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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2013 (06.09.2013)

(10) International Publication Number
WO 2013/127809 A1

(51) International Patent Classification:
C07K 14/415 (2006.01) *C12N 15/82* (2006.01)

(21) International Application Number:
PCT/EP2013/053845

(22) International Filing Date:
27 February 2013 (27.02.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12157514.6 29 February 2012 (29.02.2012) EP

(71) Applicant: **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(72) Inventors: **FINCH-SAVAGE, William, Edward**; School of Life Sciences, Wellesbourne Campus, University of Warwick, Wellesbourne CV35 9EF (GB). **MORRIS, Karl**; 4 Colbourne Grove, Leamington Spa CV32 6SL (GB). **BARKER, Guy, Cameron**; Life Sciences, Room 139/ C101, Life Sciences, Warwick University, Wellesbourne CV35 9EF (GB). **BRUGGINK, Tonko, Gerhard**; Syngenta Seeds B.V., Westeinde 62, NL-1601 BK Enkhuizen (NL). **VAN DEN WIJNGAARD, Paul**; Syngenta Seeds B.V., Westeinde 62, NL-1601 BK Enkhuizen (NL).

(74) Agent: **SLAVEN, John, William**; Syngenta International AG, Intellectual Property, WRO-1004-6-22, Schwarzwaldallee 215, CH-4058 Basel (CH).

(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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description (Rules 13bis.4(d)(i) and 48.2(a)(viii))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2013/127809 A1

(54) **Title**: MODULATION OF SEED VIGOR

(57) **Abstract**: The present invention provides a polynucleotide which enables the modulation of the seed vigour, and in particular enhances the seed vigour, and more particularly enables the modification of the speed of germination. A plant seed comprising the said polynucleotide is also provided. A method of producing the plant seed, method for improving the germination and vigour of plant seed, transgenic plant and the use of the polynucleotide of the invention for producing plants growing seeds with improved germination and vigour characteristics are also provided. The invention particularly concerns *Brassica*, more particularly *Brassica oleracea*.

MODULATION OF SEED VIGOR

FIELD OF THE INVENTION

5 The present invention generally relates to a polynucleotide which enables the modulation of the seed vigour, particularly enhances the seed vigour, and more particularly enables the modification of the speed of germination. The invention also relates to a plant seed comprising the said polynucleotide. Moreover, method of producing the plant seed, method for improving the germination and vigour of plant seed, transgenic plant and the 10 use of the polynucleotide of the invention for producing plants growing seeds with improved germination and vigour characteristics are also provided within the present invention. The plants of the present invention particularly concern *Brassica*, more particularly *Brassica oleracea*.

15 BACKGROUND OF THE INVENTION

Seed quality, as defined by the number of uniform usable plants obtainable from a seed batch, is becoming an ever more important trait in developed horticulture markets. Young plant raising is a highly technological activity in these markets and demands with respect 20 to seed quality are therefore high. Reliable and consistently high seed quality is required for this. Also, germination under adverse conditions is an important seed quality trait. This means that one requirement for commercial success of seed varieties is consistent and robust seed quality. Currently however, seed quality of commercial varieties is not always stable and predictable.

25

Seed quality parameters are highly influenced by maternal environmental conditions during seed development. Given the volumes of seed needed and commercial feasibility, only limited controls of these conditions are possible. Therefore consistency of seed quality is limited by the susceptibility for maternal conditions. Research has shown that 30 maternal environment can potentially affect all seed quality parameters, including uniformity and germination under adverse conditions. Decreasing the influence of maternal conditions would therefore lead to more consistent and robust seed quality, that provides advantages in both developed and developing markets.

Predictable and uniform seedling establishment is essential for the production of crops that are both sustainable and profitable. A key contributor to this predictability is the germination performance of seeds, which is influenced directly by seed dormancy and vigour. Dormancy *per se* (lack of germination in generally permissible conditions) is not
5 considered to be a practical problem with many crop species, but low seed vigour (poor seed performance in practice) greatly influences not only the number of seedlings that emerge, but also the timing and uniformity of seedling emergence in all crops. The effects of this have a major impact upon many aspects of crop production that determine cost effectiveness and the inputs required, and there are also direct crop specific influences on
10 marketable yield (Finch-Savage, 1995). Low seed vigour can result from seed deterioration and damage of many kinds and this has great commercial significance. However, there are also inherent differences in the initial vigour of the seed before it begins to deteriorate, but the genetic, molecular and physiological basis of this remains poorly understood.
15

Mutations in many genes have been identified that show phenotypes with altered seed germination performance and these have been instrumental in developing our current understanding of the control of germination (reviewed by Finch-Savage and Leubner-Metzger, 2006; Holdsworth *et al.*, 2008a and 2008b). However, the relative impact of these
20 genes in wild type or crop seeds is little understood and no clear candidates have been revealed that will form the basis of a discriminating test for seed vigour. An alternative source of genetic variation to laboratory induced mutations is available in natural populations and crop genotypes. Using this variation to identify QTL associated with seed vigour and then candidate genes influencing these traits may provide a route to identify
25 practically important genes.

