-terminal lipid anchor was predicted. The data suggest that the presence of an N-terminal and a C-terminal lipid anchor by S-palmitoylation, S-farnesylation and/or S-geranylgeranylation may be necessary for subcellular localization in the cytoplasmic membrane.

The presence of an N-terminal and C-terminal membrane anchor may be a useful additional criterion for the identification of functional orthologues in cereals and non-cereals, where phylogenetic analysis, even combined with the criterion of expression in the pollen, is often not sufficient.

Identification of an orthologous sequence as described herein may thus include a step of identification of a lipid anchor domain in C terminal and N terminal of the polypeptidic orthologue sequence as described herein and/or a step of identification of an ortholgoues sequence which is expressed in pollen and/or in the pollen tube of the plant.

In one embodiment, the orthologous sequenceas described therein is thus a polypeptidic sequence which comprises a lipid anchor domain in its C-terminal and N-terminal polypeptidic sequence and/or is a sequence which is expresses in pollen and/or pollen tube of the plant.

In another embodiment, the mutant of the orthologous sequence as described therein does not comprise a lipid anchor domain in its N-terminal polypeptide sequence..

For example, identification of a ZmPL functional orthologue from *Arabidopsis thaliana* by the use of a phylogenetic tree of all patatin-like phospholipases was constructed. Using preferential expression in pollen and the prediction of an N-terminal and a C-terminal lipid anchor as additional criteria, the gene At1g61850.1 was identified as the best candidate, and mutants N642695, N657713 and N596745 (obtained from NASC) are putative mutant haploid inducer plant identified by the method according to the invention.

Specific systems exist to create targeted mutations, for example the ZFN, meganuclease, TALEN or CAS9/CRISPR systems (Gaj *et al.*, 2013). They can be used to alter in a site-specific fashion the polynucleotide according to the invention or the polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4 or fragments thereof, and can be used to create new specific mutants for haploid induction ability. These systems can also be used to create targeted mutation in other species for example, wheat or sunflower by the use of these technologies on orthologous sequences.

In another embodiment, the present invention further relates to a process for determining the presence, the absence or the alteration of the polynucleotide according to the invention in a maize plant, comprising at least isolating nucleic acid molecules from said maize plant and amplifying sequences homologous to the polynucleotide according to the invention. Isolation step and amplification step can be carried out by methods well known by a man skilled in the art, such as for example those described in Dellaporta et al. 1985. Maize DNA miniprep, p. 36-37, in Molecular Biology of Plants. ,Malberg, J. Messing and I. Sussex, eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y; Dellaporta et al. 1983. A plant DNA minipreparation: Version II., Plant Molecular Biology Reporter Volume 1, Issue 4, pp 19-21; John G.K.Williams et al., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers., Nucleic Acids Research, Vol. 18, No. 22 6531.

As previously mentioned, the inventors have also identified new molecular markers. Said molecular markers allow to identify haploid inducer maize plants. This marker being specific of the allele responsible of haploid induction in maize, it opens the way for an easiest, fastest and cheapest method of control of such character.

The present invention thus also relates to a genetic marker of haploid induction in maize plants, wherein said genetic marker able to identify the polymorphism between the nucleotide sequences SEQ ID N°5 and SEQ ID N°42. In particular, said genetic marker is the genetic marker GRMZM2G471240_14. The genetic marker GRMZM2G471240_14 allows identification of the polymorphism as described in the tables 5a and 6 and is described in example 5. By “genetic marker” is meant a specific DNA sequence that can be identified within the genome of an individual and can be used to locate a particular locus of interest and/or determine the allelic origin of the plant at the locus. In the context of the invention, “genetic marker” can identify the locus responsible of haploid induction in maize, and determine the allelic variant at this loci. This DNA sequence can be a coding sequence. The marker can be read as dominant or co-dominant.

In the framework of the present invention, the genetic marker can detect plants carrying one or two alleles of the isolated polynucleotide responsible of haploid induction in maize plants according to the invention. The SEQ ID N°5 corresponds to the allelic sequence of the genetic marker carrying the 4 pb nucleotide insertion and responsible of haploid induction whereas the SEQ ID N°42 corresponds to the allelic sequence of the genetic marker without the insertion (wild type sequence). For proper inducer property the inducer line needs to be homozygous for this allele. The described genetic marker enables quick testing of seeds before sowing, seedlings or plants in development and retaining those with the desired characteristics, i.e. those which are haploid inducers.

This method could be used for example to follow the allele of the isolated polynucleotide responsible of haploid induction in maize plants according to the invention during the back cross of this allele in maize material for example exotic lines.

As a consequence, the present invention also relates to a process of identification of a haploid inducer maize plant comprising detecting in a maize plant or seed the genetic marker according to the invention.

The term "allele" refers to one of two or more different nucleotide sequences that occur at a specific locus. A first allele is found on one chromosome, while a second allele occurs at the same position on the homologue of that chromosome, e.g., as occurs for different chromosomes of a heterozygous individual, or between different homozygous or heterozygous individuals in a population. A "favorable allele" is the allele at a particular locus that confers, or contributes to, a desirable property, e.g., induction of haploid plant in maize.

As previously mentioned, the multiplication of the inducer line is critical since pollination by foreign material is frequent due to poor pollen efficiency of these lines. As a consequence, these lines need to be regularly controlled for purity. The use of the genetic marker according to the invention should drastically reduce the time and means needed to achieve it.

The present invention thus further relates to the use of the genetic marker according to the invention for quality control of seed lots in maize haploid inducer lines, comprising the steps of:

(a) taking a sample of seeds from a seed lot in a maize haploid inducer line;

(b) conducting molecular analyses to identify and quantify the presence of haploid inducer or non-inducer alleles;

(c) deducing from step b) the genetic purity value of the lot for the haploid inducer character.

The molecular analyses are well within the knowledge of the man skilled in the art. For example, the KASP method from KBioscience (LGC Group, Teddington, Middlesex, UK) can be used. The KASP™ genotyping system uses three target specific primers: two primers, each of them being specific of each allelic form and one extra primer on the reverse strand of DNA to achieve amplification, which is shared by both allelic forms. Each target specific primer also presents a tail sequence that corresponds to one of two FRET probes: one labelled with FAM® dye and the other with HEX® dye. After PCR reaction, the nature of the emitted fluorescence is used to identify the allelic form present in the mix of the studied DNA (He et al 2014).

The « genetic purity value » refers to the degree of contamination of a seed lot, in the context of the present invention, by non-inducers of haploid.

The invention will be further illustrated by the following figures and examples.

FIGURES

Figure 1: Observed segregation bias against PK6 allele in the subset of 531 recombinants using 16 SNP markers evenly distributed in the region.

Figure 2: Alignment of GRMZM2G471240 cDNA sequences obtained by canonical splicing from reference sequence B73 (SEQ ID N°23), and from genotypes HD99 (SEQ ID N°4) and PK6 (SEQ ID N°3). Among them was a 4 bp insertion in exon 4 of the GRMZM2G471240 candidate gene for PK 6.

Figure 3: Alignment of deduced amino acid sequences for gene GRMZM2G471240 from B73 (SEQ ID N°26 and SEQ ID N°27), HD99 (SEQ ID N°25) and PK6 (SEQ ID N°24). The consequence of the 4 bp insertion in exon 4 of the GRMZM2G471240 nucleic acid sequence of PK6 is a frame shift leading to 20 non conserved amino acids followed by a premature STOP codon. T01 and T02 (SEQ ID N°26 and SEQ ID N°27) correspond to alternative gene models proposed by the annotation of the B73 reference sequence, which differ in the length of exon 2 (see Fig. 1). All cDNA products cloned from pollen of genotypes PK6 and HD99 correspond to gene model T01.

Figure 4: Relative expression level by qRT-PCR with primers Pat_qRT_F1 and Pat_qRT_R1 (sequence SEQ ID N°6 and SEQ ID N°7) specific for amplification of gene GRMZM2G471240, arbitrary units. (Leaf_s : aerial parts of 5 DAS seedlings ; Leaf_b_j : leaf blade of juvenile leaf 3 ; Leaf_s_j : leaf sheath of juvenile leaf 3 ; Leaf_b_a : leaf blade of adult leaf 11 ; Leaf_s_a : leaf sheath of adult leaf 11 ; SAM : shoot apical meristem roughly dissected at 21 DAS ; Stem : stem section between leaf 9 and leaf 10 ; Root_s : roots of 5 DAS seedlings ; Root_a : roots adult at 38 DAS ; Tassel_imm : tassel immature at 45 DAS ; Tassel_mat : tassel mature with anthers and pollen ; Pollen : pollen ; Ear_imm : immature ear (2 cm) ; Ear_mat : mature ear ; Silk : silks emerged from husk leaves ; Kernel_12 : kernel 12 DAP ; Kernel_35 : kernel 35 DAP).

Figure 5: Pollen-specific expression of gene GRMZM2G471240

Promoter activity of gene GRMZM2G471240 was visualized in transgenic maize plants by histochemical detection of the GUS reporter. In mature anthers of hemizygous plants (A) blue GUS staining was found in about 50% of the pollen grains, whereas no GUS staining was observed in the anther. Observations of isolated pollen (B) and of wild type silk and ovule (C) after pollination with transgenic pollen revealed GUS activity in the pollen tube (arrow) and the fertilized embryo sac (arrow head).

Figure 6: Graphical representation of marker coverage and recombination breakpoints (determined by allelic sequencing) on the candidate gene GRMZM2G471240, for the two recombinant lines 71-19-29 and 71-16-132.

Figure 7: plasmid map for L1457

Figure 8: plasmid map for L1465

Figure 9: plasmid map for L1478

Figure 10: plasmid map for L1482

Figure 11: plasmid map for L1479

Figure 12: plasmid map for L1483

Figure 13: plasmid map for L1542 (p*ZmPL*$_{B73}$full::GUS::pmock3')

Figure 14: plasmid map for L1543 (p*ZmPL*$_{PK6}$full::GUS::pmock3')

Figure 15: Subcellular localization of ZmPL: Confocal imaging of Arabidopsis root tips expressing either truncated ZmPL$_{PK6}$ or wild-type ZmPL$_{HD99}$ fused to citrine fluorescent protein. Signal of both protein fusions can be seen in the cytosol whereas only wild-type ZmPL$_{HD99}$::citrine accumulated in the plasma membrane.

Figure 16: plasmid map for L1540 (pUBQ10::CDS-ZmPL$_{HD99}$::Citrine)

Figure 17: plasmid map for L1541 (pUBQ10::CDS-ZmPL$_{PK6}$::Citrine)

EXAMPLES

Example 1: map based cloning of the gene involved in haploid induction

Fine Mapping

For fine mapping of the *ggi1* locus, a previously identified QTL for haploid induction (Barret *et al.,* 2008), 96 highly-recombinant F2i3S2 plants (F2 intermated for 3 generations followed by 2 selfs) and 18 single seed descendants (selected for induction over 7 generations) of the HD99 x PK6 cross were genotyped with 10 markers in the umc1144/bnlg1811 interval and tested for haploid induction. Phenotyping was carried out by pollinating the hybrid F564 x HD7 homozygous for the recessive *glossy1* mutation with pollen of the plants of interest (all wildtype for the *GLOSSY1* locus) and counting the percentage of seedlings with a mutant *glossy1* phenotype (percentage of glossy) by visual inspection, assuming that the percentage of glossy reflected the percentage of haploid plants. The left border of the QTL interval was determined between markers umc1144 (pos 64.2 Mb) and AY110477 (pos 66.9) and the right border between markers GRMZM2G120587 (pos 68.1 Mb) and CL424968 (pos 70.6 Mb). N1078 (SSD) and

SMH37 (F2i3S2), the two inducing lines with the smallest PK6 region at the locus were backcrossed to HD99. After a selfing step a total of 10275 F2 have been generated and screened for recombination within the QTL interval.

First, 130 SNP markers located on chromosome 1 between pos. 65 and 70 Mbp (comprising 34 SNPs discovered by candidate gene re-sequencing) were evaluated for polymorphism between the parents using a KASP genotyping platform (http://www.lgcgenomics.com). From those, 26 polymorphic markers (Table 1) evenly distributed in the region were validated in a subset of 192 F2 plants comprising 96 derived from N1078 and 96 derived from SMH37.

QTL analysis has been reprocessed with these extra marker data on the 114 initial plants and results suggested a putative interval for the underlying determinant of the QTL between SYN24144 (67.72Mb) and PZE-101081844 (69.28Mb). Based on this information, three markers flanking the *ggi1* interval (GRMZM2G100497_10, GRMZM2G152877_6 and SYN35770) located on chromosome 1 at positions 65734188, 66958748 and 69889217 bp on the B73 reference map v2, respectively, were chosen to screen the 10275 F2 plants for recombination in the interval.

Table 1: SNP markers used in the genotyping of 114 initial recombinants (96 F2i3S2 + 18 SSD). The two SNP markers in italics mark the limit of the putative interval for the underlying determinant of the ggi1 QTL. "*" identifies markers used to screen the entire F2 population for recombinants.

KASP Markers id	**Chr1_bp (B73 RefMap V2)**	**Polymorphism**
GRMZM2G100497_10 *	65734188	A/G
GRMZM2G100497_9	65734040	T/C
GRMZM2G152877_16	66958261	G/T
GRMZM2G152877_6 *	66958748	G/T
SYN17701	67056433	A/C
PZE-101080848	67646656	A/C
GRMZM2G051879_48	67646723	/TTTGTTTTGCA
SYN24145	67727585	T/C
SYN24144	*67727790*	*T/C*
SYN24142	67727977	A/C
SYN25767	67850058	T/C
PZE-101081233	68179267	A/G

PZE-101081269	68241700	T/G
SYN6864	68437034	A/G
SYN6867	68437046	T/C
SYN20148	68555034	T/C
SYN20145	68557798	A/G
PZE-101081484	68558721	A/G
SYN25793	68670617	T/C
PZE-101081844	*69289243*	*T/G*
SYN2042	69587711	T/C
GRMZM2G117513_1	69888509	A/G
ASSAY954_00198	69887082	A/G
ASSAY954_00197	69887083	T/G
PUT-163a-74233607-3597	69887397	A/C
SYN35770 *	69889217	A/G

The screening of the 10275 F2 plants with SNP markers GRMZM2G100497_10, GRMZM2G152877_6 and SYN35770 identified 531 recombinant plants on the *ggi1* interval.

Fine mapping of the distortion bias trait

The PK6 allele is counter selected compared to a normal allele with a Mendelian segregation. Thus, progeny derived from a plant heterozygous at this locus is distorted and has less than the theoretical rate of 50 % of PK6 type alleles. Assuming that the same PK6 locus was responsible for haploid induction and distortion bias, the gene of haploid induction can be fine mapped based on the distortion score of genetic markers at this locus.

A subset of 48 recombinants plants, having a crossing over between GRMZM2G305400b_2 and SYN20148 (Table 2), have been selected and selfed. For each family a set of 48 derived seeds, or less if the selfing was not successful, have been sown and plantlets have been genotyped for the 16 SNP listed in Table 2. Analyses have been done by comparing for each marker the total number of homozygotes plants for Pk6 and HD99 alleles, respectively. The strongest bias was observed for the marker GRMZM2G471240_1 (Table 2 and Figure 1).

Table 2. : Observed segregation bias against PK6 allele for 48 recombinants in the subset of 531 recombinants using 16 SNP markers evenly distributed in the region. Column

Chr1_bp (V2) represents the physical position on the B73 reference genome v2. The segregation bias is the number of PK6 alleles divided by the number of HD99 alleles.

KASP Markers id	**Polymorphism**	**Chr1_bp (V2)**	**Segregation Bias (nb allele PK6/nb allele HD99)**
GRMZM2G100497_10	A/G	65734188	0.31
GRMZM2G152877_6	G/T	66958748	0.33
GRMZM2G051879_48	/TTTGTTTTGCA	67646723	0.33
SYN24142	A/C	67727977	0.33
SYN25767	T/C	67850058	0.33
GRMZM2G305400b_2	T/G	67993670	0.33
GRMZM2G120587_2	/GCA	68134724	0.25
PZE-101081233	A/G	68179267	0.18
GRMZM2G471240_1	TACG/	68241668	0.167
GRMZM2G003530_2	T/C	68437955	0.41
SYN20148	T/C	68555034	0.42
PZE-101081484	A/G	68558721	0.42
SYN25793	T/C	68670617	0.42
PZE-101081844	T/G	69289243	0.42
SYN2042	T/C	69587711	0.44
SYN35770	A/G	69889217	0.42

A set of 48 recombinants comprising 20 recombinants located within the interval with strong segregation bias against PK6 (within 68134724 and 68437955 on the B73 reference genome v2) and 28 extra recombinants in the flanking regions (within 67993670 and 68555034 on the B73 reference genome v2) were considered for further studies.

Analysis of the 48 recombinants at the locus

Amongst the 48 families selected to take forward, only 31 produced F2 seeds. For each one of those 31 recombinants producing seeds, 48 seeds (less in a few cases, if not available) were germinated and genotyped in order to select homozygous recombinant plants (HomoRec). Between 3 and 5 plants (preferentially homozygous recombinant plants) per recombinant family were further grown and self-pollinated.

For the initial screening of 31 families (up to 48 plants per family) only 4 markers were used: SYN25767, GRMZM2G305400b_2, SYN20148 and SYN25793 (see Table 2). The detected homozygous recombinants plants were further genotyped with a set of 78 markers.

Gene content of the ggi1 region

Preliminary analysis of the gene content in the maize B73 reference genome v2 highlighted 13 gene models (Table 3). Two of the putative gene models (GRMZM2G471240 and GRMZM2G062313) were expressed in anthers. Expression data have been obtained by interrogation of an eFP browser, Winter et al., 2007, Li et al., 2010 and Sekhon et al., 2011. Sequence analysis of the two gene models revealed the two corresponding genes are homologues from the Acyl transferase/acyl hydrolase/lysophospholipase family. The second one appeared to be a pseudogene and is then likely not expressed and the result obtained from the eFP browser should be an artefact due to the homology with the GRMZM2G471240 gene. This pseudogene may have arisen by duplication of the first one. The expression of GRMZM2G471240 in anthers (Table 3) made of it the best candidate gene which may be responsible for the ggi1 phenotype. Moreover, the marker GRMZM2G471240_1 developed on the gene GRMZM2G471240 exhibits the strongest PK6 allele segregation bias among the tested markers (Figure 1 and Table 2).

Table 3. : Genome position and expression data available for the 13 gene models identified within the ggi1 interval. The filtered gene set (FGS, solid evidence for gene) contains the 32540 genes of the maize genome published by Schnable et al., 2009. The working gene set (WGS) contains more than 30000 additional gene models, many of which are mere informatic predictions without further evidence. Expression data comes from interrogation of an eFP browser, Winter et al., 2007, Li et al., 2010 and Sekhon et al., 2011.

Id	Chr1_Pos(bp-V2)	Set	Expression data
GRMZM2G544129	68212317	WGS	not in eFP browser database
GRMZM2G544135	68215141	WGS	not in eFP browser database
GRMZM2G703616	68236616	WGS	Constitutive
GRMZM2G471240	**68240862**	**FGS**	**Anthers**
GRMZM2G062320	68318898	FGS	constitutive
GRMZM2G062313	**68323867**	**WGS**	**Anthers**
GRMZM2G062304	68365105	WGS	not in eFP browser database

AC213048.3_FG003	68398780	WGS	germinating seed
GRMZM2G520395	68404999	WGS	not in eFP browser database
AC213048.3_FG002	68409826	FGS	not in eFP browser database
GRMZM2G047877	68415868	WGS	not in eFP browser database
GRMZM2G047843	68419166	WGS	Constitutive
GRMZM2G510681	68428105	WGS	not in eFP browser database

GRMZM2G471240 specific expression is confirmed by qRT-PCR experiments in maize tissues

Since cross-hybridisation between closely related genes cannot be excluded in the micro-array data used to generate the eFP browser (Sekhon et al., 2011), gene specific primers Pat_qRT_F1 and Pat_qRT_R1 (see Table 9 for SEQ ID N°6 and SEQ ID N° 7), were used to do qRT-PCR on different maize tissues of genotype A188 (Figure 4).

Approximately 100 mg of fresh tissue was quick frozen in liquid nitrogen and ground to powder with mortar and pestle. Total RNA was extracted with 1 mL of Tri-reagent according to the instructions of the supplier (Invitrogen). After ethanol precipitation, the RNA was resuspended in 30 µL of RNase-free water and treated with RNase-free DNase. The DNase was inactivated according to the instructions of the supplier (Ambion). Approximately 5 µg of total RNA were reverse transcribed using random hexamers (Amersham Biosciences) and reverse transcriptase without RNaseH activity (Fermentas) in a final volume of 20 µL. A total of $2.5x10^5$ copies of GeneAmplimer pAW109 RNA (Applied Biosystems) were added to the RT reaction. The cDNA was diluted 50 times, and 2 µL was used in a volume of 20 µL containing 10 µL of the FastStart SYBR Green Master mix (Roche) on a StepOne Real-Time PCR System (Applied Biosystems). According to the manufacturer's protocol, the following program was used: 10 min at 95°C, followed by 40 cycles of 95°C for 10 sec and 60°C for 30 sec. Data were analysed using the StepOne Software v2.3 (Applied Biosystems). Expression levels were calculated using *Actin* as reference gene. The primers used are actin-q-F and actin-q-R respectively of SEQ ID N°8 and SEQ ID N°9 listed in Table 9.

The data confirmed the anther-specific expression suggested by the eFP browser and demonstrated in addition that the GRMZM2G471240 gene was (i) only expressed in mature tassels and not in immature tassels and (ii) expressed in pollen. The data did not allow to determine whether the GRMZM2G471240 gene was expressed only in pollen or also in other parts of mature tassels, for example the tapetum.

GRMZM2G471240 specific expression is confirmed by promoter fusion in transgenic maize plants.

The use of transgenic maize plants harboring fusions of the promoter of gene GRMZM2G471240 from genotype PK6 or B73 with the GUS reporter gene allowed to determine the spatial expression pattern of the gene during male reproduction. Based on the blue staining indicative of GUS activity, the gene has a strict gametophytic expression in pollen and pollen tube. It is not expressed in the sporophytic tissues of the anther (figure 5).

Materials and Methods

Promoter regions of 2657 bp (*ZmPL*$_{B73}$) corresponding to SEQ ID N°50 and of 2534 bp (*ZmPL*$_{PK6}$) corresponding to SEQ ID N°51 were amplified with primer pairs attB4_prom_PL_2576_B73 corresponding to SEQ ID N° 46 / attB1r_prom_PL_2576_B73 corresponding to SEQ ID N° 47 and attB4_promoPL_2534_PK6 corresponding to SEQ ID N° 48 / attB1r_promoPL_2534_PK6 corresponding to SEQ ID N°49, respectively, and introduced into pENTR P4-P1R (Invitrogen) by BP reaction. The prom *ZmPL*::*GUS* cassette in plasmids L1542 (p*ZmPL*$_{B73}$full::GUS::pmock3'), figure 13 and L1543 **(**p*ZmPL*$_{PK6}$full::GUS::pmock3') figure 14, used for maize transformation was obtained by triple LR reaction between these promoter fragments, the *GUS* gene of pEN-L1-SI-L2 (Karimi et al., 2007), pmock3' and the destination vector pB7m34GW (Karimi et al., 2005). GUS (beta-glucuronidase) histochemical staining of dissected organs of transgenic maize plants was performed by dipping tissues into the following solution: 1 mM X-Gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid), 0.05% TritonX100, 100 mM sodium phosphate (pH7), and 0.5 mM potassium ferrocyanure and 0.5 mM potassium ferricyanure. The enzymatic reaction was performed overnight at 37°C after vacuum infiltration.

SNP densification of the QTL

In order to densify the region, online available resources related to the HAPMAP2 project (Chia *et al.,* 2012) were used to develop 60 new SNP markers between 68180000 pb and 68420000 pb on the physical B73 reference genome v2. Among the 60 newly developed markers (Table 4), only 3 were polymorphic between HD99 and PK6, 18 markers were monomorphic and 39 revealed a presence/absence polymorphism (Presence/Absence Variant or PAV) without amplification in PK6. Even if the use of endpoint PCR protocol to detect absence of the locus is less accurate than use of quantitative PCR, thanks to the high density of markers and the high frequency of absence of amplification at this locus for the PK6 allele, it can be assumed that there is a deletion of at least 100 kb downstream of the GRMZM2G471240 candidate gene in PK6 compared to the reference sequences of genotype B73 and the parental line HD99. This data suggests that the pseudogene GRMZM2G062313 would be absent in PK6.

Table 4. : HAPMAP2 extra markers within the ggi1 interval and marker allele for HD99 and PK6. NA is for absence of the allele, marker type is defined between the lines HD99 and PK6.

KASP Markers id	**Chr1_bp (V2)**	**Marker type**	**HD99**	**PK6**
PZE0166358891	68180209	Monomorphic	G:G	G:G
PZE0166359231	68180549	Polymorphic	T:T	C:C
PZE0166359596	68180914	PAV	C:C	NA
PZE0166380893	68202211	PAV	G:G	NA
PZE0166393715	68215033	Monomorphic	A:A	A:A
PZE0166394286	68215604	Polymorphic	G:G	A:A
PZE0166394451	68215769	Monomorphic	T:T	T:T
PZE0166394686	68216004	PAV	C:C	NA
PZE0166394709	68216027	Monomorphic	A:A	A:A
PZE0166394976	68216294	PAV	G:G	NA
PZE0166407328	68228646	Polymorphic	T:T	G:G
PZE0166408365	68229683	PAV	A:A	A:A
PZE0166422350	68243668	PAV	C:C	NA
PZE0166435898	68257216	Monomorphic	G:G	G:G
PZE0166437099	68258417	PAV	A:A	NA
PZE0166480583	68301901	PAV	T:T	NA
PZE0166480678	68301996	PAV	C:C	NA
PZE0166480837	68302155	PAV	C:C	NA
PZE0166481099	68302417	PAV	G:G	NA
PZE0166481496	68302814	PAV	A:A	NA
PZE0166487364	68308682	PAV	G:G	NA
PZE0166488186	68309504	PAV	G:G	NA
PZE0166488255	68309573	PAV	G:G	NA
PZE0166497519	68318837	PAV	A:A	NA
PZE0166500667	68321985	Monomorphic	C:C	C:C
PZE0166502346	68323664	PAV	C:C	NA
PZE0166502392	68323710	PAV	C:C	NA
PZE0166502746	68324064	PAV	A:A	NA

PZE0166503514	68324832	PAV	C:C	NA
PZE0166504940	68326258	PAV	G:G	NA
PZE0166505171	68326489	PAV	C:C	NA
PZE0166505239	68326557	PAV	C:C	NA
PZE0166505408	68326726	PAV	C:C	NA
PZE0166505464	68326782	PAV	A:A	NA
PZE0166507774	68329092	PAV	A:A	NA
PZE0166507883	68329201	PAV	A:A	NA
PZE0166508182	68329500	PAV	A:A	NA
PZE0166508227	68329545	PAV	C:C	NA
PZE0166508604	68329922	PAV	G:G	NA
PZE0166508796	68330114	PAV	C:C	NA
PZE0166509054	68330372	PAV	T:T	NA
PZE0166540610	68361928	PAV	T:T	NA
PZE0166540778	68362096	PAV	C:C	NA
PZE0166540846	68362164	PAV	T:T	NA
PZE0166542974	68364292	PAV	G:G	NA
PZE0166544802	68366120	PAV	C:C	NA
PZE0166544973	68366291	Monomorphic	A:A	A:A
PZE0166546986	68368304	PAV	A:A	NA
PZE0166576167	68397485	PAV	T:T	NA
PZE0166578457	68399775	Monomorphic	G:G	G:G
PZE0166578493	68399811	Monomorphic	A:A	A:A
PZE0166586633	68407951	Monomorphic	G:G	G:G
PZE0166587074	68408392	Monomorphic	G:G	G:G
PZE0166588267	68409585	Monomorphic	C:C	C:C
PZE0166588497	68409815	Monomorphic	C:C	C:C
PZE0166588562	68409880	Monomorphic	G:G	G:G
PZE0166594311	68415629	Monomorphic	C:C	C:C
PZE0166594370	68415688	Monomorphic	G:G	G:G
PZE0166597249	68418567	Monomorphic	G:G	G:G
PZE0166598227	68419545	Monomorphic	G:G	G:G

Re-sequencing data of the candidate gene GRMZM2G471240 in the PK6 and HD99 parental lines allowed to develop 17 markers on the gene in addition to the previously developed marker GRMZM2G471240_1 (Table 5a). Moreover, a quantitative real time PCR marker named qPCR2313 was developed to follow the presence/absence of the pseudogene GRMZM2G062313 (not present in PK6).

Table 5a.: Molecular markers developed on GRMZM2G471240 by re-sequencing.

KASP Markers id	Chr1_bp (V2)	HD99	PK6
GRMZM2G471240_2	68241270	C:C	T:T
GRMZM2G471240_3	68241339	T:T	C:C
GRMZM2G471240_1	68241668	TACG:TACG	-
GRMZM2G471240_5	68241700	A:A	C:C
GRMZM2G471240_7	68241902	G:G	C:C
GRMZM2G471240_8	68241910	G:G	A:A
GRMZM2G471240_9	68242070	G:G	C:C
GRMZM2G471240_10	68242166	T:T	C:C
GRMZM2G471240_11	68242296	G:G	A:A
GRMZM2G471240_13	68242401	A:A	C:C
GRMZM2G471240_14	68242433-68242434	-	CGAG:CGAG
GRMZM2G471240_15	68242448	T:T	C:C
GRMZM2G471240_16	68242452	C:C	A:A
GRMZM2G471240_17	68242547	A:A	C:C
GRMZM2G471240_18	68242553	G:G	A:A
GRMZM2G471240_19	68242566	T:T	C:C
GRMZM2G471240_20	68242569	G:G	C:C
GRMZM2G471240_21	68242602	G:G	A:A

Table 5b.: Molecular markers developed on GRMZM2G382717 and GRMZM5G866758 by re-sequencing.

KASP Markers id	Chr1_bp (V2)	Polymorphism
GRMZM2G382717_1	68113455	G/T
GRMZM5G866758_1	68430654	C/A

Marker analysis on a diversity panel

A diversity panel was assembled, consisting of 127 lines containing 116 lines from diverse origins representative of the diversity in the maize lines gene-pool, and 11 inducer lines of various origin including PK6 and stock6. The panel was genotyped with a subset of 59 markers (Table 6) previously described (from HAPMAP2 project data and genes re-sequencing data, Table 5a and 5b) between positions 67850001 and 68430000.

Table 6.: SNP Marker set used on the diversity panel. Marker type refers to the polymorphism between line PK6 and HD99.

Marker id (KASP and real time PCR)	**Chr1_bp (V2)**	**Marker type**
SYN25767	67850058	SNP
GRMZM2G382717_1	68113455	SNP
PZE0166359596	68180914	PAV
PZE0166380893	68202211	PAV
PZE0166394686	68216004	PAV
PZE0166394976	68216294	PAV
PZE0166408365	68229683	PAV
GRMZM2G471240_2	68241270	SNP
GRMZM2G471240_3	68241339	SNP
GRMZM2G471240_7	68241902	SNP
GRMZM2G471240_8	68241910	SNP
GRMZM2G471240_9	68242070	SNP
GRMZM2G471240_10	68242166	SNP
GRMZM2G471240_11	68242296	SNP
GRMZM2G471240_13	68242401	SNP
GRMZM2G471240_14	68242433	INDEL
GRMZM2G471240_15	68242448	SNP
GRMZM2G471240_16	68242452	SNP
GRMZM2G471240_17	68242547	SNP
GRMZM2G471240_18	68242553	SNP
GRMZM2G471240_19	68242566	SNP
GRMZM2G471240_20	68242568	SNP
GRMZM2G471240_21	68242602	SNP
PZE0166422350	68243668	PAV
PZE0166437099	68258417	PAV
PZE0166480583	68301901	PAV

PZE0166480678	68301996	PAV
PZE0166480837	68302155	PAV
PZE0166481099	68302417	PAV
PZE0166481496	68302814	PAV
PZE0166487364	68308682	PAV
PZE0166488186	68309504	PAV
PZE0166488255	68309573	PAV
PZE0166497519	68318837	PAV
PZE0166502346	68323664	PAV
PZE0166502392	68323710	PAV
PZE0166502746	68324064	PAV
qPCR2313	68324620	PAV
PZE0166503514	68324832	PAV
PZE0166504940	68326258	PAV
PZE0166505171	68326489	PAV
PZE0166505239	68326557	PAV
PZE0166505408	68326726	PAV
PZE0166505464	68326782	PAV
PZE0166507774	68329092	PAV
PZE0166507883	68329201	PAV
PZE0166508182	68329500	PAV
PZE0166508227	68329545	PAV
PZE0166508604	68329922	PAV
PZE0166508796	68330114	PAV
PZE0166509054	68330372	PAV
PZE0166540610	68361928	PAV
PZE0166540778	68362096	PAV
PZE0166540846	68362164	PAV
PZE0166542974	68364292	PAV
PZE0166544802	68366120	PAV
PZE0166546986	68368304	PAV
PZE0166576167	68397485	PAV
GRMZM5G866758_1	68430654	SNP

Genotyping data obtained on the panel indicated that the large deletion downstream the GRMZM2G471240 candidate gene is present on 11.5% of the 116 maize lines, which are not haploid inducers, suggesting that this large deletion and the presence/absence of the pseudogene GRMZM2G062313 contained in it is not the causal factor of haploid induction.

Among the markers developed on the candidate gene GRMZM2G471240 and tested on the diversity panel, the marker GRMZM2G471240_14 (INS/DEL of 4 bp CGAG/) is the only one having an allele exclusively present on the PK6 genotype and the 10 other inducing lines, which could mean that this is the causal polymorphism of the ggi1 QTL. The insertion of 4 bp targeted by the marker GRMZM2G471240_14 is specific to the original inducing line stock6 and all its tested derivatives including PK6 and is absent in all the non-inducer lines tested. This insertion of 4 bp in the 4th intron of the GRMZM2G471240 gene PK6 causes a translation frameshift resulting in a shorter protein with the last 20 amino acids differing completely from the original protein sequence. None of the other polymorphisms tested were exclusive to PK6 or absent from all non-inducer lines.

Description of recombinants for the candidate gene GRMZM2G471240

Homozygous recombinant plants were obtained and genotyped for 28 of the 31 recombinant families. Two recombinants families appeared to have recombination breakpoints inside the candidate gene GRMZM2G471240 (see Figures 2 and 3). The recombinant family 71-19-29 has a recombination breakpoint in exon 4 upstream of the GRMZM2G471240_14 marker (insertion of four bp specific to inductors of haploids, see Figure 6). The recombinant family 71-16-132 has a recombination breakpoint in a region which encompasses exon 2 (see Figure 6). Other families exhibit recombination breakpoints on the right and on the left of the candidate gene GRMZM2G471240.

These two new recombinants are new alleles at the *ggi1* locus and have been evaluated for their inducing ability.

Phenotyping for haploid induction capacity of recombinants families

All the 28 homozygous recombinant families from the previous step have been evaluated for inducing ability together with two positive controls (PK6 and RWS), and five negative controls (HD99, F2, WPP112, Nys302 and EM1201).

Each line has been crossed with a tester (female line), and the resulting ear obtained from each cross has been evaluated for their rate of haploid and diploid kernels by counting the percentage of ligule-less or glossy seedlings amongst the progeny (Lashermes and beckert 1988, Neuffer, 1997; Prigge, 2011). The glossy test has been described within example 1. The ligule-less test involves a recessive trait that can be used as alternative to the glossy test as a visual marker for the identification of haploid plants. The induction tests were duplicated and carried out in Limagrain and INRA using the ligule-less and glossy systems respectively.

Both duplicates provided the same results. The phenotypic results of both tests show that all the plants that are PK6 at the marker GRMZM2G471240_14 (INS:INS) are clear haploid inducers and that the plants that are HD99 at the marker GRMZM2G471240_14 (DEL:DEL) are not haploid inducers.

<u>Example 2</u>: **comparison of the *GRMZM2G471240 alleles***

Sequencing of GRMZM2G471240 alleles

The genomic sequence of the GRMZM2G471240 candidate gene in genotype PK6 was determined by PCR amplification of overlapping fragments from reference sequence B73, which were arbitrarily named A, C and D, with primers G471240_A_F and G471240_A_R (sequences SEQ ID N°10 and SEQ ID N°11), G471240_C_F and G471240_C_R (sequences SEQ ID N°12 and SEQ ID N°13), and G471240_D_F and G471240_D_R3 (or G471240_B_R), respectively of sequences SEQ ID N°14, SEQ ID N°15 and SEQ ID N°16 (see Table 9 for primer sequences). PCR fragments obtained with a proof reading enzyme were either sequenced directly or after prior subcloning (SEQ ID N° 1).

The genomic sequence of the GRMZM2G471240 candidate gene in genotype HD99 was determined by PCR amplification of overlapping fragments, which were arbitrarily named A, C and B with primers G471240_A_F and G471240_A_R, G471240_C_F and G471240_C_R and G471240_B_F and G471240_B_R, respectively sequences SEQ ID N°10, SEQ ID N°11, SEQ ID N°12, SEQ ID N°13, SEQ ID N°17 and SEQ ID N°16 (see Table 9 for primer sequences). PCR fragments obtained with a proof reading enzyme were either sequenced directly or after prior subcloning (SEQ ID N° 2). An extraction of the corresponding region from the reference genome of genotype B73 (Schnable et al., 2009) is also presented in the sequence listing (SEQ ID N° 23).

Sequence alignment of alleles at the GRMZM2G471240 locus

Alignment with the reference sequence of genotype B73 (SEQ ID N° 23) revealed numerous polymorphisms between PK6, HD99 and B73. Among them was a 4 bp insertion in exon 4 of the GRMZM2G471240 candidate gene in genotype PK 6 (Figure 2). The consequence of the 4 bp insertion in exon 4 is a frame shift leading to 20 non conserved amino acids followed by a premature STOP codon (SEQ ID N° 24). This may be the causal mutation explaining the PK6 phenotype (Figure 3). GRMZM2G471240 is predicted to code for a phospholipase (PL) and will be named *ZmPL* for the further experimental genetic constructs. T01 and T02 correspond to alternative gene models proposed by the annotation of the B73 reference sequence, which differ in the length of exon 2 (SEQ ID N° 26, SEQ ID N°27; see Fig. 1). All cDNA products cloned from pollen of genotypes PK6 and HD99 correspond to gene model T01.