Both natural and crop plant variation has been exploited in quantitative genetic analyses of a range of seed vigour traits in tomato (Foolad *et al.*, 1999), *Brassica oleracea* (Bettey *et al.*, 2000, Finch-Savage *et al.*, 2005) and *Arabidopsis* (Groot *et al.*, 2000, Clerkx *et al.*,
30 2004). Speed of seed germination QTLs have been identified in all three species.
The distinction between dormancy and low seed vigour in healthy non-aged seeds in terms of speed of germination, if one exists, is not understood and may have the same basis (Hilhorst and Toorop, 1997). In most situations, for example in *Arabidopsis*, physiological dormancy is not absolute, but seeds are conditionally dormant *i.e.*

germination tends to be slow and is only possible in a limited range of environments. As dormancy is progressively lost, germination tends to speed up and becomes possible in a wider range of environments and can therefore appear like an increase in vigour.

5 Among the factors accounting for the establishment of seed germination and the regulation of seed dormancy, abscisic acid (ABA), a well-known plant hormone, plays an important role. ABA is in particular essential for the seed germination and seed maturation processes (for review, see Finkelstein *et al.* 2002) as it is responsible for the establishment of a period of seed dormancy. As for buds, it is important that the seeds do not germinate
10 prematurely, for example, during unseasonably mild conditions prior to the onset of winter or a dry season. ABA in the seed enforces this dormancy. The dormancy is lifted only if the seed has been exposed to a prolonged cold spell and/or other appropriate environmental signal and if there is sufficient water to support germination. Besides its role in seed vigour, ABA also regulates many important aspects of plant life including the physiological
15 responses to biotic threats and abiotic stresses like drought and dessication

There is thus a long-standing need for seed with a more reliable and constant seed vigour; especially with a timely-defined and uniform speed of germination, in order to provide seeds that germinate at a more constant rate, independently of the maternal conditions
20 and whatever the external environmental conditions are. Such increased seed vigour would be of particular interest in cases where the seed is coated with a given preparation (chemical, biological), as it is usually observed a delay in seed germination. In consequence, seeds comprising sequences which enhance seed vigour, more particularly enhance the speed and the uniformity of the seed germination would be of primary
25 importance to counteract the effect of the coating treatment, while still applying insecticides and fungicides.
Furthermore, while in many aspects, increasing the seed vigour would be a very useful and desired trait, it appears that, in some cases, decreasing the seed vigour would be of great interest. In particular in viviparous seed, decreasing seed vigour could be beneficial.
30 Vivipary is defined as the germination of the seed while still on the mother plant or before drying and can occur both in immature and fully mature seed. Vivipary has been observed in many different crop species including *Brassica* crops (Ruan *et al.* 2000). Seeds comprising sequences which are able to decrease the seed vigour would thus be capable of delaying, if not removing, the non-desired vivipary phenotype.

SUMMARY OF THE INVENTION

It is therefore to the inventors' credit, in such a state of the art, to have identified a plant from a population that contained a small introgressed region spanning *SOG1 (Speed Of Germination 1)*, a speed of seed germination QTL identified in *B. oleracea* by Bettey *et al.* (2000) and particularly to have demonstrated and identified the corresponding genes that are involved in the modulation of the seed vigour, in particular involved in the regulation of the speed of germination of the seed. It is in particular demonstrated that these genes, and their corresponding sequences, can be used as enabling tools to obtain seeds (or plants that deliver seeds) that exhibit a modified seed vigour phenotype. In particular, gene sequences can be introduced into a new background in order to modulate the seed vigour. More particularly, gene sequences can be used to engineer a novel plant, whose seeds will emerge earlier, more uniformly, independently of the external environmental conditions, and regardless of the maternal conditions. Furthermore, the said gene sequences can be used as tools to modify ABA content in the seed and/or the seed response to ABA, thereby affecting seed behavior regarding to seed dormancy and seed storage proteins and lipids synthesis.

EMBODIMENTS

Accordingly, in a 1st embodiment, the present invention provides a polynucleotide, particularly an isolated polynucleotide, comprising a nucleic acid molecule selected from
5 the group consisting of

a. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 1 (A12 version of BolC.VG1.a);

b. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 2 (A12 version of BolC.VG2.a);

c10 a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 3 (GD33 version of BolC.VG1.a);

d. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 4 (GD33 version of BolC.VG2.a);

e. a nucleic acid molecule comprising a nucleotide sequence the complementary strand of
15 which hybridizes to the nucleic acid molecule of any of a) – d);

f. a nucleic acid molecule comprising a nucleotide sequence that deviates from the nucleotide sequence defined in any of a) – e) by the degeneracy of the genetic code;

wherein said nucleic acid molecule as defined in any of a) – f) upon expression in a plant or plant part, leads to a modified seed vigour.

20 In a 2nd embodiment, the present invention provides a polynucleotide, particularly an isolated polynucleotide, according to embodiment 1, comprising a nucleic acid molecule selected from the group further consisting of

a. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 5 (truncated A12 allele of BolC.VG2.b);

25b. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 6 (truncated GD33 allele of BolC.VG2.b);

c. a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) – b);

d. a nucleic acid molecule comprising a nucleotide sequence that deviates from the
30 nucleotide sequence defined in any of a) – c) by the degeneracy of the genetic code;

wherein said nucleic acid molecule as defined in any of a) – d) upon expression in a plant or plant part, leads to a modified seed vigour.

In a 3rd embodiment, the present invention provides a polynucleotide, particularly an isolated polynucleotide, comprising a nucleic acid molecule selected from the group
5 consisting of

a. nucleic acid molecule comprising a nucleotide sequence that has at least 60% sequence identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

b. nucleic acid molecule comprising a nucleotide sequence that has at least 80% sequence
10 identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

c. nucleic acid molecule comprising a nucleotide sequence that has at least 90% sequence identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

15 d. nucleic acid molecule comprising a nucleotide sequence that has at least 95% sequence identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

e. nucleic acid molecule comprising a nucleotide sequence that has at least 98% sequence identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ
20 ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

f. nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) – e);

g. nucleic acid molecule comprising a nucleotide sequence that deviates from the nucleotide sequence defined in any of a) – f) by the degeneracy of the genetic code;

25 wherein said nucleic acid molecule as defined in any of a) – g) upon expression in a plant or plant part, leads to a modified seed vigour.

In a 4th embodiment, the present invention relates to a polynucleotide according to first, second or third embodiment, wherein the modified seed vigour phenotype is characterized by a further phenotype selected in the group comprising : modified speed of germination,
30 modified speed of seedling emergence, modified uniformity of seed germination, modified

uniformity of seedling emergence, modified percentage of seed germination, modified tolerance of the seed vis-a-vis external environmental and/or maternal conditions, modified sensitivity to ABA or modified content of ABA.

In a 5th embodiment, the present invention relates to a polynucleotide, particularly an
5 isolated polynucleotide, comprising a nucleic acid molecule selected in the group comprising:

a. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 3;

b. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO:
10 4;

c. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 6;

d. a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) – c) ;

15 e. a nucleic acid molecule comprising a nucleotide sequence that deviates from the nucleotide sequence defined in any of a) – d) by the degeneracy of the genetic code;

wherein said nucleic acid molecule as defined in any of a) – e) upon expression in a plant or plant part, leads to an increased seed vigour.

In the context of the present invention, the expression “increased seed vigour” means that
20 the speed of germination of the seed may be modified in the sense that the seed germinates faster and is thus more vigorous. In one particular embodiment the speed of germination is increased allowing seeds to be obtained that germinates faster thanks to the presence of a polynucleotide according to the present invention in its genome as compared to a seed that does not comprise said polynucleotide. Such an increased speed
25 of germination results in a significant earlier seedling emergence and thus provides the seeds with enhanced flexibility and better adaptation to various environments. The 2 genes that were identified were found to have a significant germination phenotype indicating these genes are regulators of germination to different extent. The allele corresponding to the polynucleotide of SEQ ID NO: 3 and SEQ ID NO: 4 (GD33 alleles of
30 the genes) according to the present invention has been proven to correspond to

SUBSTITUTE SHEET (RULE 26)

phenotype of increased seed vigour and thus allow to modify the speed of germination of plant seed in which it is introgressed into towards and increased speed of germination.

In a 6th embodiment, the present invention relates to a polynucleotide according to embodiment 5, wherein the increased seed vigour phenotype is characterized by a further phenotype selected in the group comprising: increased speed of germination, increased speed of seedling emergence, increased uniformity of seed germination, increased uniformity of seedling emergence, increased percentage of seed germination, increased tolerance of the seed vis-a-vis external environmental and/or maternal conditions, decreased sensitivity to ABA or decreased content of ABA.

In a 7th embodiment, the present invention relates to a polynucleotide, particularly an isolated polynucleotide, comprising a nucleic acid molecule selected in the group comprising:

a. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 1;

b. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 2;

c. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 5;

d. a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) to c) ;

e. a nucleic acid molecule comprising a nucleotide sequence that deviates from the nucleotide sequence defined in any of a) – d) by the degeneracy of the genetic code;

wherein said nucleic acid molecule as defined in any of a) – e) upon expression in a plant or plant part, leads to a decreased seed vigour.

In an 8th embodiment, the present invention provides an expression cassette comprising a polynucleotide of any of the preceding embodiments.

In a 9th embodiment, the present invention provides a vector molecule comprising the expression cassette according to the 8th embodiment.

In a 10th embodiment, the present invention relates to the use of a polynucleotide according to embodiments 1 to 3 for modifying seed vigour.

In a 11th embodiment, the present invention relates to a method for modifying the seed vigour comprising introgressing through crossing or by plant transformation techniques to

SUBSTITUTE SHEET (RULE 26)

and expressing in a plant or plant part a polynucleotide, an expression cassette or a vector molecule of any of embodiments 1 to 10.

In a 12th embodiment, the present invention relates to a method for producing seed with modified seed vigour comprising:

5 a. obtaining a first plant verified to contain the polynucleotide of any of embodiments 1 to 7;
b. crossing said first plant with a second plant verified to lack the said polynucleotide ; and
c. identifying a plant seed resulting from the cross exhibiting a modified seed vigour as compared to seeds delivered by the second plant.

In a 13th embodiment, the present invention relates to a plant or plant part which contains
10 within its genome an introgression comprising the polynucleotide, the expression cassette or the vector molecule of any of embodiments 1 to 9 and exhibits a modification of seed vigour as compared to a plant or plant part that does not comprise the said polynucleotide, expression cassette or vector molecule.

In a 14th embodiment, the present invention relates to a plant or plant part which contains
15 within its genome an introgression comprising a polynucleotide according to embodiment 5 and exhibits an increased seed vigour as compared to a seed delivered by a plant or plant part that does not comprise the said polynucleotide.

In a 15th embodiment, the present invention relates to a plant or plant part which contains within its genome an introgression comprising a polynucleotide according to embodiment 7
20 and exhibits a decreased seed vigour as compared to a seed delivered by a plant or plant part that does not comprise the said polynucleotide.