Subcellular localization of the ZmPL protein

The subcellular localization of the ZmPL protein from genotypes PK6 and HD99 was determined in a heterologous system by in frame fusions to citrine fluorescent protein, stable transformation of *Arabidopsis thaliana* and observation of transgenic root tips by confocal microscopy. Based on the fluorescence of the chimeric proteins the wildtype $ZmPL_{HD99}$ protein was mainly located in the cytoplasmic membrane but also present in the cytoplasm, whereas the truncated $ZmPL_{PK6}$ protein was absent from the cytoplasmic membrane and almost entirely located in the cytoplasm (figure 15).. This difference may be caused either by the loss of the 49 C-terminal amino acid residues in the $ZmPL_{PK6}$ protein or the presence of the 20 unrelated amino acid residues These results suggest that (i) the 3' end of the protein is important for its localization in the cytoplasmic membrane and (ii) that the mis-localization of the $ZmPL_{PK6}$ protein is the cause for the haploid inducing capacity of genotype PK6.

Materials and Methods: Arabidopsis ecotype Col-0 was transformed with the floral-dip method (Clough and Bent, 1998) and a modified procedure for *Agrobacterium* preparation (Logemann et al., 2006) ; *ZmPL* CDS (without STOP codon) from HD99 and PK6 were PCR amplified from cDNA made from mature anther tissues using respectively primers pair PL-CDS-F_Dtopo corresponding to SEQ ID N° 52 / PL-CDS-HD99-R2 corresponding to SEQ ID N° 53 and PL-CDS-F_Dtopo / PL-CDS-PK6-R2 corresponding to SEQ ID N° 54, and cloned into pENTR/D-topo (Invitrogen). The cassettes pUBQ10::ZmPL-CDS::CITRINE were obtained by LR reaction between the fragments pUBQ10, *ZmPL* CDS, mCITRINE and the destination vector pK7m34GW (Karimi et al., 2005). The fragments for pUBQ10 and mCITRINE were gifts of Yvon Jallais (Jaillais et al., 2011) plasmid L1540 (pUBQ10::CDS-$ZmPL_{HD99}$::Citrine) is described in figure 16 and plasmid L1541 (pUBQ10::CDS-$ZmPL_{PK6}$::Citrine) in figure 17.

.

Fluorescence was detected with a Zeiss LSM 710 Laser Scanning Microscope: mCITRINE was excited with a 510-nm laser signal and fluorescence was detected using 520–580 nm bandpass filters. Image data were analyzed by using Image J software.

<u>Example 3</u>: Creation of new inducer alleles by transgenesis

Transformation protocol

The plasmids used for the production of *ZmPL*$_{PK6}$-OE, *ZmPL*$_{HD99}$-OE and *ZmPL*$_{A188}$-RNAi plants contained the backbone of vector pSB11 (Ishida *et al.,* 1996), a Basta resistance cassette (*Oryza sativa* (rice) *Actin* promoter and intron, *Bar* gene and *Nos* terminator) next to the right border, a GFP cassette (*CsVMV* promoter and *FAD2* intron, *GFP* gene and *Nos* terminator) and either the *ZmPL* coding sequence (primers see Table 9) under the control of the constitutive rice *Actin* promoter and intron, or a unique gene fragment (primers see Table 9) separated by the rice *Tubulin* intron in hairpin configuration and followed by the *AtSac66* terminator. *Agrobacterium*-mediated transformation of maize inbred line A188 was executed according to a published protocol (Ishida *et al.*, 2007). For each transformation event the number of T-DNA insertions was evaluated by qPCR, and the integrity of the transgene was verified by PCR with primers situated in the *AtSac66* terminator near the end of the construct of interest next to the left border.

Inhibition of the PL gene by RNAi (L1465) ZmPL$_{A188}$-RNAi

The gene encoded by gene model GRMZM2G471240 was named PL (phospholipase) and the respective genotype indicated by an extension in subscript.

The PL$_{A188}$-RNAi construct tests whether (i) the inducing capacity of genotype PK6 was linked to a loss-of-function of the PL gene and whether a (ii) a knockdown of the PL gene could yield a stronger phenotype or other phenotypes than the 4 bp insertion in the PK6 allele.

For the RNAi construct primers G471240-attB1 and G471240-attB2 of SEQ ID N°18 and SEQ ID N°19 (see Table 9) were used to amplify an intronless fragment of 363 bp in exon 4 on genomic DNA of genotype A188, the genotype used for maize transformation. Contrary to the beginning of the *PL* gene, the chosen fragment shared no sequence homology with the related pseudogene GRMZM2G062313. The fragment was recombined (BP Gateway reaction) into the vector pDONR221 (Invitrogen) to yield the entry clone L1457 (Figure 7).

Subsequently the fragment was recombined (LR Gateway reaction) into the vector pBIOS 898 (gift of W. Paul, Biogemma). The resulting plasmid L1465 was used for maize transformation (Figure 8). Transgenic plants have been tested for their capacity to induce gynogenesis and (ii) examined for other phenotypes linked to pollen maturation and fertilisation.

The lines to be tested for haploid induction were crossed as male parent with hybrid F564 x DH7 homozygous for the glossy1 mutation. A minimum of 100 kernels was germinated and scored for the glossy (bright leaf surface, adhering water droplets) phenotype indicative of haploid plantlets. Haploid induction rate was determined as the percentage of glossy plantlets among germinated plantlets.

Hemizygous transgenic lines were either selfed or crossed with the glossy tester. At least 50 kernels were sown and sprayed with Basta herbicide (glufosinate–ammonium solution at 1.5 g/L) 10 days after germination. Segregation distortion was determined as the number of Basta resistant transgenic plantlets divided by the number of wild-type plantlets.

For five independent transformation events (U261 to U265) fertile plants with offspring were obtained. Event U265 provoked some haploid induction both using a heterozygous T1 plant (1 haploid plantlet among 197 T2 seedlings) and a homozygous T2 plant (1 haploid plantlet among 1000 T3 seedlings). It needs to be recalled that other loci than ggi1 influence haploid induction rate and that the genotype A188 used for maize transformation has an extremely low haploid induction rate (0 haploid plantlets among over 3000 seedlings tested) compared to other genotypes. Finally, event U264 showed a strong segregation bias (8.3% and 3.6% instead of the 50% transgenics expected) both in T1 and T2 seedlings and some haploid induction (1 haploid plantlet among 1177 T2 seedlings). All results for event U264 are based on heterozygous plants, since so far no confirmed homozygous plants have been obtained for this event.

Ectopic over expression (OE) of the PL_{PK6} allele (L1482)

For the PL_{PK6}–OE construct genomic DNA containing the entire coding sequence of genotype PK6 was amplified with primers PLPK6_HDPK_F1 and PLPK6_PK6_R1, of sequences SEQ ID N°20 and SEQ ID N°21 (see Table 9). The PCR product was recombined (BP Gateway reaction) into the vector pDONR221 (Invitrogen) to yield the entry clone L1478 (Figure 9). In this intermediary vector the coding sequence (CDS) of PL_{PK6} is flanked by attL sites (CDS is the genomic sequence from Start to Stop codon).

Subsequently the CDS of PL_{PK6} was recombined (LR Gateway reaction) into the vector pBIOS 895 (gift of W. Paul, Biogemma). The resulting plasmid L1482, in which the PL_{PK6} CDS is placed under the control of a rice *Actin* promoter and followed by a At*Sac66* terminator, was used for maize transformation (Figure 10). Transgenic plants have been tested for their capacity to induce gynogenesis and (ii) examined for other phenotypes linked to pollen maturation and fertilisation.

The lines to be tested for haploid induction were crossed as male parent with hybrid F564 x DH7 homozygous for the *glossy1* mutation. A minimum of 100 kernels was germinated and scored for the *glossy* (bright leaf surface, adhering water droplets) phenotype indicative of haploid plantlets. Haploid induction rate was determined as the percentage of *glossy* plantlets among germinated plantlets.

Hemizygous transgenic lines were either selfed or crossed with the *glossy* tester. At least 50 kernels were sown and sprayed with Basta herbicide (glufosinate–ammonium solution at 1.5 g/L) 10 days after germination. Segregation distortion was determined as the number of Basta resistant transgenic plantlets divided by the number of wild-type plantlets.

Ectopic over expression of the PL$_{HD99}$ allele (L1483), complementation of PK6 allele

In genotype PK6 a 4 bp insertion in the *PL* gene causes a truncation of the predicted protein. If this truncated protein was not functional at all, then the introduction of the non-inducing wild type allele PL$_{HD99}$ should complement the mutation and block the induction of gynogenesis.

For the PL$_{HD99}$–OE construct genomic DNA containing the entire coding sequence of genotype HD99 was amplified with primers PLPK6_HDPK_F1 and PLPK6_HD99_R1 of sequences SEQ ID N°20 and SEQ ID N°22 (Table 9). The PCR product was recombined (BP Gateway reaction) into the vector pDONR221 (Invitrogen) to yield the entry clone L1479 (Figure 11). In this intermediary vector the CDS (genomic sequence from Start to Stop codon) of PL$_{HD99}$ is flanked by attL sites.

Subsequently the CDS of PL$_{HD99}$ was recombined (LR Gateway reaction) into the vector pBIOS 895 (gift of W. Paul, Biogemma). The resulting plasmid L1483, in which the PL$_{HD99}$ CDS is placed under the control of a rice *Actin* promoter and followed by a At*Sac66* terminator, was used for maize transformation (Figure 12). Transgenic plants have been crossed with genotype PK6 and offspring have been tested for their capacity to induce gynogenesis.

The lines to be tested for haploid induction were crossed as male parent with hybrid F564 x DH7 homozygous for the *glossy1* mutation. A minimum of 100 kernels was germinated and scored for the *glossy* (bright leaf surface, adhering water droplets) phenotype indicative of haploid plantlets. Haploid induction rate was determined as the percentage of *glossy* plantlets among germinated plantlets.

Hemizygous transgenic lines were either selfed or crossed with the *glossy* tester. At least 50 kernels were sown and sprayed with Basta herbicide (glufosinate–ammonium solution at 1.5 g/L) 10 days after germination. Segregation distortion was determined as the number of Basta resistant transgenic plantlets divided by the number of wild-type plantlets.

Example 4: Creation of new inducer alleles by targeted transgenesis *Mutation of the* PL_{A188} *allele*

The CAS9/CRISPR system allows to create small deletions at nearly any site in the genome. The system have been used to create two distinct deletions: one at the beginning of the gene to obtain a true knockout (as compared to the knockdown by RNAi) and a second one to obtain an independent frameshift towards the end of the gene copying the 4 bp insertion in the PL_{PK6} allele. TALEN system has also been used to create new mutants (Gaj et al., 2013).

Example 5: Use of GRMZM2G471240_14 marker for control quality of seed lots.

Four seed stocks for inducer hybrids (RWS x RWK76) and five for inducer lines (RWS or RWK76) from different years and seasons of production have been tested for seed purity at the locus of *PL* gene, about 350 kernel by lot have been randomly selected from each lot, and tested with the GRMZM2G471240_14 marker by the KASP method (with the three primers of the following sequences: SEQ ID N°43: GAGGGCATCGGCATTGCTTCCTT (Common); SEQ ID N°44: GTCAACGTGGAGACAGGGAGC and SEQ ID N°45: GTCAACGTGGAGACAGGGAGG). The results are shown in Table 7 below.

Table 7.: wt means wild type and NA reflects missing data

	RWS x RWK76	RWS x RWK76	RWS x RWK76	RWS x RWK76	RWS	RWS	RWK76	RWK76	RWK76
Lot reference	2010 winter	2010 summer	2012 winter	2012 summer	2011 summer	2012 summer	2012 winter	2012 summer	2013 winter
Number of tested kernels	344	354	372	372	305	372	372	372	369
wt:wt (Homozygous wt)	9	10	0	0	0	0	0	0	0
pk6:wt (Heterozygous pk6/wt)	50	56	1	4	0	0	0	0	0
pk6:pk6 (homozygous for pk6)	284	282	371	368	305	368	367	368	360
NA	1	6	0	0	0	4	5	4	9

According to these results the two first lots show an insufficient purity to be further used to induce haploids. These findings corroborate the low induction rate of these lots that was below 5% when the normal rate is about 10%. Moreover the more recent lots can be used for inducing new haploid lines.

This molecular test can be done on each kernel individually, but also on a pool of kernel samples obtained by a puncher according to the method described in the patent application FR No.1450486.

The method currently used for purity control is based on the sowing of these seeds, crossing of each plant with a female to test inducer ability, and sowing of seed obtained from this cross to identify haploid and diploids plants. The phenotypic markers ligule-less or glossy (Lashermes and Beckert 1988, Neuffer, 1997) are currently used for this last test. With the new method according to the invention, the time needed to obtain information on quality control of the lot is drastically reduced and thanks to the number of seeds that can be investigated the accuracy of the results is much higher.

Example 6: Identification of a Zm PL orthologues

Analysis of the deduced amino acid sequences of the wildtype $ZmPL_{b73}$ protein corresponding to SEQ ID N°26, revealed the presence of two S-palmitoylation sites at position 10 and position 423 as well as an S-farnesylation site at position 423 (Table 8a). The second site (423) is missing in the truncated $ZmPL_{PK6}$ protein. Together with the fact that the truncated $ZmPL_{PK6}$ protein is no longer localized in the cytoplasmis membrane, this suggests that the presence of a lipid anchor at the C-terminus of ZmPL may be essential for the correct subcellular localization of the protein.

A survey of the deduced amino acids from cereal orthologues showed that the proteins from *Brachypodium dystachion* (SEQ ID 29), *Sorghum bicolor* (SEQ ID 30), *Panicum virgatum* (SEQ ID 31 and SEQ ID 32), *Setaria italica* (SEQ ID 33) and *Oryza sativa* (SEQ ID 34) are all predicted to have lipid anchors at their N-terminus and C-terminus, whereas for *Hordeum vulgare* (SEQ ID 28) only a N-terminal lipid anchor was predicted. The data suggest that the presence of an N-terminal and a C-terminal lipid anchor by S-palmitoylation, S-farnesylation and/or S-geranylgeranylation may be necessary for subcellular localization in the cytoplasmic membrane (Tables 8b to 8h)

The presence of an N-terminal and C-terminal membrane anchor may be a useful additional criterion for the identification of functional orthologues in non-cereals, where phylogenetic analysis, even combined with the criterion of expression in the pollen, is often not sufficient.

Materials and Methods: For the prediction of lipid anchors allowing membrane association, the "GPS-lipid" web site predictor (http://lipid.biocuckoo.org/webserver.php) was used with the following parameters: "Search for palmitoylation, N-myristoylation, farnesylation geranylgeranylation post-translational modifications" and threshold setting "High".

Table 8a

ID	Position	Peptide	Score	Cutoff	Type
SEQ ID N°26	10	SYSSRRPCNTCSTKA	3.817	3.076	S-Palmitoylation: Cluster B
SEQ ID N°26	423	INPRGSRCASYDI**	5.806	1.983	S-Palmitoylation: Cluster A
SEQ ID N°26	423	INPRGSRCASYDI**	12.801	4.003	S-Farnesylation: Non-consensus

Table 8b

ID	Position	Peptide	Score	Cutoff	Type
SEQ ID N°30	11	YYSSRRPCNACSTKA	4.969	3.076	S-Palmitoylation: Cluster B
SEQ ID N°30	14	SRRPCNACSTKAMAG	3.224	3.076	S-Palmitoylation: Cluster B
SEQ ID N°30	141	KPRYNGKCLRNLIMS	2.223	1.983	S-Palmitoylation: Cluster A
SEQ ID N°30	430	GGASRRTCASKVSNV	6.338	4.003	S-Farnesylation: Non-consensus
SEQ ID N°30	430	GGASRRTCASKVSNV	6.329	1.617	S-Geranylgeranylation: Non-consensus

Table 8c

ID	Position	Peptide	Score	Cutoff	Type
SEQ ID N°33	10	SYSSRRPCNACRTKA	4.493	3.076	S-Palmitoylation: Cluster B
SEQ ID N°33	13	SRRPCNACRTKAMAG	1.979	1.396	S-Palmitoylation: Cluster C
SEQ ID N°33	415	ACAGGSRCCSPVKT*	3.977	1.983	S-Palmitoylation: Cluster A
SEQ ID N°33	415	ACAGGSRCCSPVKT*	8.327	4.003	S-Farnesylation: Non-consensus
SEQ ID N°33	415	ACAGGSRCCSPVKT*	1.112	0.486	S-Geranylgeranylation: CC/CXC
SEQ ID N°33	416	CAGGSRCCSPVKT**	20.687	4.003	S-Farnesylation: Non-consensus
SEQ ID N°33	416	CAGGSRCCSPVKT**	4.008	0.486	S-Geranylgeranylation: CC/CXC

Table 8d

ID	Position	Peptide	Score	Cutoff	Type
SEQ ID N°31	10	SYSSRRPCSVCRTKA	4.212	3.076	S-Palmitoylation: Cluster B
SEQ ID N°31	13	SRRPCSVCRTKAMAG	3.754	3.076	S-Palmitoylation: Cluster B
SEQ ID N°31	422	GAAGGSRCCSPVKLY	3.323	1.983	S-Palmitoylation: Cluster A
SEQ ID N°31	422	GAAGGSRCCSPVKLY	5.441	4.003	S-Farnesylation: Non-consensus
SEQ ID N°31	422	GAAGGSRCCSPVKLY	4.063	0.486	S-Geranylgeranylation: CC/CXC
SEQ ID N°31	423	AAGGSRCCSPVKLY*	21.309	4.003	S-Farnesylation: Non-consensus
SEQ ID N°31	423	AAGGSRCCSPVKLY*	4.572	0.486	S-Geranylgeranylation: CC/CXC

Table 8e

ID	Position	Peptide	Score	Cutoff	Type
SEQ ID N°32	10	SYSSRRPCSVCRTKA	4.212	3.076	S-Palmitoylation: Cluster B
SEQ ID N°32	13	SRRPCSVCRTKAMAR	3.82	3.076	S-Palmitoylation: Cluster B
SEQ ID N°32	422	GCAGGSTCCSPVKT*	4.031	1.983	S-Palmitoylation: Cluster A
SEQ ID N°32	422	GCAGGSTCCSPVKT*	8.189	4.003	S-Farnesylation: Non-consensus
SEQ ID N°32	422	GCAGGSTCCSPVKT*	2.906	0.486	S-Geranylgeranylation: CC/CXC
SEQ ID N°32	423	CAGGSTCCSPVKT**	19.819	4.003	S-Farnesylation: Non-consensus
SEQ ID N°32	423	CAGGSTCCSPVKT**	4.822	0.486	S-Geranylgeranylation: CC/CXC

Table 8f

ID	Position	Peptide	Score	Cutoff	Type
SEQ ID N°29	6	**MASYACRRPCESC	1.559	1.396	S-Palmitoylation: Cluster C
SEQ ID N°29	10	SYACRRPCESCRTRA	1.793	1.396	S-Palmitoylation: Cluster C
SEQ ID N°29	13	CRRPCESCRTRAMAG	1.727	1.396	S-Palmitoylation: Cluster C
SEQ ID N°29	423	PANGKSRC*******	10.896	4.003	S-Farnesylation: Non-consensus
SEQ ID N°29	423	PANGKSRC*******	4.688	1.617	S-Geranylgeranylation: Non-consensus

Table 8g

ID	Position	Peptide	Score	Cutoff	Type
SEQ ID N°34	7	*MAASYSCRRTCEAC	1.542	1.396	S-Palmitoylation: Cluster C
SEQ ID N°34	11	SYSCRRTCEACSTRA	7.03	3.076	S-Palmitoylation: Cluster B
SEQ ID N°34	197	NALLSDICISTSAAP	3.092	3.076	S-Palmitoylation: Cluster B
SEQ ID N°34	430	GEPSGVACKR*****	8.222	4.003	S-Farnesylation: Non-consensus

Table 8h

ID	Position	Peptide	Score	Cutoff	Type
SEQ ID N°28	6	**MASYWCRRPCESC	1.698	1.396	S-Palmitoylation: Cluster C
SEQ ID N°28	10	SYWCRRPCESCSTRA	6.059	3.076	S-Palmitoylation: Cluster B
SEQ ID N°28	13	CRRPCESCSTRAMAG	1.855	1.396	S-Palmitoylation: Cluster C
SEQ ID N°28	195	NARLADICIGTSAAP	3.494	3.076	S-Palmitoylation: Cluster B

Example 7: Identification of a ZmPL functional orthologue from *Arabidopsis thaliana*.

To identify the functional ortholog of ZmPL in the genome of the model plant *Arabidopsis thaliana* a phylogenetic tree of all patatin-like phospholipases was constructed. Using preferential expression in pollen and the prediction of an N-terminal and a C-terminal lipid anchor as additional criteria, the gene At1g61850.1 was identified as the best candidate.

The mutants N642695, N657713 and N596745 (obtained from NASC), which were in a Col-0 wildtype background, were crossed as male parent to a *glabra1* mutant line.

Table 9.: primers identification

SEQ ID name	Primer name	Primer sequence (5' to 3')	Primer use
18	G471240-attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTCGACGTCCGAGCAGGGCC	RNAi fragment
19	G471240-attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTCACCGAGGGCATCGGCATTGCTTCC	RNAi fragment
20	PLPK6_HDPK_F1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGGCAATGGCGAGCTACTC	PL$_{PK6}$-OE, PL$_{HD99}$-OE and PL$_{HD99}$-fusion
21	PLPK6_PK6_R1	GGGGACCACTTTGTACAAGAAAGCTGGGTAGCCTTGTTCTCCTCTCCT	PL$_{PK6}$-OE
22	PLPK6_HD99_R1	GGGGACCACTTTGTACAAGAAAGCTGGGTAGCCACTTGTCTTAGATAT	PL$_{HD99}$-OE
6	Pat_qRT_F1	GAGAAGGAAGCAATGCCGATGCC	qRT-PCR
7	Pat_qRT_R1	TGATTGACAGTAAAGCCACTTGTCTTAGATATC	qRT-PCR
8	actin-q-F	TACCCGATTGAGCATGGCA	qRT-PCR
9	actin-q-R	TCTTCAGGCGAAACACGGA	qRT-PCR
10	G471240_A_F	AGTTCATCACTAATCACACTTATTGTGCC	allelic sequencing
11	G471240_A_R	GGCGGACGTGCCAATGCA	allelic sequencing
17	G471240_B_F	GAACGCTCTGCTCTCGGACG	allelic sequencing
16	G471240_B_R	TATATTCAAGAACATATA	allelic sequencing
12	G471240_C_F	ATGTCCGCGCTGAGGAAGCCA	allelic sequencing
13	G471240_C_R	GAGCACTGCCGCGCCGTGTA	allelic sequencing
14	G471240_D_F	GGAGCTGTACCCAGTGAAGCCG	allelic sequencing
15	G471240_D_R3	TTAGATATCGTACGACGCACATCTAGA	allelic sequencing
46	attB4_prom_PL_2576_B73	GGGGACAACTTTGTATAGAAAAGTTGCTTCAAAATGGTTATGCGTAGGTTGAA	
47	attB1r_prom_PL_25	GGGGACTGCTTTTTTGTACAAACTTGCTGCCGCCTTCGACAACAC	

	76_B73		
48	attB4_promoPL_2534_PK6	GGGGACAACTTTGTATAGAAAAGTTGCTTTTAGGATAAGCCAGAGTTTGT	
49	attB1r_promoPL_2534_PK6	GGGGACTGCTTTTTTGTACAAACTTGCTGCCGCCTTCGACCGCAC	
52	PL-CDS-F_Dtopo	CACCATGGCGAGCTACTCGTCGCGGCGT	
53	PL-CDS-HD99-R2	GATATCGTACGACGCACATCTAGAG	
54	PL-CDS-PK6-R2	GCGAGCCCACCGAGGGCAT	

References

- Eyal, Y., Curie, C., and McCormick, S. (1995)., Pollen specificity elements reside in 30 bp of the proximal promoters of two pollen-expressed genes., Plant Cell *7*, 373–384;
- Barret P; Brinkmann M ; Beckert M., (2008) A major locus expressed in the male gametophyte with incomplete penetrance is responsible for in situ gynogenesis in maize, *Theoretical and Applied Genetics, 117(4) : 581-594;*
- Bordes J., Création de lignées haploïdes doublées de maïs par gynogenèse induite in situ :amélioration de la méthode et intégration dans les schémas de sélection.,Thèse, 2006;
- Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D., Maize HapMap2 identifies extant variation from a genome in flux., Nat Genet. 2012 Jun 3;44(7):803-7.;
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacteriummediated transformation of Arabidopsis thaliana. Plant J 16: 735-743.- Coe 1959., A line of maize with high haploid frequency., The American Naturalist, November-december 1959, p381;
- Deimling S, Röber FK, Geiger HH, (1997), Methodology and genetics of *in vivo* haploid induction in maize [in German]. Vortr Pflanzenz üchtg, 38: 203–224.;
- Dong X ; Xu X ; Miao J ; LI L ; Zhang D ; Mi X ; Liu C ; Tian X ; Melchinger AE ; Chen S., (2013) Fine mapping of qhir1 influencing in vivo haploid induction in maize., *Theoretical And Applied Genetics, 126(7) : 1713-1720;*
- Gaj T., Gersbach C.A., and C.F. Barbas., ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering., Trends in Biotechnology, July 2013, Vol. 31, No. 7;
- Geiger HH; Gordillo GA., (2009) Doubled Haploids In Hybrid Maize Breeding, *Maydica, 54(4) : 485-499. ;*
- Greenblatt IM ; Bock M., A commercially desirable procedure for detection of monoploids in maize., *JOURNAL OF HEREDITY, 58 : 9-13;*
- Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T., and Kumashiro, T. (1996)., High efficiency transformation of maize (Zea mays L) mediated by Agrobacterium tumefaciens., Nat. Biotechnol., *14*, 745–750.;
- Ishida, Y., Hiei, Y., and Komari, T., (2007)., Agrobacterium-mediated transformation of maize., Nat Protoc *2*, 1614–1621.;

- Jaillais, Y., Hothorn, M., Belkhadir, Y., Dabi, T., Nimchuk, Z.L., Meyerowitz, E.M., and Chory, J., (2011)., Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor., Genes Dev., *25*, 232–237.;

- Karimi, M., Bleys, A., Vanderhaeghen, R., and Hilson, P., (2007)., Building blocks for plant gene assembly., Plant Physiol., *145*, 1183–1191.;

- Lashermes P ; Beckert M., Genetic Control Of Maternal Haploidy In Maize (Zea Maysl.) And Selection Of Haploid Inducing Lines., *Theoritical And Applied Genetics, 76(3) : 405-410;*

- Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., -Provart, N., Patel, R., Myers, C.R., Reidel, E.J., Turgeon, R., Liu, P., Sun, Q., Nelson T. & T.P. Brutnell., The developmental dynamics of the maize leaf transcriptome, NATURE GENETICS, 2010, 42(12) : 10p.;

Logemann, E., Birkenbihl, R.P., Ulker, B., and Somssich, I.E. (2006). An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. Plant Methods 2: 16.

- Needleman and Wunsch, 1970 A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48:443-453

- Neuffer MG ; Coe EH ; Wessler SR. *Mutants of maize; Neuffer MG;Coe EH; Wessler ER SR - 1997 - Vol : 468p;*

- Prigge V ; Sanchez C ; Dhillon BS ; Schipprack W ; Araus JL ; Banziger M ; Melchinger AE., (2012) Doubled Haploids in Tropical Maize: I. Effects of Inducers and Source Germplasm on in vivo Haploid Induction Rates., *Crop Science, 51(4) : 1498-1506;*

- Prigge V ; Xu XW ; Li L ; Babu R ; Chen SJ ; Atlin GN ; Melchinger AE,. (2011) New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize., *Genetics, 190(2) : 781-793;*

- Rober FK; Gordillo GA ; Geiger HH., (2005) In vivo haploid induction in maize – Performance of new inducers and significance of doubled haploid lines in hybrid breeding.,
Maydica, 50(3-4) : 275-283.;

- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al., (2009)., The B73 maize genome: complexity, diversity, and dynamics., Science *326*, 1112–1115.;

- Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Buell, C.R., de Leon, N., and Kaeppler, S.M. (2011)., Genome-wide atlas of transcription during maize development. Plant J. Cell Mol. Biol., *66*, 553–563.;

- Weber DF., (2014) Today's Use of Haploids in Corn Plant Breeding *Advances In Agronomy, 123 : 123-144. ;*
- Winter D., Vinegar B.., Nahal. H., Ammar R., Wilson G.V., and N. J. Provar., An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets. 2007, PLoS One 2(8):e718.;
- Xu XW ; Li L ; Dong X ; Jin WW ; Melchniger AE ; Chen SJ., (2013) Gametophytic and zygotic selection leads to segregation distortion through in vivo induction of a maternal haploid in maize., Journal Of Experimental Botany, 64(4) : 1083-1096.

CLAIMS

1. An isolated polynucleotide responsible of haploid induction in maize plants comprising or consisting of:

(a) a nucleotide sequence which comprises or consists in SEQ ID N°1 or 3;

(b) a nucleotide sequence at least 80% identical to the nucleotide sequence of (a) and which is responsible of haploid induction in maize plants ; or

(c) a complementary sequence of the nucleotide sequence of (a) or (b), wherein the complementary sequence and the nucleotide sequence consist of the same number of nucleotides and are 100% complementary.

2. A recombinant DNA construct comprising the polynucleotide according to claim 1 operably linked to at least one regulatory sequence.

3. The recombinant DNA construct according to claim 2, wherein said at least one regulatory sequence is responsible for an expression of the polynucleotide sequence according to claim 1 in maize plants.

4. The recombinant DNA construct according to any of claims 2 or 3, wherein said regulatory sequence is chosen from the rice actin promoter (Act1) and intron, the ubiquitin promoter of maize or the CsVMV promoter, the 35S promoter, the 19S promoter, the pCRV promoter and FAD_2 intron.

5. A recombinant DNA construct, comprising:

(a) a fragment of the polynucleotide according to claim 1 or of a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4 ; and

(b) the complementary sequence thereof;

said fragment and complementary sequence thereof being transcribed in an hairpin RNA to induce RNA interference.

6. A vector comprising the recombinant DNA construct according to any one of claims 2 to 5.

7. A process for transforming a host cell, comprising transforming the host cell with the recombinant DNA construct according to any of claims 2 to 5 or the vector of claim 6.

8. A haploid inducer maize plant or seed comprising the recombinant DNA construct according to any of claims 2 to 5.

9. A process for producing haploid inducer maize plant or enhancing haploid induction ability of a maize plant comprising:
(a) transforming a maize plant cell with the recombinant DNA construct according to any of claims 2 to 5;
(b) regenerating a maize plant from the transformed maize cell; and
(c) growing the transformed maize plant under conditions that are suitable for expression of the recombinant DNA construct, wherein expression of the recombinant DNA construct results in haploid induction or enhancement of haploid induction by pollen of the transformed maize plant.

10. A process for inducing haploid maize plant lines comprising:
(a) growing transformed plants by the process according to claim 9;
(b) using said plants as pollinators during the crossing with a female plant; and
(c) screening the progeny of the cross for haploid plants.

11. A process for identification of a mutant haploid inducer plant, said process comprising the step of identifying an orthologous sequence of the sequence consisting in:
(a) a polynucleotide according to claim 1 a) or (b or a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4; and/or
(b) a polypeptide encoded by a polynucleotide of (a).

12. A process according to claim 11 comprising the steps of:
(1) Identifying an orthologous sequence of the sequence consisting in:
(a) a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4;
(b) a polypeptide encoded by a polynucleotide of (a); and
(2a) Screening a mutant library to identify a mutant of the orthologous sequence identified in step (1); or
(2b) generating a mutant of the orthologous sequence identified in step (1) by using a gene editing methods.

13. A process according to claim 12, wherein the orthologous sequence identified in step (1) code for polypeptidic sequence which comprises a lipid anchor domain in its C-terminal and N-terminal polypeptide sequence and/or is a sequence which is expressed in pollen and/or pollen tube of the plant.

14. A process according to claim 12 or 13, wherein step (2a) comprises the use of a sequence consisting in:
(i) the orthologous sequence identified in step (1); or
(ii) a fragment thereof.

15. A process for identification of a mutant haploid inducer plant, said process comprising:
(a) screening a mutant library to identify a mutant of the sequence consisting in:
(i) a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 ; and/or
(b) generating a mutant of the sequence of step (i) by using a gene editing methods.

16. A process according to claim 15, wherein the step of screening comprises the use of a sequence consisting in:
(i) a polynucleotide according to claim 1 or a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4; or
(ii) a fragment thereof.

17. A process according to any of claims 15 or16, wherein the mutant library is a mutant maize library.

18. A process according to any of claims 11 to 17, wherein the orthologous sequence or identified mutated sequence code for a polypeptidic sequence which does not comprise a lipid anchor domain in its N-terminal polypeptide sequence.

19. A process according to claim 18, wherein said mutant maize library is screened for modified expression of a polynucleotide according to claim 1 (a), (b) or of a polynucleotide comprising or consisting in the nucleotide sequence of SEQ ID N°2 or 4 or of a nucleotide sequence at least 80% identical to said nucleotide sequence of SEQ ID N°2 or 4.

20. A process for determining the presence, the absence or the alteration of the polynucleotide according to claim 1 in a maize plant, comprising at least isolating nucleic acid molecules from said maize plant and amplifying sequences homologous to the polynucleotide according to claim 1.

21. The mutant haploid inducer plant identified by the method according to any of claims 11 to 19.

22. The use of a recombinant DNA comprising an orthologous sequence as identified in any of claims 11 to 14, for transforming a host cell and/ and obtaining a transgenic plant.

23. The use of a recombinant DNA comprising:
(a) a fragment of the orthologous sequences as identified in any of claims 11 to 14 or of a nucleotide sequence at least 80% identical to said nucleotide sequence;
(b) the complementary sequence thereof;
said fragment and complementary sequence thereof being transcribed in an hairpin RNA to induce RNA interference,
for transforming a host cell and obtaining a transgenic plant.

24. A genetic marker of haploid induction in maize plants, wherein said genetic marker identify the polymorphism between the nucleotide sequence SEQ ID N°5 and the nucleotide sequence SEQ ID N°42.

25. A genetic marker according to claim 24, wherein said genetic marker is the genetic marker GRMZM2G471240_14.

26. A process of identification of a haploid inducer maize plant comprising detecting in a maize plant or seed the genetic marker of claim 24 or 25.

27. Use of the genetic marker according to claim 24 or 25 for quality control of seed lots in maize haploid inducer lines, comprising the steps of:
(a) taking a sample of seeds from a seed lot in a maize haploid inducer line;
(b) conducting molecular analyses to identify and quantify the presence of haploid inducer or non-inducer alleles;

(c) deducing from step b) the genetic purity value of the lot for the haploid inducer character.

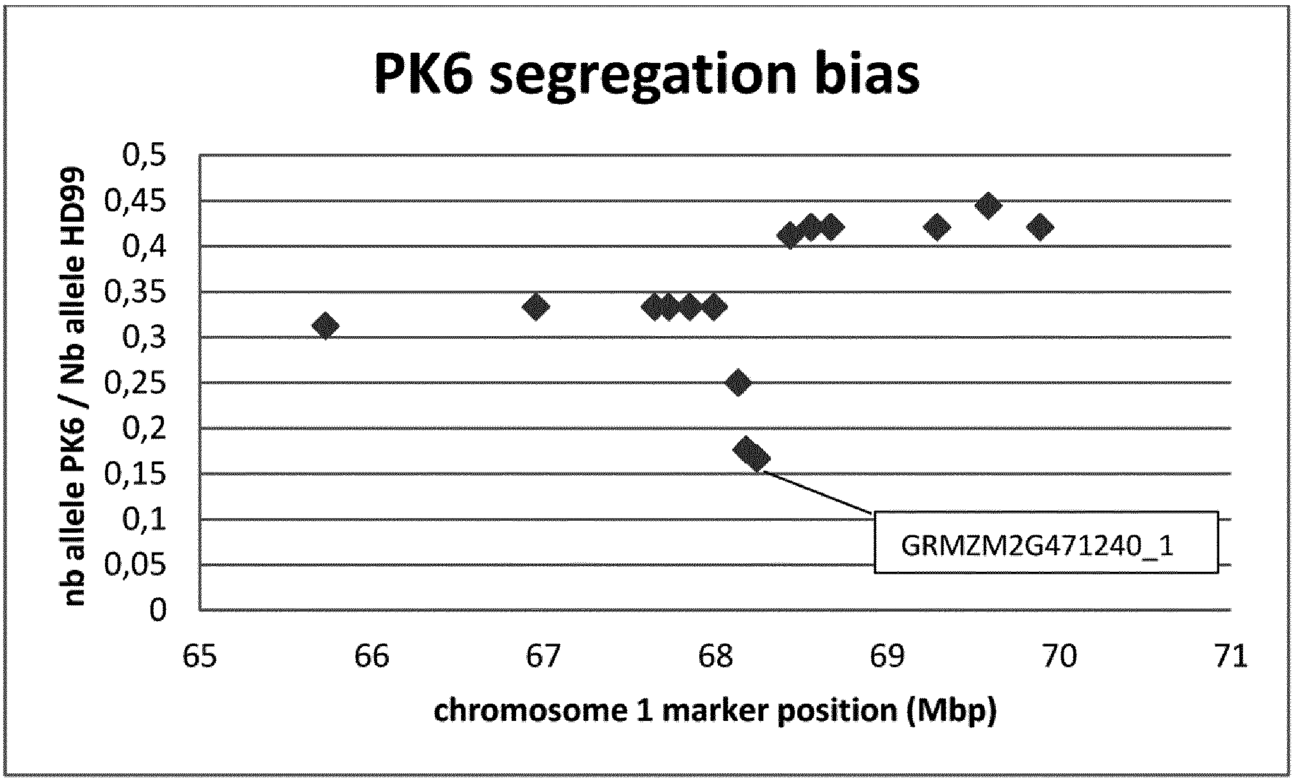

PK6 segregation bias
nb allele PK6 / Nb allele HD99
0,5
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
65
66
67
68
69
70
71
chromosome 1 marker position (Mbp)
GRMZM2G471240_1