In a 16th embodiment, the invention provides a method for selecting plant or plant part with modified seed vigour, comprising the detection in the plant or plant part to be tested of the presence or absence of a polynucleotide according to any of embodiments 1 to 7.

25 In a 17th embodiment, the invention provides a method for selecting plant or plant parts with modified seed vigour, comprising contacting candidate plant or plant part with a selection tool selected from the group comprising the polynucleotides of any of embodiments 1 to 7.

In a 18th embodiment, the present invention relates to plant or plant part according to any
30 of the preceding claims that are cultivated plant or cultivated plant part and are selected in the group comprising *Brassica oleracea*, *Brassica napus*, *Brassica rapa, Brassica campestris, Brassica juncea, Brassica nigra, Brassica pekinensis, Brassica chinensis,*

Brassica rosularis, Eruca vesicaria, Eruca sativa, Raphanus sativus, Lepidium sativum, Nasturtium officinale, Wasabia japonica.

In a 19th embodiment, the invention provides a non biological method for obtaining plant or plant part with modified seed vigour, comprising introducing a polynucleotide according to any of embodiments 1 to 7 into the genome of said plant or plant part.

In an embodiment 20, the invention relates to the method according to embodiment N°16 comprising (a) obtaining a first plant verified to contain the polynucleotide of any of embodiments 1 to 7; (b) crossing said first plant with a second plant verified to lack the said polynucleotide; and (c) identifying a plant resulting from the cross exhibiting modified seed vigour, and containing the said polynucleotide.

In an embodiment 21, the invention relates to the method according to embodiment 20, wherein presence of the polynucleotide is verified by use of a molecular marker, particularly by a molecular marker physically located in a position that is within or outside the genetic locus containing the polynucleotide.

In an embodiment 22, the invention relates to the method according to embodiment 20, wherein presence of the polynucleotide is verified by use of at least two molecular markers, particularly by at least two molecular markers physically located in a position that is flanking the genetic locus containing the polynucleotide.

In an embodiment 23, the invention relates to a seed comprising a polynucleotide according to embodiments 1 to 7, wherein the said seed is coated with any type of coating.

In an embodiment 24, the invention relates to a plant or plant part according to embodiments 13 to 18, wherein the plant is a hybrid plant.

In an embodiment 25, the invention relates to a plant or plant part according to embodiment 24 obtainable from seed deposited at NCIMB under deposit number NCIMB 41951, or progeny thereof.

DETAILED DESCRIPTION OF THE INVENTION

The polynucleotides sequences according to the previous embodiment have been identified in and isolated from a *Brassica oleracea* plant with different seed vigour phenotypes. Alleles with the polynucleotide sequences according to the present invention were introgressed into a plant with different seed quality and different seed vigour. Thanks to the introgression of any of the polynucleotide sequence according to the present invention, seed vigour is significantly modified and consequently plant or plant part quality, particularly seed quality of the plant was significantly modified, particularly seed vigour, as highlighted by the speed of germination. In one embodiment the seed was able to germinate better and quicker, particularly in cold conditions, showing more robust seed quality. The introgression of a polynucleotide sequence according to embodiment 5 allows to obtain seed quality of the plant less susceptible to seed production environment. This means that the introgression of the polynucleotide sequences according to the present invention enables delivery of consistent seed quality from conventional seed production.

Introduction of the polynucleotide sequences according to the present invention in commercial cultivated plant, particularly cultivated *Brassica* plants, more particularly cultivated *Brassica oleracea* plants, will ensure more reliable production of seed with sufficient seed quality. This will significantly increase operational flexibility.

Hybrid seeds that comprise any of the polynucleotide sequences according to the present invention exhibit a more uniform stand establishment, especially under adverse conditions. This will definitively add value to the commercial seed product.

Seed Deposit details

The following seed samples were deposited with NCIMB, Ferguson Building, Craibstone Estate, Bucksburn, Aberdeen AB21 9YA, Scotland, UK, on April 4th, 2012 under the provisions of the Budapest Treaty in the name of Syngenta Participations AG:

NCIMB 41950 Brassica oleracea A12
NCIMB 41951 Brassica oleracea SL101

Definitions

The technical terms and expressions used within the scope of this application are
5 generally to be given the meaning commonly applied to them in the pertinent art of genetic engineering in plants, plant breeding and cultivation if not otherwise indicated herein below.

As used in this specification and the appended claims, the singular forms "a", "an", and
10 "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a plant" includes one or more plants, and reference to "a cell" includes mixtures of cells, tissues, and the like.

A "cultivated plant", particularly "cultivated *Brassica* plant", more particularly "cultivated *Brassica oleracea* plant" is understood within the scope of the invention to refer to a plant
15 that is no longer in the natural state but has been developed by human care and for human use and/or growing purposes and/or consumption. "Cultivated plant", particularly "cultivated *Brassica* plant", more particularly "cultivated *Brassica oleracea* plant" are further understood to exclude those wild-type species which comprise the trait being subject of this invention as a natural trait and/or part of their natural genetics.

20 An "allele" is understood within the scope of the invention to refer to alternative or variant forms of various genetic units identical or associated with different forms of a gene or of any kind of identifiable genetic element, which are alternative in inheritance because they are situated at the same locus in homologous chromosomes. Such alternative or variant forms may be the result of single nucleotide polymorphisms, insertions, inversions,
25 translocations or deletions, or the consequence of gene regulation caused by, for example, by chemical or structural modification, transcription regulation or post-translational modification/regulation. In a diploid cell or organism, the two alleles of a given gene or genetic element typically occupy corresponding loci on a pair of homologous chromosomes.