FIG.1

```
                                     1                                                50
  GRMZM2G471240_B73_ACB     (1)  AGTTCATCACTAATCACACTTATTGTGCCCTCGACGAGTATCTATAGCTA
 GRMZM2G471240_HD99_ACB     (1)  --------------------------------------------------
  GRMZM2G471240_PK6_ACD     (1)  --------------------------------------------------
                                 51                                              100
  GRMZM2G471240_B73_ACB    (51)  GCTCATTAATCGATTCGGGGGTGTGTTGTCGAAGGCGGCAATGGCGAGCT
 GRMZM2G471240_HD99_ACB     (1)  ---------------------------GTCGAAGGCGGCAATGGCGAGCT
  GRMZM2G471240_PK6_ACD     (1)  -----------------GGGGTGTGCGGTCGAAGGCGGCAATGGCGAGCT
                                 101                                             150
  GRMZM2G471240_B73_ACB   (101)  ACTCGTCGCGGCGTCCATGCAATACCTGTAGCACGAAGGCGATGGCCGGG
 GRMZM2G471240_HD99_ACB    (24)  ACTCGTCGCGGCGTCCATGCAATACCTGTAGCACGAAGGCGATGGCCGGG
  GRMZM2G471240_PK6_ACD    (34)  ACTCGTCGCGGCGTCCATGCAATACCTGTAGCACGAAGGCGATGGCCGGG
                                 151                                             200
  GRMZM2G471240_B73_ACB   (151)  AGCGTGGTCGGCGAGCCCGTCGTGCTGGGGCAGAGGGTGACGGTGCTGAC
 GRMZM2G471240_HD99_ACB    (74)  AGCGTGGTCGGCGAGCCCGTCGTGCTGGGGCAGAGGGTGACGGTGCTGAC
  GRMZM2G471240_PK6_ACD    (84)  AGCGTGGTCGGCGAGCCCGTCGTGCTGGGGCAGAGGGTGACGGTGCTGAC
                                 201                                             250
  GRMZM2G471240_B73_ACB   (201)  GGTGGACGGCGGCGGCGTCCGGGGTCTCATCCCGGGAACCATCCTCGCCT
 GRMZM2G471240_HD99_ACB   (124)  GGTGGACGGCGGCGGCGTCCGGGGTCTCATCCCGGGAACCATCCTCGCCT
  GRMZM2G471240_PK6_ACD   (134)  GGTGGACGGCGGCGGCGTCCGGGGTCTCATCCCGGGAACCATCCTCGCCT
                                 251                                             300
  GRMZM2G471240_B73_ACB   (251)  TCCTGGAGGCCAGGCTGCAGGAGCTGGACGGACCGGAGGCGAGGCTGGCG
 GRMZM2G471240_HD99_ACB   (174)  TCCTGGAGGCCAGGCTGCAGGAGCTGGACGGACCGGAGGCGAGGCTGGCG
  GRMZM2G471240_PK6_ACD   (184)  TCCTGGAGGCCAGGCTGCAGGAGCTGGACGGACCGGAGGCGAGGCTGGCG
                                 301                                             350
  GRMZM2G471240_B73_ACB   (301)  GACTACTTCGACTACATCGCCGGAACCAGCACCGGCGGTCTCATCACCGC
 GRMZM2G471240_HD99_ACB   (224)  GACTACTTCGACTACATCGCCGGAACCAGCACCGGCGGTCTCATCACCGC
  GRMZM2G471240_PK6_ACD   (234)  GACTACTTCGACTACATCGCCGGAACCAGCACCGGCGGTCTCATCACCGC
                                 351                                             400
  GRMZM2G471240_B73_ACB   (351)  CATGCTCACCGCGCCCGGCAAGGACAAGCGGCCTCTCTACGCTGCCAAGG
 GRMZM2G471240_HD99_ACB   (274)  CATGCTCACCGCGCCCGGCAAGGACAAGCGGCCTCTCTACGCTGCCAAGG
  GRMZM2G471240_PK6_ACD   (284)  CATGCTCACCGCGCCCGGCAAGGACAAGCGGCCTCTCTACGCTGCCAAGG
                                 401                                             450
  GRMZM2G471240_B73_ACB   (401)  ACATCAACCACTTTTACATGCAGAACTGCCCGCGCATCTTTCCTCAGAAG
 GRMZM2G471240_HD99_ACB   (324)  ACATCAACCACTTTTACATGGAGAACTGCCCGCGCATCTTCCCTCAGAAG
  GRMZM2G471240_PK6_ACD   (334)  ACATCAACTACTTTTACATGGAGAACTGCCCGCGCATCTTCCCTCAGAAG
                                 451                                             500
  GRMZM2G471240_B73_ACB   (451)  TGAGTCCGATGCTGCCGCCATTGTTCTTGCATCCATCCAGCATCGTACGT
 GRMZM2G471240_HD99_ACB   (374)  TGAGTCCGATGCTGCCGCCATTGTTCTTGCATCCATGCATCCAGCATCGT
  GRMZM2G471240_PK6_ACD   (384)  TGAGTCCGATGCTGCCGCCATTGTTCTCGCATCCATCCAGCAT----CGT
                                 501                                             550
  GRMZM2G471240_B73_ACB   (501)  ACGTCCTCTATACATCTGCGGATCATCATGTGCGCATGTTTGTGGCATGC
 GRMZM2G471240_HD99_ACB   (424)  ACGTCCTCTATACATCTGCGGATGATCATTTGCGCATGTTTGTGGCATGC
  GRMZM2G471240_PK6_ACD   (430)  ACGTCCTCTATACATCTGCGGATGATCATTTGCGCATGTTTGTGGCATGC
                                 551                                             600
  GRMZM2G471240_B73_ACB   (551)  ATGCATGC-ATGTGAGCAGGAGCAGGCTTGCGGCCGCCATGTCCGCGCTG
 GRMZM2G471240_HD99_ACB   (474)  ATGCATGTGATGTGAGCAGGAGCAGGCTTGCGGCCGCCATGTCCGCGCTG
  GRMZM2G471240_PK6_ACD   (480)  AT---------GTGAGCAGGAGCAGGCTTGCGGCCGCCATGTCCGCGCTG
                                 601                                             650
  GRMZM2G471240_B73_ACB   (600)  AGGAAGCCAAAGTACAACGGCAAGTGCATGCGCAGCCTGATTAGGAGCAT
 GRMZM2G471240_HD99_ACB   (524)  AGGAAGCCAAAGTACAACGGCAAGTGCATGCGCAGCCTGATTAGGAGCAT
  GRMZM2G471240_PK6_ACD   (521)  AGGAAGCCAAAGTACAACGGCAAGTGCATGCGCAGCCTGATTAGGAGCAT
                                 651                                             700
  GRMZM2G471240_B73_ACB   (650)  CCTCGGCGAGACGAGGGTAAGCGAGACGCTGACCAACGTCATCATCCCTG
 GRMZM2G471240_HD99_ACB   (574)  CCTCGGCGAGACGAGGGTAAGCGAGACGCTGACCAACGTCATCATCCCTG
  GRMZM2G471240_PK6_ACD   (571)  CCTCGGCGAGACGAGGGTAAGCGAGACGCTGACCAACGTCATCATCCCTG
                                 701                                             750
  GRMZM2G471240_B73_ACB   (700)  CCTTCGACATCAGGCTGCTGCAGCCTATCATCTTCTCTACCTACGACGTA
 GRMZM2G471240_HD99_ACB   (624)  CCTTCGACATCAGGCTGCTGCAGCCTATCATCTTCTCTACCTACGACGTA
  GRMZM2G471240_PK6_ACD   (621)  CCTTCGACATCAGGCTGCTGCAGCCTATCATCTTCTCTACCTACGACGTA
                                 751                                             800
  GRMZM2G471240_B73_ACB   (750)  CGTACGTCGTCACGAATGATTCATCTGTACGTCGTCGCATGCGAATGGCT
```

FIG.2 (beginning)

```
GRMZM2G471240_HD99_ACB  (674) CGTACGTCGTCACGAATGATTCATCTGTACGTCGTCGCATGCGAATGGCT
 GRMZM2G471240_PK6_ACD  (671) CGTACGTCGTCACGAATGATTCATCTGTACGTCGTCGCATGCGAATGGCT
                              801                                            850
 GRMZM2G471240_B73_ACB  (800) GCCTACGTACGCCGTGCGCTAACATACTCAGCTCTTTCCTATCTGCTGCG
GRMZM2G471240_HD99_ACB  (724) GCCTACGTACGCCGTGCGCTAACATACTCAGCTCTTTCCTATCTGCTGCG
 GRMZM2G471240_PK6_ACD  (721) GCCT----ACGCCGTGCGCTAACATACTCAGCTCTTTCCGATCTGCTGCG
                              851                                            900
 GRMZM2G471240_B73_ACB  (850) CCAATTTGCAGGCCAAGAGCACGCCTCTGAAGAACGCTCTGCTCTCGGAC
GRMZM2G471240_HD99_ACB  (774) CCAATTTGCAGGCCAAGAGCACGCCTCTGAAGAACGCTCTGCTCTCGGAC
 GRMZM2G471240_PK6_ACD  (767) CCAATTTGCAGGCCAAGAGCACGCCTCTGAAGAACGCGCTGCTCTCGGAC
                              901                                            950
 GRMZM2G471240_B73_ACB  (900) GTGTGCATTGGCACGTCCGCCGCGCCGACCTACCTCCCGGCGCACTACTT
GRMZM2G471240_HD99_ACB  (824) GTGTGCATTGGCACGTCCGCCGCGCCGACCTACCTCCCGGCGCACTACTT
 GRMZM2G471240_PK6_ACD  (817) GTGTGCATTGGCACGTCCGCCGCGCCGACCTACCTCCCGGCGCACTACTT
                              951                                           1000
 GRMZM2G471240_B73_ACB  (950) CCAGACTGAAGACGCCAACGGCAAGGAGCGCGAATACAACCTCATCGACG
GRMZM2G471240_HD99_ACB  (874) CCAGACTGAAGACGCCAACGGCAAGGAGCGCGAATACAACCTCATCGACG
 GRMZM2G471240_PK6_ACD  (867) CCAGACTGAAGACGCCAACGGCAAGGAGCGCGAATACAACCTCATCGACG
                              1001                                          1050
 GRMZM2G471240_B73_ACB (1000) GCGGTGTGGCGGCCAACAACCCGGTAACTGACTAGCTAACTGGAAAACGG
GRMZM2G471240_HD99_ACB  (924) GCGGTGTGGCGGCCAACAACCCGGTAACTGACTAGCTAACTGGAAAACGG
 GRMZM2G471240_PK6_ACD  (917) GCGGTGTGGCGGCCAACAACCCGGTAACTGACTAGCTAACTGCAAAACGA
                              1051                                          1100
 GRMZM2G471240_B73_ACB (1050) ACGCACAGACTCCATGTCCATGGCGGCCCACAAGGTCGATGCTAATTGTT
GRMZM2G471240_HD99_ACB  (974) ACGCACAGACTCCATGTCCATGGCGGCCCACAAGGTCGATGCTAATTGTT
 GRMZM2G471240_PK6_ACD  (967) ACGCACAGACTCCATGTCCATGGCGGCCCACAAGGTCGATGCTAATTGTT
                              1101                                          1150
 GRMZM2G471240_B73_ACB (1100) GCTTATGTATGTCGCCCGATTGCACATGCGTAGACGATGGTTGCGATGAC
GRMZM2G471240_HD99_ACB (1024) GCTTATGTATGTCGCCCGATTGCACATGCGTAGACGATGGTTGCGATGAC
 GRMZM2G471240_PK6_ACD (1017) GCTTATGTATGTCGCCCGATTGCACATGCGTAGACGATGGTTGCGATGAC
                              1151                                          1200
 GRMZM2G471240_B73_ACB (1150) GCAGATCACCAAAAAGATGCTTGCCAGCAAGGACAAGGCCGAGGAGCTGT
GRMZM2G471240_HD99_ACB (1074) GCAGATCACCAAAAAGATGCTTGCCAGCAAGGACAAGGCCGAGGAGCTGT
 GRMZM2G471240_PK6_ACD (1067) GCAGATCACCAAAAAGATGCTTGCCAGCAAGGACAAGGCCGAGGAGCTGT
                              1201                                          1250
 GRMZM2G471240_B73_ACB (1200) ACCCAGTGAAGCCGTCGAACTGCCGCAGGTTCCTGGTGCTGTCCATCGGG
GRMZM2G471240_HD99_ACB (1124) ACCCAGTGAAGCCGTCGAACTGCCGCAGGTTCCTGGTGCTGTCCATCGGG
 GRMZM2G471240_PK6_ACD (1117) ACCCAGTGAACCCGTCGAACTGCCGCAGGTTCCTGGTGCTGTCCATCGGG
                              1251                                          1300
 GRMZM2G471240_B73_ACB (1250) ACGGGGTCGACGTCCGAGCAGGGCCTCTACACGGCGCGGCAGTGCTCCCG
GRMZM2G471240_HD99_ACB (1174) ACGGGGTCGACGTCCGAGCAGGGCCTCTACACGGCGCGGCAGTGCTCCCG
 GRMZM2G471240_PK6_ACD (1167) ACGGGGTCGACGTCCGAGCAGGGCCTCTACACGGCGCGGCAGTGCTCCCG
                              1301                                          1350
 GRMZM2G471240_B73_ACB (1300) GTGGGGTATCTGCCGGTGGCTCCGCAACAACGGCATGGCCCCCATCATCG
GRMZM2G471240_HD99_ACB (1224) GTGGGGTATCTGCCGGTGGCTCCGCAACAACGGCATGGCCCCCATCATCG
 GRMZM2G471240_PK6_ACD (1217) GTGGGGCATCTGCCGGTGGCTCCGCAACAACGGCATGGCCCCCATCATCG
                              1351                                          1400
 GRMZM2G471240_B73_ACB (1350) ACATCTTCATGGCGGCCAGCTCGGACCTGGTGGACATCCACGTCGCCGCG
GRMZM2G471240_HD99_ACB (1274) ACATCTTCATGGCGGCCAGCTCGGACCTGGTGGACATCCACGTCGCCGCG
 GRMZM2G471240_PK6_ACD (1267) ACATCTTCATGGCGGCCAGCTCGGACCTGGTGGACATCCACGTCGCCGCG
                              1401                                          1450
 GRMZM2G471240_B73_ACB (1400) ATGTTCCAGTCGCTCCACAGCGACGGCGACTACCTGCGCATCCAGGACAA
GRMZM2G471240_HD99_ACB (1324) ATGTTCCAGTCGCTCCACAGCGACGGCGACTACCTGCGCATCCAGGACAA
 GRMZM2G471240_PK6_ACD (1317) ATGTTCCAGTCGCTCCACAGCGACGGCGACTACCTACGCATCCAGGACAA
                              1451                                          1500
 GRMZM2G471240_B73_ACB (1450) CTCGCTCCGTGGCGCCGCGGCCACCGTGGACGCGGCGACGCCGGAGAACA
GRMZM2G471240_HD99_ACB (1374) CTCGCTCCGTGGCGCCGCGGCCACCGTGGACGCGGCGACGCCGGAGAACA
 GRMZM2G471240_PK6_ACD (1367) CTCGCTCCGTGGCGCCGCGGCAACCGTGGACGCGGCGACGCCGGAGAACA
```

FIG.2 (continuation 1)

```
                                   1501                                             1550
  GRMZM2G471240_B73_ACB   (1500)   TGCGGACGCTCGTCGGGATCGGGGAGCGGATGCTGGCACAGAGGGTGTCC
 GRMZM2G471240_HD99_ACB   (1424)   TGCGGACGCTCGTCGGGATCGGGGAGCGGATGCTGGCACAGAGGGTGTCC
  GRMZM2G471240_PK6_ACD   (1417)   TGCGGACGCTCGTCGGGATCGGGGAGCGGATGCTGGCACAGCGGGTGTCC
                                   1551                                             1600
  GRMZM2G471240_B73_ACB   (1550)   AGGGTCAACGTGGAGACAGGGAG----GTACGAACCGGTGACTGGCGAAG
 GRMZM2G471240_HD99_ACB   (1474)   AGGGTCAACGTGGAGACAGGGAG----GTACGAACCGGTGACTGGCGAAG
  GRMZM2G471240_PK6_ACD   (1467)   AGGGTCAACGTGGAGACAGGGAGCGAGGTACGAACCGGTGACCGGAGAAG
                                   1601                                             1650
  GRMZM2G471240_B73_ACB   (1596)   GAAGCAATGCCGATGCCCTCGGTGGGCTCGCTAGGCAGCTCTCCGAGGAG
 GRMZM2G471240_HD99_ACB   (1520)   GAAGCAATGCCGATGCCCTCGGTGGGCTCGCTAGGCAGCTCTCCGAGGAG
  GRMZM2G471240_PK6_ACD   (1517)   GAAGCAATGCCGATGCCCTCGGTGGGCTCGCTAGGCAGCTCTCCGAGGAG
                                   1651                                             1700
  GRMZM2G471240_B73_ACB   (1646)   AGGAGAACAAGGCTCGCGCCGCGTCTCTGCCATCAACCCAAGAGGCTC
 GRMZM2G471240_HD99_ACB   (1570)   AGGAGAACAAGGCTCGCGCCGCGTCTCTGCCATCAACCCAAGAGGCTC
  GRMZM2G471240_PK6_ACD   (1567)   AGGAGAACAAGGCTCGCGCCGCGTCTCTGCCATCAACCCCAGAAGCTC
                                   1701                                             1750
  GRMZM2G471240_B73_ACB   (1696)   TAGATGTGCGTCGTACGATATCTAAGACAAGTGGCTTTACTGTCAGTCAC
 GRMZM2G471240_HD99_ACB   (1620)   TAGATGTGCGTCGTACGATATCTAAGACAAGTGGCTTTACTGTCAGTCAC
  GRMZM2G471240_PK6_ACD   (1617)   TAGATGTGCGCCCTACGATATCTAAGACAAGTGGCTTTACTGTCAATCAC
                                   1751                                             1800
  GRMZM2G471240_B73_ACB   (1746)   ATGCTTGTAAATAAGTAGACTTTATTTTAATAAAACATAAAAATATATAT
 GRMZM2G471240_HD99_ACB   (1670)   ATGCTTGTAAATAAGTAGACTTTATTTTAATAAAACATAAAAATATATAT
  GRMZM2G471240_PK6_ACD   (1667)   ATGCTTGTAAATAAGTAGACTTTATTTTAATAAAATATAA--ATATATAT
                                   1801                                             1850
  GRMZM2G471240_B73_ACB   (1796)   ATGTTCTTGAATATAAAATTGATAACCAAATTAAAATTCGAACCATCACT
 GRMZM2G471240_HD99_ACB   (1720)   ATGTTCTTGAATATAAAATTGATAACCAAATTAAAA--------------
  GRMZM2G471240_PK6_ACD   (1715)   ATATTCT-------------GATAACCAAG-----ATTCGAACCCTCACT
                                   1851                                         1897
  GRMZM2G471240_B73_ACB   (1846)   TATACATAATTTTACTTTATTTTTTATAAAACGTGAACGGGAAGGAC
 GRMZM2G471240_HD99_ACB   (1756)   -----------------------------------------------
  GRMZM2G471240_PK6_ACD   (1747)   TATACACAATTTTATCTTATTTTTTATAAAACGTGAACG--------
```