30 A "truncated" allele of a gene is meant within the present invention to represent an allele of a gene that has lost single or multiple partial nucleotide sequences when compared with its full-length gene allele counterpart.

An allele associated with a qualitative trait may comprise alternative or variant forms of various genetic units including those that are identical or associated with a single gene or multiple genes or their products or even a gene disrupting or controlled by a genetic factor contributing to the phenotype represented by the locus.

5 As used herein, the term "marker allele" refers to an alternative or variant form of a genetic unit as defined herein above, when used as a marker to locate genetic loci containing alleles on a chromosome that contribute to variability of phenotypic traits.

As used herein, the term "breeding", and grammatical variants thereof, refer to any process that generates a progeny individual. Breedings can be sexual or asexual, or any 10 combination thereof. Exemplary non-limiting types of breedings include crossings, selfings, doubled haploid derivative generation, and combinations thereof, which all are known techniques to the person skilled in the art.

"Backcrossing" is understood within the scope of the invention to refer to a process in which a hybrid progeny is repeatedly backcrossed back to one of the parents. Different 15 recurrent parents may be used in subsequent backcrosses. Recombinant lines can be produced by selfing the offspring resulting from backcrossing.

"Locus" is understood within the scope of the invention to refer to a region on a chromosome, which comprises a gene or any other genetic element or factor contributing to a trait.

20 "Introgression" (or introgressed) is understood within the scope of the present invention to refer to a movement of gene or segment(s) of nucleic acid from a species into the gene pool of another species, or from a line into the gene pool of another line within the same species. Introgression may be achieved by sexual crossing, sexual hybridization or by genetic transformation.

25 As used herein, "marker locus" refers to a region on a chromosome, which comprises a nucleotide or a polynucleotide sequence that is present in an individual's genome and that is associated with one or more loci of interest, which may which comprise a gene or any other genetic element or factor contributing to a trait. "Marker locus" also refers to a region on a chromosome, which comprises a polynucleotide sequence complementary to a 30 genomic sequence, such as a sequence of a nucleic acid used as probes.

For the purpose of the present invention, the term "segregation" or "co-segregation" refers to the fact that the allele for the trait and the allele(s) for the marker(s) tend to be

transmitted together because they are physically close together on the same chromosome (reduced recombination between them because of their physical proximity) resulting in a non-random association of their alleles as a result of their proximity on the same chromosome. "Co-segregation" also refers to the presence of two or more traits within a single plant of which at least one is known to be genetic and which cannot be readily explained by chance.

As used herein, the phrase "genetic marker" or "marker" refers to a feature of an individual's genome (*e.g.*, a nucleotide or a polynucleotide sequence that is present in an individual's genome) that is associated with one or more loci of interest. In some embodiments, a genetic marker is polymorphic in a population of interest, or the locus occupied by the polymorphism, depending on context. Genetic markers include, for example, single nucleotide polymorphisms (SNPs), indels (*i.e.*, insertions/deletions), simple sequence repeats (SSRs), restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), cleaved amplified polymorphic sequence (CAPS) markers, Diversity Arrays Technology (DArT) markers, and amplified fragment length polymorphisms (AFLPs), among many other examples. Genetic markers can, for example, be used to locate genetic loci containing alleles on a chromosome that contribute to variability of phenotypic traits. The phrase "genetic marker" or "marker" can also refer to a polynucleotide sequence complementary to a genomic sequence, such as a sequence of a nucleic acid used as probes.

A "genetic marker" or "marker" can be physically located in a position on a chromosome that is within or outside of to the genetic locus with which it is associated (*i.e.*, is intragenic or extragenic, respectively). Stated another way, whereas "genetic markers" or "markers" are typically employed when the location on a chromosome of the gene or of a functional mutation, e.g. within a control element outside of a gene, that corresponds to the locus of interest has not been identified and there is a non-zero rate of recombination between the "genetic marker" or "marker" and the locus of interest, the presently disclosed subject matter can also employ "genetic markers" or "markers" that are physically within the boundaries of a genetic locus (*e.g.*, inside a genomic sequence that corresponds to a gene such as, but not limited to a polymorphism within an intron or an exon of a gene). In some embodiments of the presently disclosed subject matter, the one or more "genetic markers" or "markers" comprise between one and ten markers, and in some embodiments the one or more genetic markers comprise more than ten genetic markers.

As used herein, the term "genotype" refers to the genetic constitution of a cell or organism. An individual's "genotype for a set of genetic markers" includes the specific alleles, for one or more genetic marker loci, present in the individual's haplotype. As is known in the art, a genotype can relate to a single locus or to multiple loci, whether the loci are related or
5 unrelated and/or are linked or unlinked. In some embodiments, an individual's genotype relates to one or more genes that are related in that the one or more of the genes are involved in the expression of a phenotype of interest. Thus, in some embodiments a genotype comprises a summary of one or more alleles present within an individual at one or more genetic loci of a quantitative trait. In some embodiments, a genotype is expressed
10 in terms of a haplotype.

As used herein, the term "linkage", and grammatical variants thereof, refers to the tendency of alleles at different loci on the same chromosome to segregate together more often than would be expected by chance if their transmission were independent, in some embodiments as a consequence of their physical proximity.