FIG.2 (continuation 2)

```
                   1                                                         60
CDS_B73_T01    (1) MASYSSRRPCNTCSTKAMAGSVVGEPVVLGQRVTVLTVDGGGVRGLIPGTILAFLEARLQ
CDS_B73_T02    (1) MASYSSRRPCNTCSTKAMAGSVVGEPVVLGQRVTVLTVDGGGVRGLIPGTILAFLEARLQ
CDS_HD99_ACB   (1) MASYSSRRPCNTCSTKAMAGSVVGEPVVLGQRVTVLTVDGGGVRGLIPGTILAFLEARLQ
CDS_PK6_ACD    (1) MASYSSRRPCNTCSTKAMAGSVVGEPVVLGQRVTVLTVDGGGVRGLIPGTILAFLEARLQ
                   61                                                       120
CDS_B73_T01   (61) ELDGPEARLADYFDYIAGTSTGGLITAMLTAPGKDKRPLYAAKDINHFYMQNCPRIFPQK
CDS_B73_T02   (61) ELDGPEARLADYFDYIAGTSTGGLITAMLTAPGKDKRPLYAAKDINHFYMQNCPRIFPQK
CDS_HD99_ACB  (61) ELDGPEARLADYFDYIAGTSTGGLITAMLTAPGKDKRPLYAAKDINHFYMENCPRIFPQK
CDS_PK6_ACD   (61) ELDGPEARLADYFDYIAGTSTGGLITAMLTAPGKDKRPLYAAKDINYFYMENCPRIFPQK
                   121                                                      180
CDS_B73_T01  (121) SRLAAAMSALRKPKYNGKCMRSLIRSILGETRVSETLTNVIIPAFDIRLLQPIIFSTYDA
CDS_B73_T02  (121) SRLAAAMSALRKPKYNGKCMRSLIRSILGETR---------------------------A
CDS_HD99_ACB (121) SRLAAAMSALRKPKYNGKCMRSLIRSILGETRVSETLTNVIIPAFDIRLLQPIIFSTYDA
CDS_PK6_ACD  (121) SRLAAAMSALRKPKYNGKCMRSLIRSILGETRVSETLTNVIIPAFDIRLLQPIIFSTYDA
                   181                                                      240
CDS_B73_T01  (181) KSTPLKNALLSDVCIGTSAAPTYLPAHYFQTEDANGKEREYNLIDGGVAANNPTMVAMTQ
CDS_B73_T02  (154) KSTPLKNALLSDVCIGTSAAPTYLPAHYFQTEDANGKEREYNLIDGGVAANNPTMVAMTQ
CDS_HD99_ACB (181) KSTPLKNALLSDVCIGTSAAPTYLPAHYFQTEDANGKEREYNLIDGGVAANNPTMVAMTQ
CDS_PK6_ACD  (181) KSTPLKNALLSDVCIGTSAAPTYLPAHYFQTEDANGKEREYNLIDGGVAANNPTMVAMTQ
                   241                                                      300
CDS_B73_T01  (241) ITKKMLASKDKAEELYPVKPSNCRRFLVLSIGTGSTSEQGLYTARQCSRWGICRWLRNNG
CDS_B73_T02  (214) ITKKMLASKDKAEELYPVKPSNCRRFLVLSIGTGSTSEQGLYTARQCSRWGICRWLRNNG
CDS_HD99_ACB (241) ITKKMLASKDKAEELYPVKPSNCRRFLVLSIGTGSTSEQGLYTARQCSRWGICRWLRNNG
CDS_PK6_ACD  (241) ITKKMLASKDKAEELYPVNPSNCRRFLVLSIGTGSTSEQGLYTARQCSRWGICRWLRNNG
                   301                                                      360
CDS_B73_T01  (301) MAPIIDIFMAASSDLVDIHVAAMFQSLHSDGDYLRIQDNSLRGAAATVDAATPENMRTLV
CDS_B73_T02  (274) MAPIIDIFMAASSDLVDIHVAAMFQSLHSDGDYLRIQDNSLRGAAATVDAATPENMRTLV
CDS_HD99_ACB (301) MAPIIDIFMAASSDLVDIHVAAMFQSLHSDGDYLRIQDNSLRGAAATVDAATPENMRTLV
CDS_PK6_ACD  (301) MAPIIDIFMAASSDLVDIHVAAMFQSLHSDGDYLRIQDNSLRGAAATVDAATPENMRTLV
                   361                                                      420
CDS_B73_T01  (361) GIGERMLAQRVSRVNVETGRYEPVTGEGSNADALGGLARQLSEERRTRLARRVSAINPRG
CDS_B73_T02  (334) GIGERMLAQRVSRVNVETGRYEPVTGEGSNADALGGLARQLSEERRTRLARRVSAINPRG
CDS_HD99_ACB (361) GIGERMLAQRVSRVNVETGRYEPVTGEGSNADALGGLARQLSEERRTRLARRVSAINPRG
CDS_PK6_ACD  (361) GIGERMLAQRVSRVNVETGSEVRTGDRRRKQCRCPRWAR---------------------
                        421
CDS_B73_T01  (421) SRCASYDI
CDS_B73_T02  (394) SRCASYDI
CDS_HD99_ACB (421) SRCASYDI
CDS_PK6_ACD  (400) --------
```

FIG.3

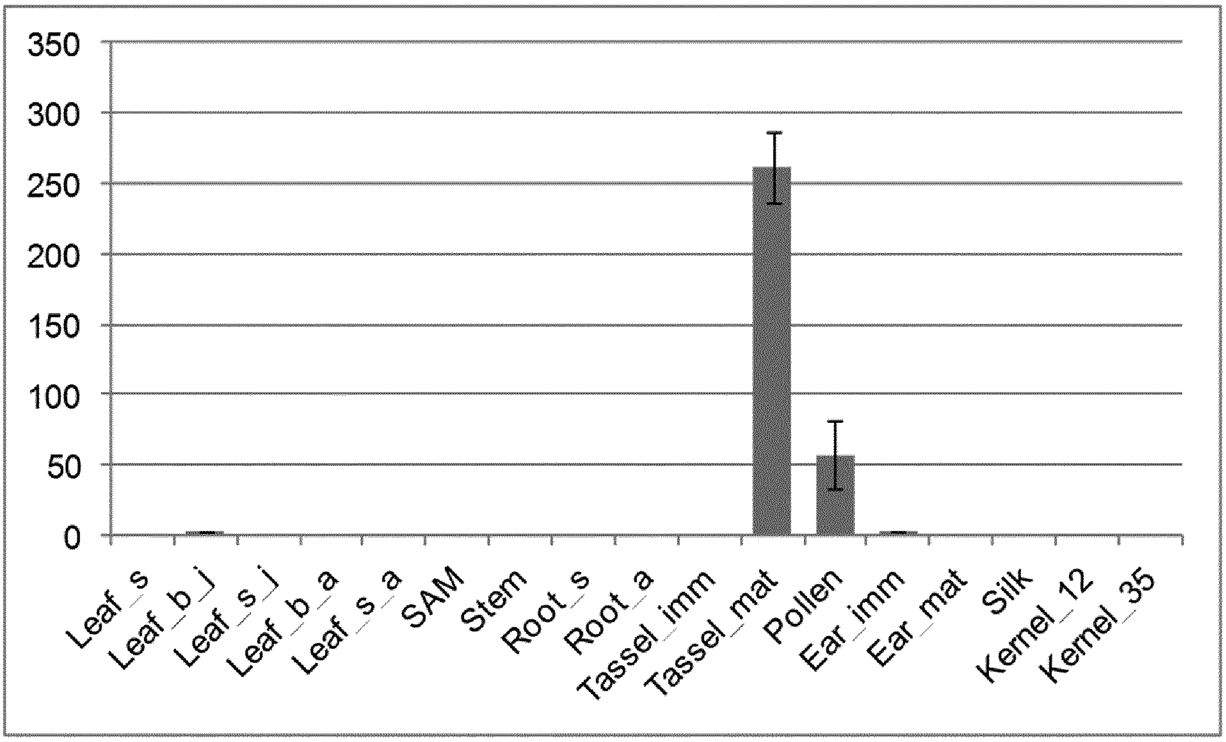

350
300
250
200
150
100
50
0
Leaf_s
Leaf_b_j
Leaf_s_j
Leaf_b_a
Leaf_s_a
SAM
Stem
Root_s
Root_a
Tassel_imm
Tassel_mat
Pollen
Ear_imm
Ear_mat
Silk
Kernel_12
Kernel_35

FIG.4

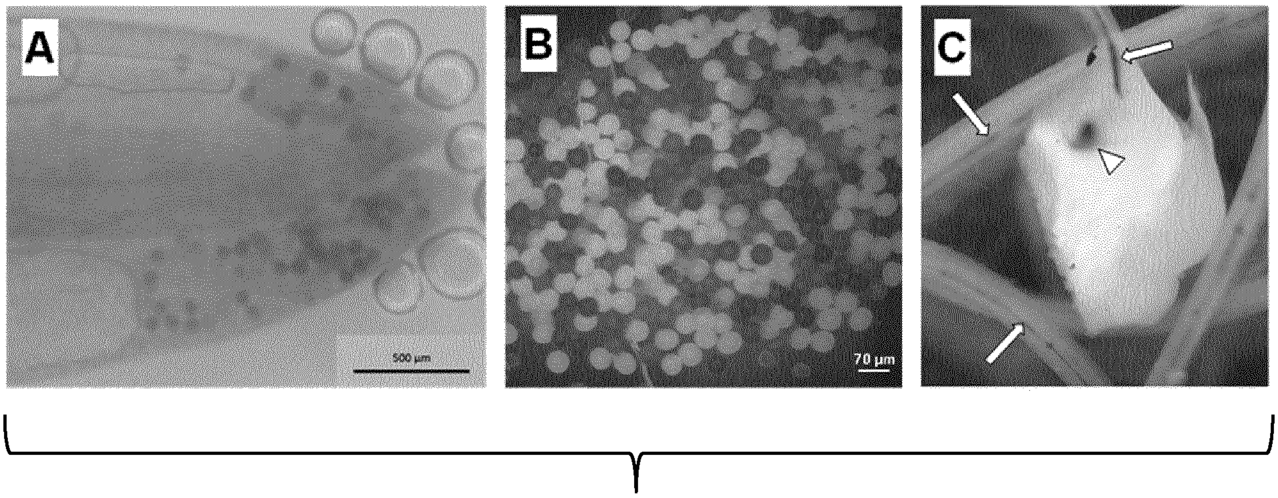
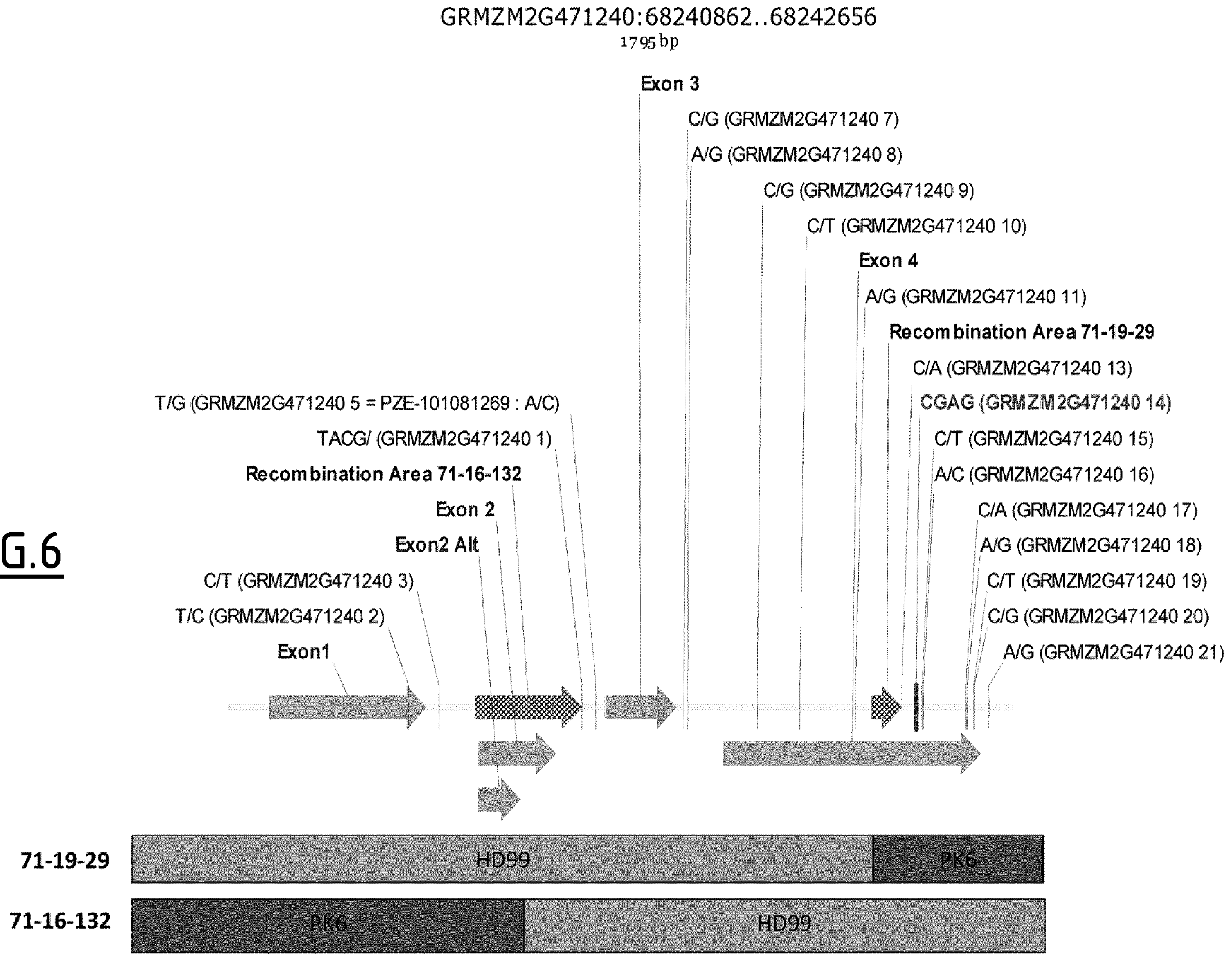

A
500 µm
B
70 µm
C

FIG.5

FIG.6


GRMZM2G471240:68240862..68242656
1795 bp
Exon 3
C/G (GRMZM2G471240 7)
A/G (GRMZM2G471240 8)
C/G (GRMZM2G471240 9)
C/T (GRMZM2G471240 10)
Exon 4
A/G (GRMZM2G471240 11)
Recombination Area 71-19-29
C/A (GRMZM2G471240 13)
CGAG (GRMZM2G471240 14)
C/T (GRMZM2G471240 15)
A/C (GRMZM2G471240 16)
C/A (GRMZM2G471240 17)
A/G (GRMZM2G471240 18)
C/T (GRMZM2G471240 19)
C/G (GRMZM2G471240 20)
A/G (GRMZM2G471240 21)
T/G (GRMZM2G471240 5 = PZE-101081269 : A/C)
TACG/ (GRMZM2G471240 1)
Recombination Area 71-16-132
Exon 2
Exon2 Alt
C/T (GRMZM2G471240 3)
T/C (GRMZM2G471240 2)
Exon1
71-19-29
HD99
PK6
71-16-132
PK6
HD99

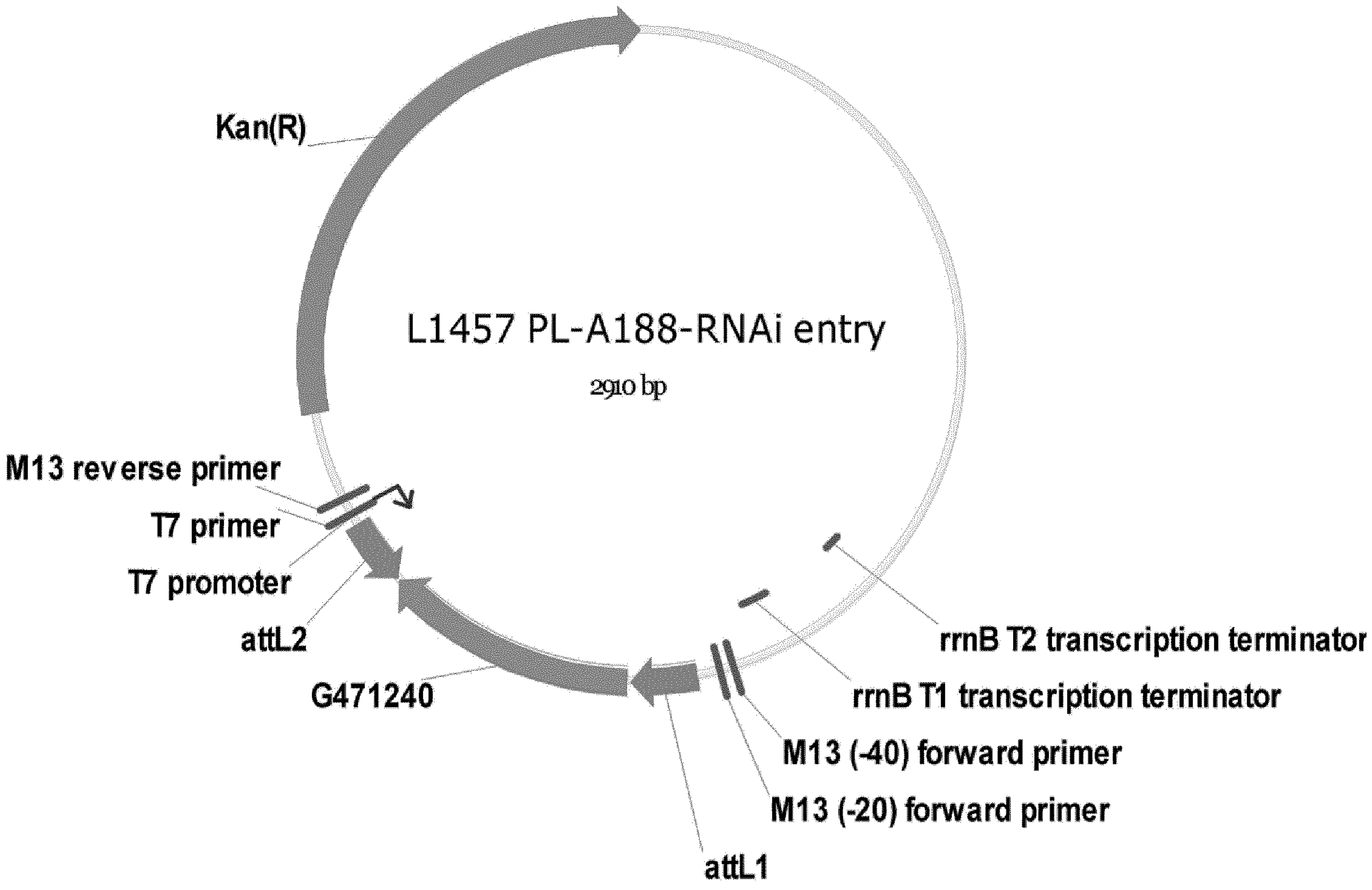

L1457 PL-A188-RNAi entry
2910 bp
Kan(R)
M13 reverse primer
T7 primer
T7 promoter
attL2
G471240
rrnB T2 transcription terminator
rrnB T1 transcription terminator
M13 (-40) forward primer
M13 (-20) forward primer
attL1