15 For example, "polynucleotide sequence" as used herein refers to all forms of naturally occurring or recombinantly generated types of nucleic acids and/or nucleotide sequences as well as to chemically synthesized nucleic acids/nucleotide sequences. This term also encompasses nucleic acid analogs and nucleic acid derivatives such as, e. g., locked DNA, RNA, cDNA, PNA, oligonucleotide thiophosphates and substituted ribo-
20 oligonucleotides. Furthermore, the term "polynucleotide sequence" also refers to any molecule that comprises nucleotides or nucleotide analogs. The phrase "nucleic acid" or "polynucleotide" refers to any physical string of monomer units that can be corresponded to a string of nucleotides, including a polymer of nucleotides (*e.g.*, a typical DNA, cDNA or RNA polymer), optionally containing synthetic, non-natural or altered nucleotide bases
25 capable of incorporation into DNA or RNA polymers, modified oligonucleotides (*e.g.*, oligonucleotides comprising bases that are not typical to biological RNA or DNA, such as 2'-*O*-methylated oligonucleotides), and the like. The term "polynucleotide" is understood herein to refer to polymeric molecule of high molecular weight which can be single-stranded or double-stranded, multi-stranded, or combinations thereof, composed of
30 monomers (nucleotides) containing a sugar, phosphate and a base which is either a purine or pyrimidine. Unless otherwise indicated, a particular nucleic acid sequence of the presently disclosed subject matter optionally comprises or encodes complementary sequences, in addition to any sequence explicitly indicated.

Preferably, the term "polynucleotide sequence" refers to a nucleic acid molecule, i.e.

deoxyribonucleic acid (DNA) and/ or ribonucleic acid (RNA). The "polynucleotide sequence" in the context of the present invention may be made by synthetic chemical methodology known to one of ordinary skill in the art, or by the use of recombinant technology, or may be isolated from natural sources, or by a combination thereof. The
5 DNA and RNA may optionally comprise unnatural nucleotides and may be single or double stranded. "Polynucleotide sequence" also refers to sense and anti-sense DNA and RNA, that is, a polynucleotide sequence which is complementary to a specific sequence of nucleotides in DNA and/or RNA. Furthermore, the term "polynucleotide sequence" may refer to DNA or RNA or hybrids thereof or any modification thereof that is known in the
10 state of the art (see, e.g., US 5525711, US 4711955, US 5792608 or EP 302175 for examples of modifications). The polynucleotide sequence may be single- or double-stranded, linear or circular, natural or synthetic, and without any size limitation. For instance, the polynucleotide sequence may be genomic DNA, cDNA, mRNA, antisense RNA, ribozymal or a DNA encoding such RNAs or chimeroplasts (Gamper, Nucleic Acids
15 Research, 2000, 28, 4332 – 4339). Said polynucleotide sequence may be in the form of a plasmid or of viral DNA or RNA. "Polynucleotide sequence" may also refer to (an) oligonucleotide(s), wherein any of the state of the art modifications such as phosphothioates or peptide nucleic acids (PNA) are included.

A "polynucleotide fragment" is a fraction of a given polynucleotide molecule or of a
20 "polynucleotide sequence". In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. A "genome" is the entire body of genetic material contained in each cell of an organism.

Unless otherwise indicated, a particular nucleic acid sequence of this invention also
25 implicitly encompasses conservatively modified variants thereof (e.g. degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, 1991; Ohtsuka *et al.*, 1985;
30 Rossolini *et al.*, 1994). The term polynucleotide is used interchangeably with nucleic acid, nucleotide sequence and may include genes, cDNAs, and mRNAs encoded by a gene, etc.

The polynucleotide of the invention is understood to be provided in isolated form. The term

"isolated" means that the polynucleotide disclosed and claimed herein is not a polynucleotide as it occurs in its natural context, if it indeed has a naturally occurring counterpart. Accordingly, the other compounds of the invention described further below are understood to be isolated. If claimed in the context of a plant genome, the
5 polynucleotide of the invention is distinguished over naturally occurring counterparts by the insertion side in the genome and the flanking sequences at the insertion side.

As used herein, the term "gene" refers to any segment of nucleic acid associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. For example, gene refers to a nucleic acid
10 fragment that expresses mRNA or functional RNA, or encodes a specific protein, and which includes regulatory sequences. Genes also include non-expressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have
15 desired parameters.

"Marker-based selection" is understood within the scope of the invention to refer to e.g. the use of genetic markers to detect one or more nucleic acids from the plant, where the nucleic acid is associated with a desired trait to identify plants that carry genes for
20 desirable (or undesirable) traits, so that those plants can be used (or avoided) in a selective breeding program. A "marker gene" encodes a selectable or screenable trait.

Suitable markers used within the invention may, for example, be selected from the group consisting of single nucleotide polymorphism (SNP) markers, indel (*i.e.*, insertions/deletions) markers, simple sequence repeat (SSR) markers, restriction fragment
25 length polymorphism (RFLP) markers, random amplified polymorphic DNA (RAPD) markers, cleaved amplified polymorphic sequence (CAPS) markers, Diversity Arrays Technology (DArT) markers, and amplified fragment length polymorphisms (AFLPs) markers.

For example, RFLP involves the use of restriction enzymes to cut chromosomal DNA at
30 specific short restriction sites, polymorphisms result from duplications or deletions between the sites or mutations at the restriction sites.

RAPD utilizes low stringency polymerase chain reaction (PCR) amplification with single primers of arbitrary sequence to generate strain-specific arrays of anonymous DNA fragments. The method requires only tiny DNA samples and analyses a large number of polymorphic loci.

5 AFLP requires digestion of cellular DNA with a restriction enzyme(s) before using PCR and selective nucleotides in the primers to amplify specific fragments. With this method, using electrophoresis techniques to visualize the obtained fragments, up to 100 polymorphic loci can be measured per primer combination and only small DNA sample are required for each test.

10 SSR analysis is based on DNA micro-satellites (short-repeat) sequences that are widely dispersed throughout the genome of eukaryotes, which are selectively amplified to detect variations in simple sequence repeats. Only tiny DNA samples are required for an SSR analysis. SNPs use PCR extension assays that efficiently pick up point mutations. The procedure requires little DNA per sample. One or two of the above methods may be used 15 in a typical marker-based selection breeding program.

The most preferred method of achieving amplification of nucleotide fragments that span a polymorphic region of the plant genome employs the polymerase chain reaction ("PCR") (Mullis *et al.*, Cold Spring Harbor Symp. Quant. Biol. 51:263 273 (1986)), using primer pairs involving a forward primer and a backward primer that are capable of hybridizing to 20 the proximal sequences that define a polymorphism in its double-stranded form. As disclosed herein, such primers may be used for fine mapping, map-based cloning and for expression analysis.

"Microsatellite or SSRs (Simple sequence repeats) Marker" is understood within the scope of the invention to refer to a type of genetic marker that consists of numerous repeats of 25 short sequences of DNA bases, which are found at loci throughout the plant's genome and have a likelihood of being highly polymorphic.

"PCR (Polymerase chain reaction)" is understood within the scope of the invention to refer to a method of producing relatively large amounts of specific regions of DNA or subset(s) of the genome, thereby making possible various analyses that are based on those regions.

30 "PCR primer" is understood within the scope of the invention to refer to relatively short fragments of single-stranded DNA used in the PCR amplification of specific regions of DNA.

"Phenotype" is understood within the scope of the invention to refer to a distinguishable characteristic(s) of a genetically controlled trait.

As used herein, the phrase "phenotypic trait" refers to the appearance or other detectable characteristic of an individual, resulting from the interaction of its genome, proteome
5 and/or metabolome with the environment.

"Polymorphism" is understood within the scope of the invention to refer to the presence in a population of two or more different forms of a gene, genetic marker, or inherited trait or a gene product obtainable, for example, through alternative splicing, DNA methylation, etc.

"Probe" as used herein refers to a group of atoms or molecules which is capable of
10 recognising and binding to a specific target molecule or cellular structure and thus allowing detection of the target molecule or structure. Particularly, "probe" refers to a labelled DNA or RNA sequence which can be used to detect the presence of and to quantify a complementary sequence by molecular hybridization.

Such polynucleotide sequences being capable of hybridizing may be identified and
15 isolated by using the polynucleotide sequences described herein or parts or reverse complements thereof, for instance by hybridization according to standard methods (see for instance Sambrook and Russell (2001), Molecular Cloning: A Laboratory Manual, CSH Press, Cold Spring Harbor, NY, USA). Nucleotide sequences comprising the same or substantially the same nucleotide sequences as indicated in the listed SEQ ID NO 1, 2 or
20 3, or parts/fragments thereof, can, for instance, be used as hybridization probes. The fragments used as hybridization probes can also be synthetic fragments which are prepared by usual synthesis techniques, the sequence of which is substantially identical with that of a nucleotide sequence according to the invention.

The term "hybridize" as used herein refers to conventional hybridization conditions,
25 preferably to hybridization conditions at which 5xSSPE, 1% SDS, 1xDenhardts solution is used as a solution and/or hybridization temperatures are between 35°C and 70°C, preferably 65°C. After hybridization, washing is preferably carried out first with 2xSSC, 1% SDS and subsequently with 0.2xSSC at temperatures between 35°C and 75°C, particularly between 45°C and 65°C, but especially at 59°C (regarding the definition of
30 SSPE, SSC and Denhardts solution see Sambrook *et al.* loc. cit.). High stringency hybridization conditions as for instance described in Sambrook et al, supra, are particularly preferred. Particularly preferred stringent hybridization conditions are for instance present

if hybridization and washing occur at 65°C as indicated above. Non-stringent hybridization conditions for instance with hybridization and washing carried out at 45°C are less preferred and at 35°C even less.

"Sequence Homology or Sequence Identity" is used herein interchangeably. The terms
5 "identical" or percent "identity" in the context of two or more nucleic acid or protein sequences, refer to two or more sequences or sub-sequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. For example, this term
10 is used herein in the context of a nucleotide sequence which has a homology, that is to say a sequence identity, of at least 50%, 55%, 60%, preferably of at least 70%, 75% more preferably of at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94, 95%, 96%, 97%, 98%, and even most preferably of at least 99% to another, preferably entire, nucleotide sequence.