FIG.7

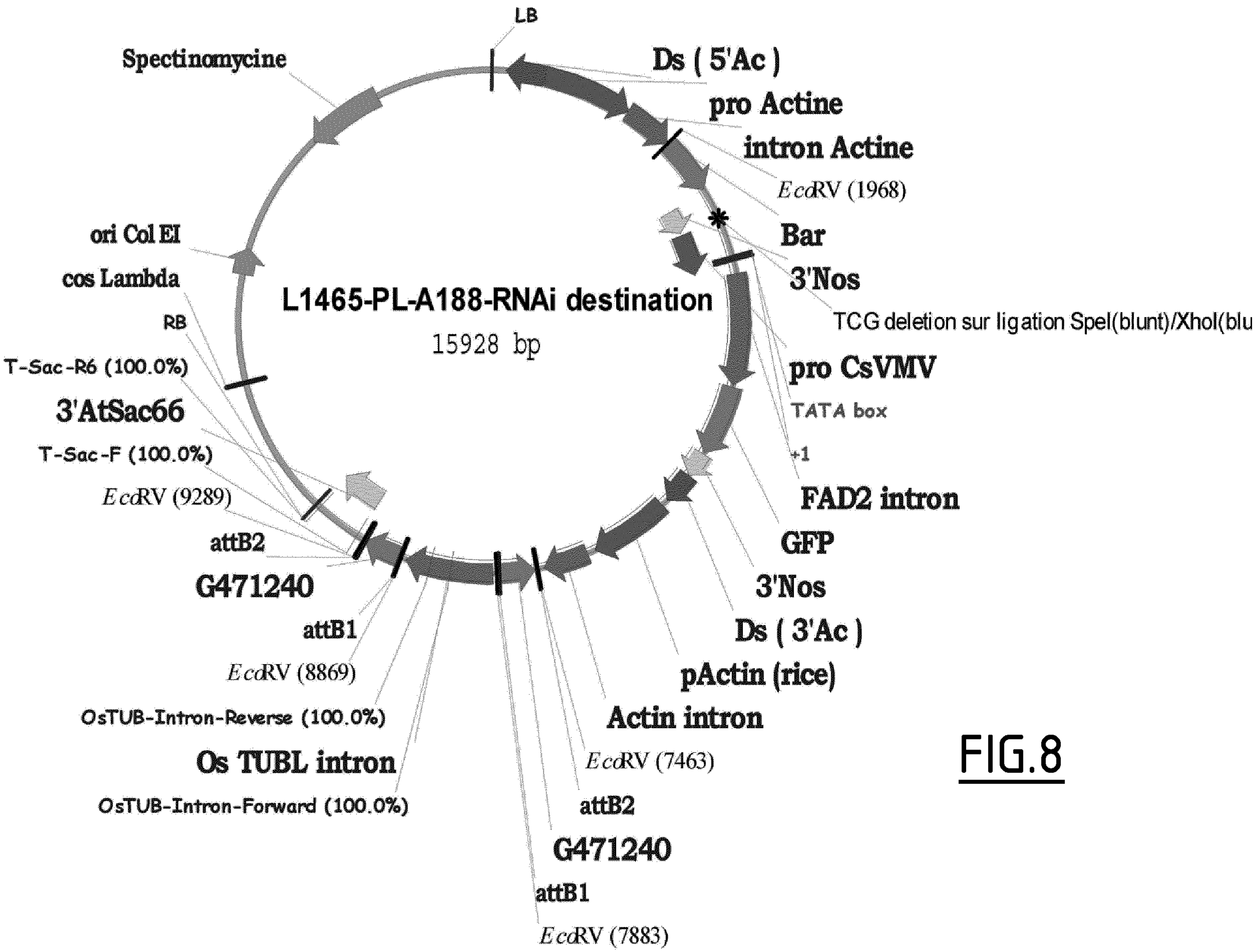

L1465-PL-A188-RNAi destination
15928 bp
LB
Ds (5'Ac)
pro Actine
intron Actine
EcdRV (1968)
Bar
3'Nos
TCG deletion sur ligation SpeI(blunt)/XhoI(blu
pro CsVMV
TATA box
+1
FAD2 intron
GFP
3'Nos
Ds (3'Ac)
pActin (rice)
Actin intron
EcdRV (7463)
attB2
G471240
attB1
EcdRV (7883)
Spectinomycine
ori Col EI
cos Lambda
RB
T-Sac-R6 (100.0%)
3'AtSac66
T-Sac-F (100.0%)
EcdRV (9289)
attB2
G471240
attB1
EcdRV (8869)
OsTUB-Intron-Reverse (100.0%)
Os TUBL intron
OsTUB-Intron-Forward (100.0%)

FIG.8

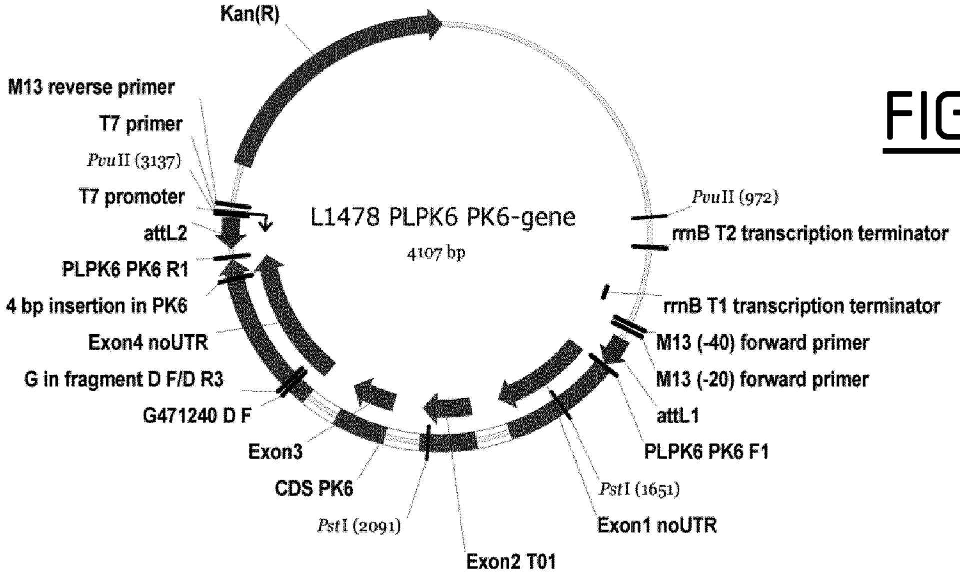

Kan(R)
M13 reverse primer
T7 primer
*Pvu*II (3137)
T7 promoter
attL2
PLPK6 PK6 R1
4 bp insertion in PK6
Exon4 noUTR
G in fragment D F/D R3
G471240 D F
Exon3
CDS PK6
*Pst*I (2091)
Exon2 T01
Exon1 noUTR
*Pst*I (1651)
PLPK6 PK6 F1
attL1
M13 (-20) forward primer
M13 (-40) forward primer
rrnB T1 transcription terminator
rrnB T2 transcription terminator
*Pvu*II (972)
L1478 PLPK6 PK6-gene
4107 bp

FIG.9

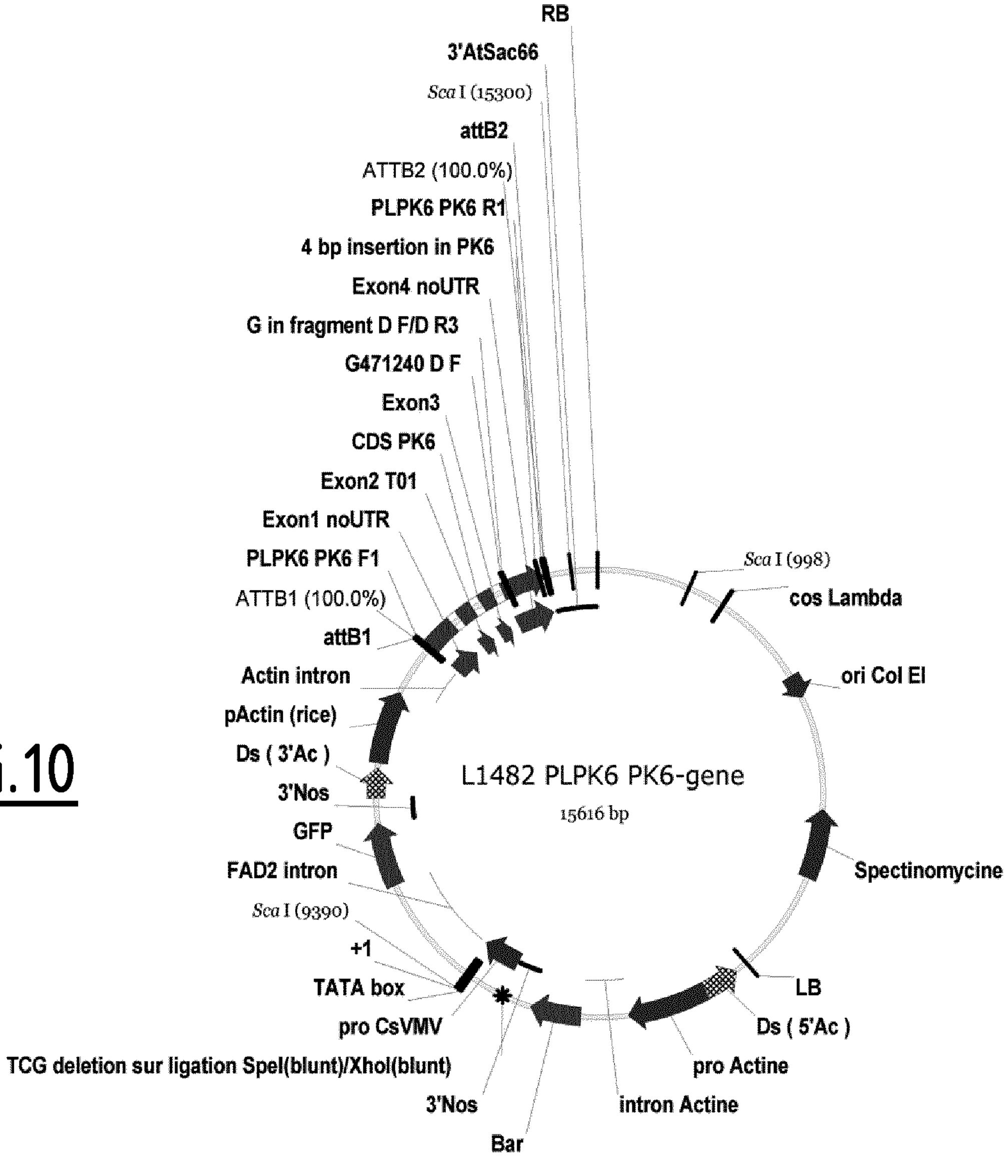

RB
3'AtSac66
Sca I (15300)
attB2
ATTB2 (100.0%)
PLPK6 PK6 R1
4 bp insertion in PK6
Exon4 noUTR
G in fragment D F/D R3
G471240 D F
Exon3
CDS PK6
Exon2 T01
Exon1 noUTR
PLPK6 PK6 F1
ATTB1 (100.0%)
attB1
Actin intron
pActin (rice)
Ds (3'Ac)
3'Nos
GFP
FAD2 intron
Sca I (9390)
+1
TATA box
pro CsVMV
TCG deletion sur ligation SpeI(blunt)/XhoI(blunt)
3'Nos
Bar
intron Actine
pro Actine
Ds (5'Ac)
LB
Spectinomycine
ori Col El
cos Lambda
Sca I (998)
L1482 PLPK6 PK6-gene
15616 bp

FIG.10

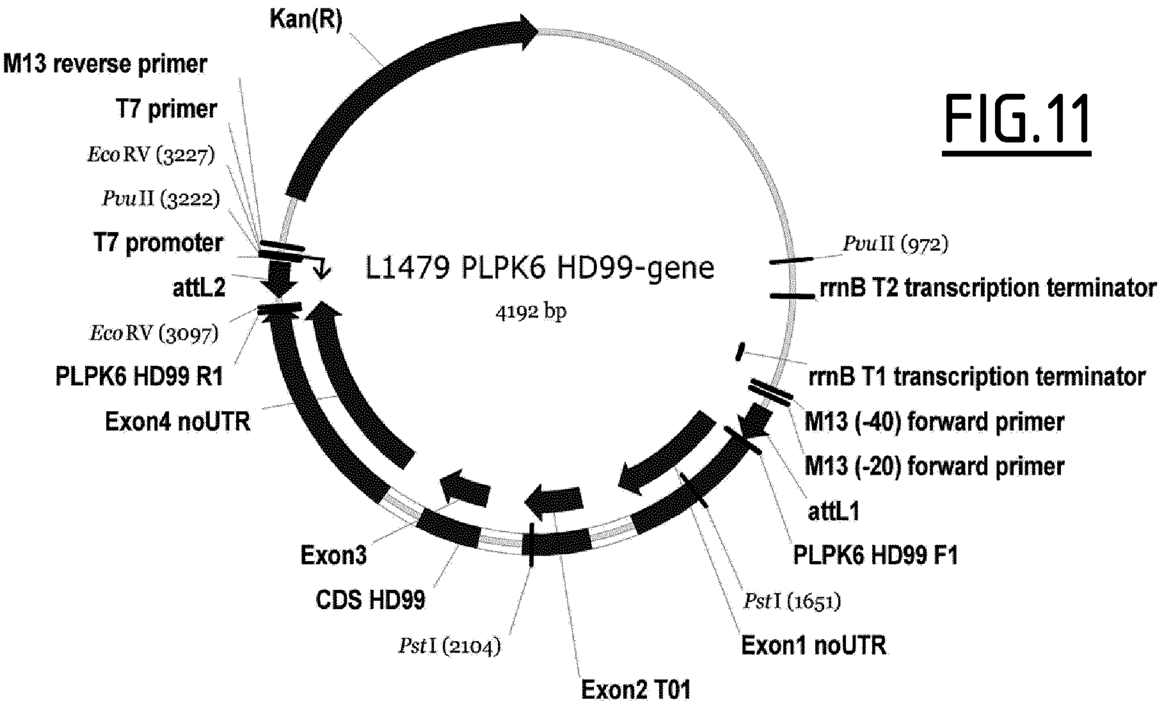
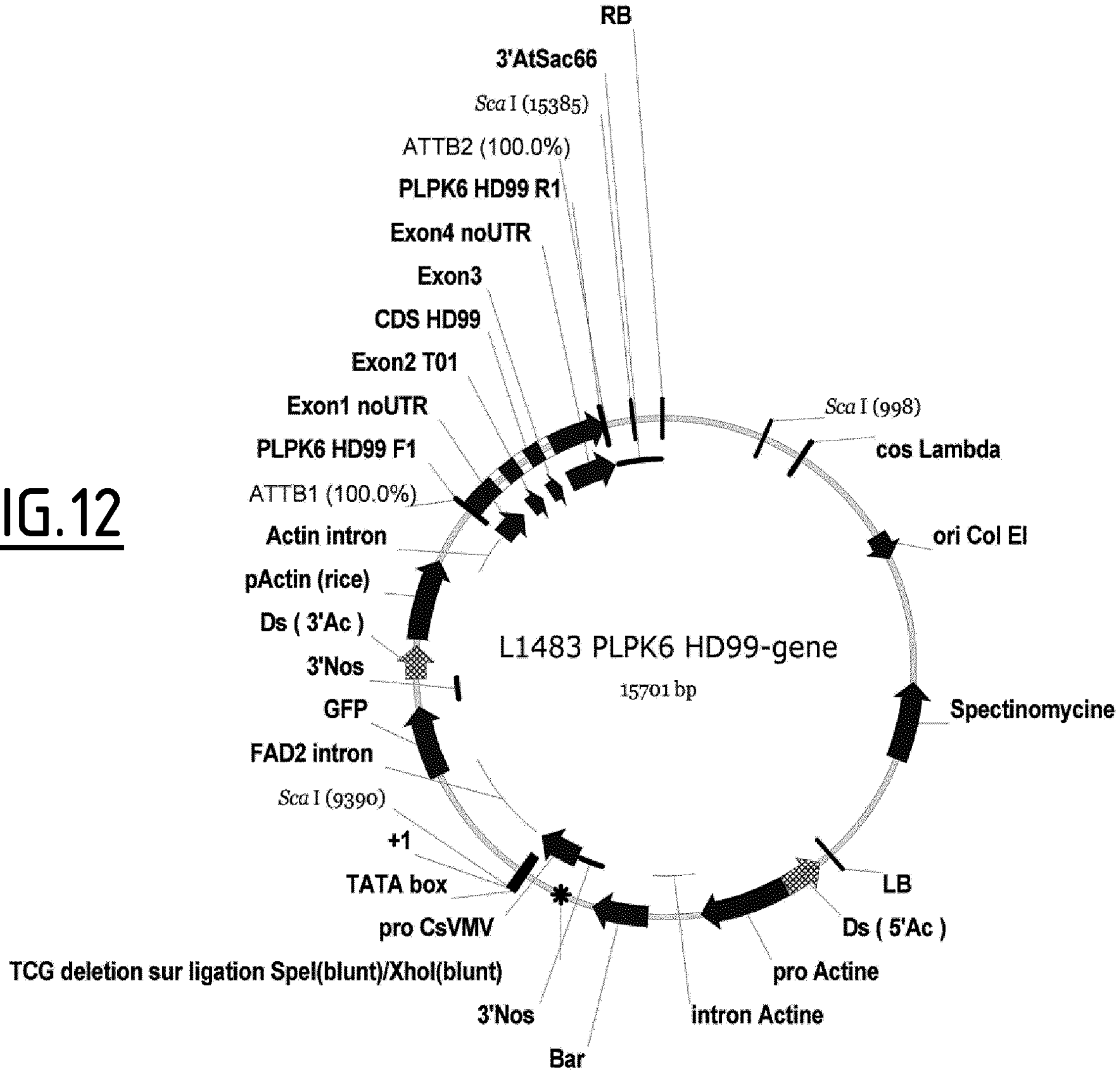

Kan(R)
M13 reverse primer
T7 primer
*Eco*RV (3227)
*Pvu*II (3222)
T7 promoter
attL2
*Eco*RV (3097)
PLPK6 HD99 R1
Exon4 noUTR
Exon3
CDS HD99
*Pst*I (2104)
Exon2 T01
Exon1 noUTR
*Pst*I (1651)
PLPK6 HD99 F1
attL1
M13 (-20) forward primer
M13 (-40) forward primer
rrnB T1 transcription terminator
rrnB T2 transcription terminator
*Pvu*II (972)
L1479 PLPK6 HD99-gene
4192 bp

FIG.11

FIG.12


RB
3'AtSac66
*Sca*I (15385)
ATTB2 (100.0%)
PLPK6 HD99 R1
Exon4 noUTR
Exon3
CDS HD99
Exon2 T01
Exon1 noUTR
PLPK6 HD99 F1
ATTB1 (100.0%)
Actin intron
pActin (rice)
Ds (3'Ac)
3'Nos
GFP
FAD2 intron
*Sca*I (9390)
+1
TATA box
pro CsVMV
TCG deletion sur ligation SpeI(blunt)/XhoI(blunt)
3'Nos
Bar
intron Actine
pro Actine
Ds (5'Ac)
LB
Spectinomycine
ori Col EI
cos Lambda
*Sca*I (998)
L1483 PLPK6 HD99-gene
15701 bp

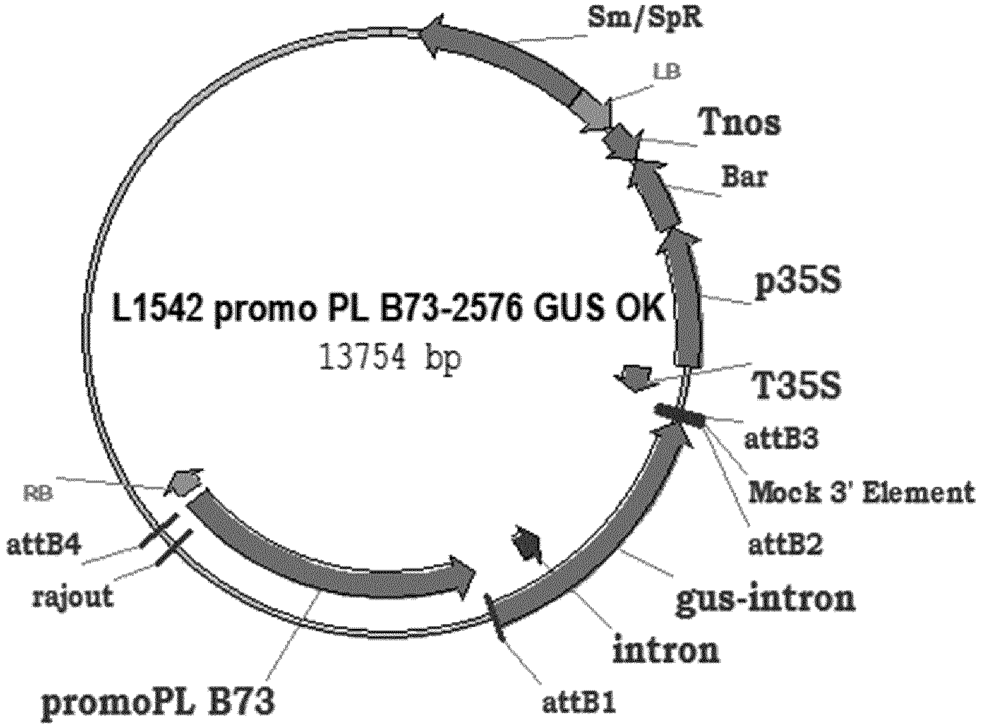

Sm/SpR
LB
Tnos
Bar
p35S
L1542 promo PL B73-2576 GUS OK
13754 bp
T35S
attB3
Mock 3' Element
attB2
RB
attB4
rajout
gus-intron
intron
promoPL B73
attB1

FIG.13

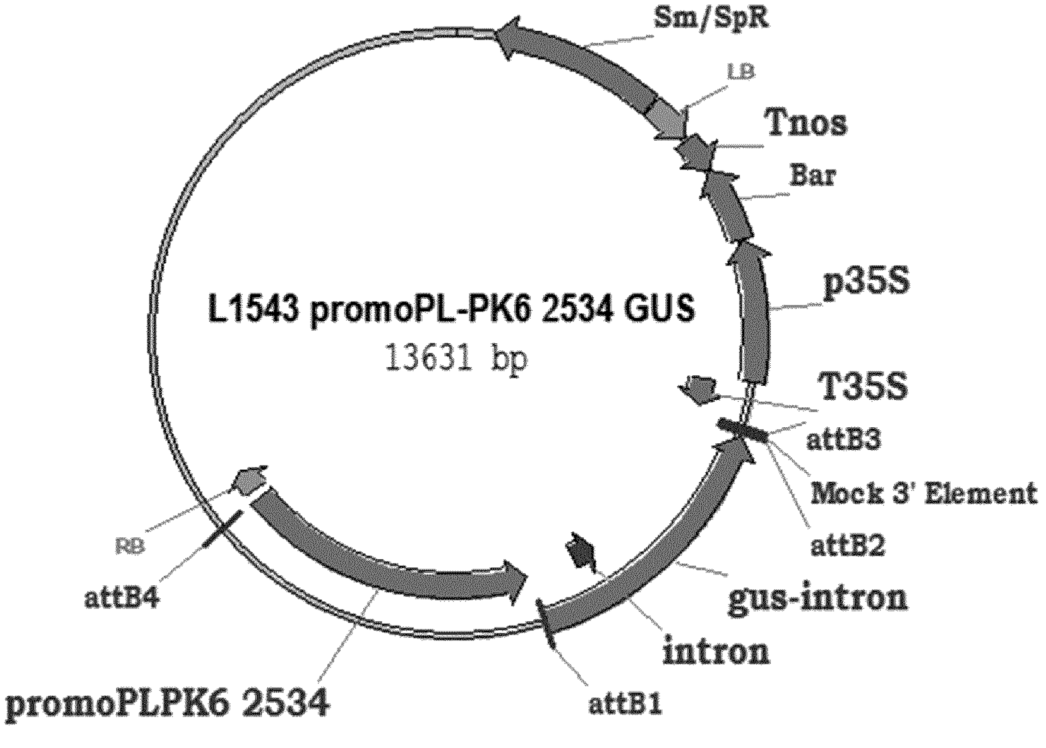

Sm/SpR
LB
Tnos
Bar
p35S
L1543 promoPL-PK6 2534 GUS
13631 bp
T35S
attB3
Mock 3' Element
attB2
RB
attB4
gus-intron
intron
promoPLPK6 2534
attB1

FIG.14

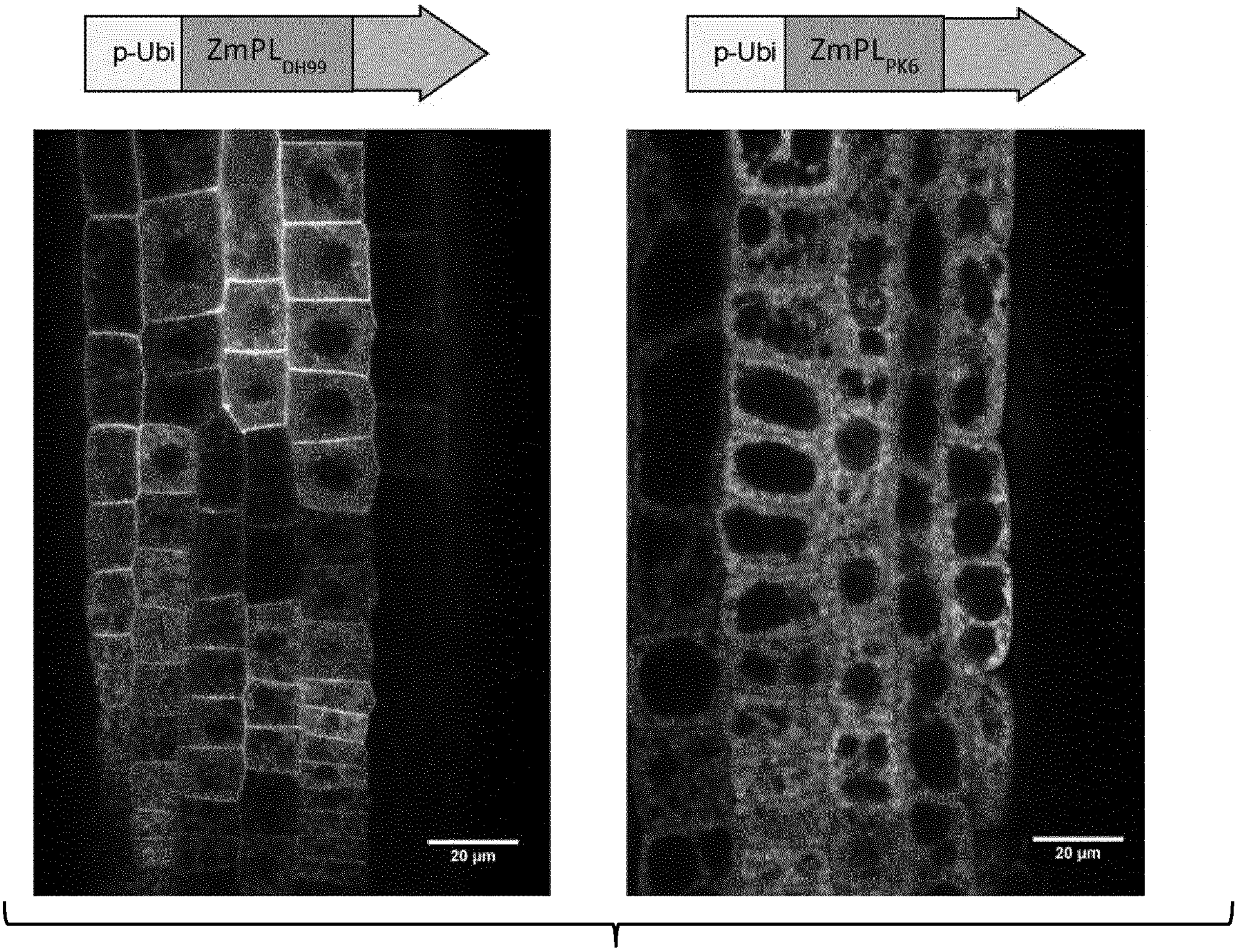

p-Ubi
ZmPL$_{DH99}$
p-Ubi
ZmPL$_{PK6}$
20 µm
20 µm

FIG.15

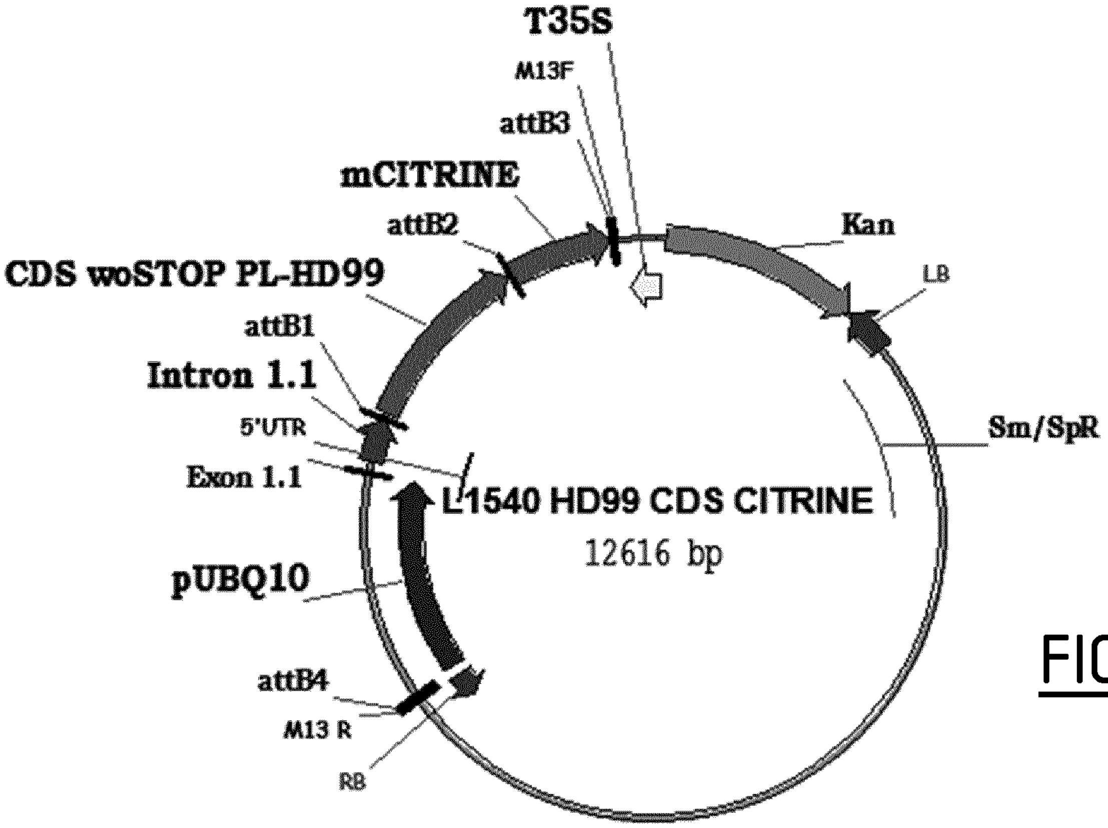

T35S
M13F
attB3
mCITRINE
attB2
Kan
LB
CDS woSTOP PL-HD99
attB1
Intron 1.1
5'UTR
Exon 1.1
Sm/SpR
L1540 HD99 CDS CITRINE
12616 bp
pUBQ10
attB4
M13 R
RB

FIG.16

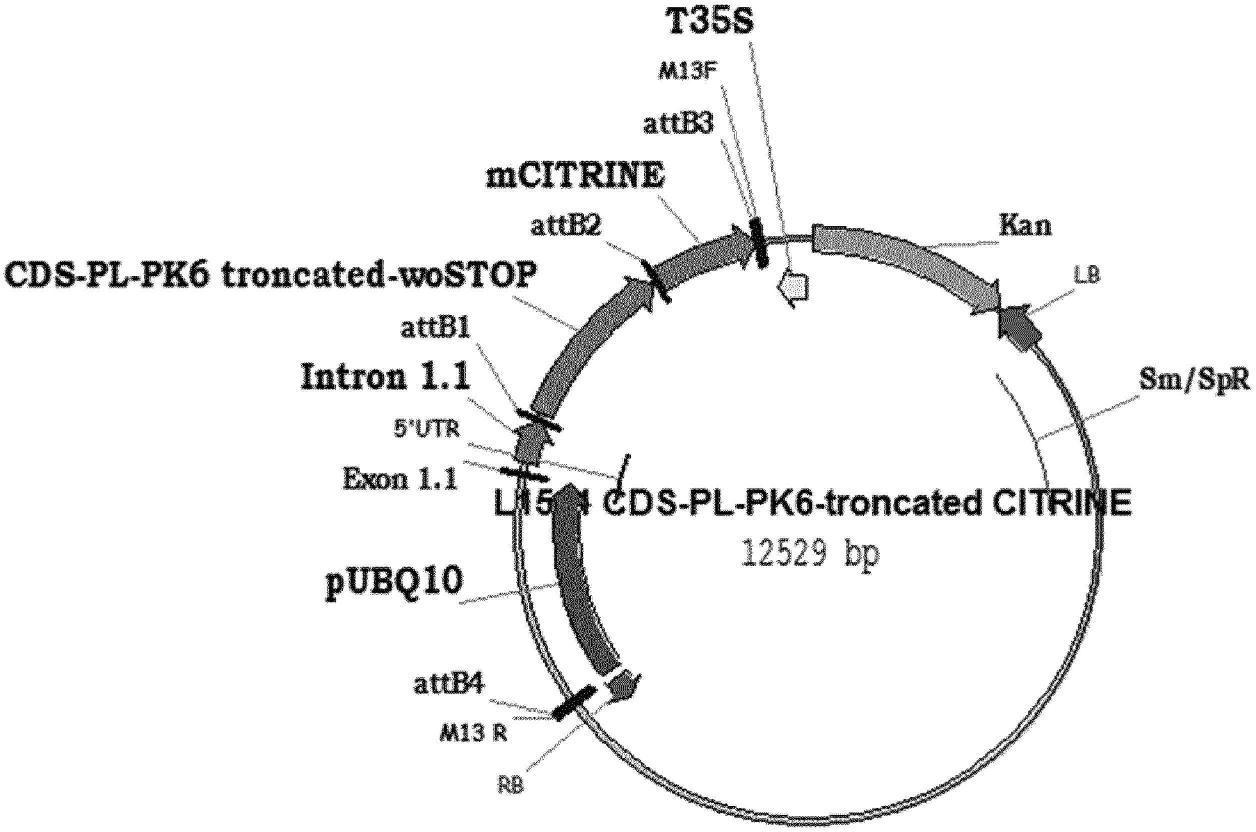

T35S
M13F
attB3
mCITRINE
attB2
Kan
CDS-PL-PK6 troncated-woSTOP
LB
attB1
Intron 1.1
Sm/SpR
5'UTR
Exon 1.1
L15[illegible]4 CDS-PL-PK6-troncated CITRINE
12529 bp
pUBQ10
attB4
M13 R
RB

FIG.17

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/060202

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/82 C07K14/415 A01H5/00 A01H5/08 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)
C07K C12N

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, CHEM ABS Data, WPI Data, BIOSIS, Sequence Search, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2010/083178 A1 (MONSANTO TECHNOLOGY LLC [US]; KOVALIC DAVID K [US]; QIU BO-XING [US];) 22 July 2010 (2010-07-22) the whole document -----	1-7, 21-23
X	WO 2014/151749 A1 (PIONEER HI BRED INT [US]; AYELE MULU [US]; HAYES KEVIN R [US]; SIMMONS) 25 September 2014 (2014-09-25) the whole document ----- -/--	5-7

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

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
25 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. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer
Puonti-Kaerlas, J

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BARRET P ET AL: "A major locus expressed in the male gametophyte with incomplete penetrance is responsible for in situ gynogenesis in maize", THEORETICAL AND APPLIED GENETICS ; INTERNATIONAL JOURNAL OF PLANT BREEDING RESEARCH, SPRINGER, BERLIN, DE, vol. 117, no. 4, 31 May 2008 (2008-05-31), pages 581-594, XP019618403, ISSN: 1432-2242 the whole document -----	1-9, 11-27
Y	ANDREA GALLAVOTTI ET AL: "Positional Cloning in Maize (Zea mays subsp. mays , Poaceae)", APPLICATIONS IN PLANT SCIENCES, vol. 3, no. 1, 1 January 2015 (2015-01-01) , page 1400092, XP055218553, DOI: 10.3732/apps.1400092 the whole document -----	1-9, 11-27
A	DONG X ET AL: "Fine mapping ofqhir1influencing in vivo haploid induction in maize", THEORETICAL AND APPLIED GENETICS, SPRINGER, BERLIN, DE, vol. 126, no. 7, 29 March 2013 (2013-03-29), pages 1713-1720, XP035333157, ISSN: 0040-5752, DOI: 10.1007/S00122-013-2086-9 [retrieved on 2013-03-29] -----	1-9, 11-27
A	BORTIRI E ET AL: "Advances in maize genomics: the emergence of positional cloning", CURRENT OPINION IN PLANT BIOLOGY, QUADRANT SUBSCRIPTION SERVICES, GB, vol. 9, no. 2, 1 April 2006 (2006-04-01), pages 164-171, XP028014917, ISSN: 1369-5266, DOI: 10.1016/J.PBI.2006.01.006 [retrieved on 2006-04-01] -----	1-9, 11-27
A	ANDREA GALLAVOTTI ET AL: "sparse inflorescence1 encodes a monocot-specific YUCCA-like gene required for vegetative and reproductive development in maize", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 105, no. 39, 1 January 2008 (2008-01-01), XP055218572, US ISSN: 0027-8424 -----	1-9, 11-27

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2016/060202

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 10
because they relate to subject matter not required to be searched by this Authority, namely:
essentially biological process for producing plants

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2016/060202

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2010083178 A1	22-07-2010	US 2012017292 A1	19-01-2012
		US 2015113680 A1	23-04-2015
		WO 2010083178 A1	22-07-2010
WO 2014151749 A1	25-09-2014	US 2016017349 A1	21-01-2016
		WO 2014151749 A1	25-09-2014