15 If two sequences which are to be compared with each other differ in length, sequence identity preferably relates to the percentage of the nucleotide residues of the shorter sequence which are identical with the nucleotide residues of the longer sequence. As used herein, the percent identity/homology between two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/
20 total # of positions x 100), taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm, as described herein below. For example, sequence identity can be determined conventionally with the use of computer
25 programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive Madison, WI 53711). Bestfit utilizes the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2 (1981), 482-489, in order to find the segment having the highest sequence identity between two sequences. When using Bestfit or another
30 sequence alignment program to determine whether a particular sequence has for instance 95% identity with a reference sequence of the present invention, the parameters are preferably so adjusted that the percentage of identity is calculated over the entire length of the reference sequence and that homology gaps of up to 5% of the total number of the

nucleotides in the reference sequence are permitted. When using Bestfit, the so-called optional parameters are preferably left at their preset ("default") values. The deviations appearing in the comparison between a given sequence and the above-described sequences of the invention may be caused for instance by addition, deletion, substitution,
5 insertion or recombination. Such a sequence comparison can preferably also be carried out with the program "fasta20u66" (version 2.0u66, September 1998 by William R. Pearson and the University of Virginia; see also W.R. Pearson (1990), Methods in Enzymology 183, 63-98, appended examples and http://workbench.sdsc.edu/). For this purpose, the "default" parameter settings may be used.

10 Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase: "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s) substantially" refers to
15 complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern
20 hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid
25 probe assays" Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5° C lower than the thermal melting point for the specific sequence at a defined ionic strength and pH. Typically, under "stringent conditions" a probe will hybridize to its target subsequence, but to no other sequences.

The thermal melting point is the temperature (under defined ionic strength and pH) at
30 which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the melting temperature (T.sub.m) for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern

or northern blot is 50% formamide with 1 mg of heparin at 42°C., with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.1 5M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2 times SSC wash at 65°C for 15 minutes (see, Sambrook, infra, for a description of SSC buffer). Often, 5 a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1 times SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6 times SSC at 40°C for 15 minutes. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt 10 concentrations of less than about 1.0M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2 times (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates 15 detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g. when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

A "plant" is any plant at any stage of development, particularly a seed plant.

20 A "plant part" as used herein refers to structural and/or functional sub-units of a plant including, but not limited to, plant cells, plant tissues, plant material, plant organs, harvestable plant parts, etc..., as defined herein below.

A "harvestable plant part" is a part of a plant refers to those parts of the plant that are harvested at any suitable time and may be further processed for industrial use or 25 consumption including flowers, fruits, leafs, seeds, fibres, etc.

A "plant cell" is a structural and physiological unit of a plant, comprising a protoplast and a cell wall. The plant cell may be in form of an isolated single cell or a cultured cell, or as a part of higher organized unit such as, for example, plant tissue, a plant organ, or a whole plant.

30 "Plant cell culture" means cultures of plant units such as, for example, protoplasts, cell culture cells, cells in plant tissues, pollen, pollen tubes, ovules, embryo sacs, zygotes and embryos at various stages of development.

"Plant material" or "plant material obtainable from a plant" refers to leaves, stems, roots, flowers or flower parts, fruits, pollen, egg cells, zygotes, seeds, cuttings, cell or tissue cultures, or any other part or product of a plant.

A "plant organ" is a distinct and visibly structured and differentiated part of a plant such as 5 a root, stem, leaf, flower bud, or embryo.

"Plant tissue" as used herein means a group of plant cells organized into a structural and functional unit. Any tissue of a plant in planta or in culture is included. This term includes, but is not limited to, whole plants, plant organs, plant seeds, tissue culture and any groups of plant cells organized into structural and/or functional units. The use of this term in 10 conjunction with, or in the absence of, any specific type of plant tissue as listed above or otherwise embraced by this definition is not intended to be exclusive of any other type of plant tissue.

As used herein, the term "population" means a genetically heterogeneous collection of plants sharing a common genetic derivation.

15 As used herein, the term "variety" or "cultivar" means a group of similar plants that by structural features and performance can be identified from other varieties within the same species.

As used herein, the term "plant transformation techniques" relates to the introduction of a transgene, conferring a specific trait, into the host plant. The transgene is incorporated into 20 the host plant genome and stably inherited through future generations. The correct regulatory sequences are added to the gene of interest i.e. promoters and terminators, then the DNA is transferred to the plant cell culture using an appropriate vector. In some embodiments, the gene is attached to a selectable marker which allows selection for the presence of the transgene as described herein above. Once the plant tissue has been 25 transformed, the cells containing the transgene are selected and regeneration back into whole plants is carried out.

Plant transformation can be carried out in a number of different ways for generating plant hosts with delegated functions required making plant host competent for expression of the sequence of interest, depending on the species of plant in question. For example, 30 *Agrobacterium* mediated transformation may be used to transform plants according to the invention. Within this transformation method, plant or plant tissue (*e.g.* leaves) is cut into small pieces, e.g. 10x10mm, and soaked for 10 minutes in a fluid containing suspended

Agrobacterium containing a Ti-plasmid vector carried by *Agrobacterium* (US 5,591,616; US 4,940,838; US 5,464,763). Placed on selectable rooting and shooting media, the plants will regrow.

Non-*Agrobacterium* techniques involve the uptake of exogenous genetic material directly
5 by protoplasts or cells. These techniques include, but are not limited to, PEG or electroporation mediated uptake, particle bombardment-mediated delivery and microinjection. Examples of these techniques are described in Paszkowski *et al.*, EMBO J 3, 2717-2722 (1984), Potrykus *et al.*, Mol. Gen. Genet. 199,169-177 (1985), Reich *et al.*, Biotechnology 4:1001-1004 (1986), and Klein *et al.*, Nature 327,70-73 (1987). In each
10 case, the transformed cells are regenerated to whole plants using standard techniques. For example, within particle bombardment particles of gold or tungsten are coated with DNA and then shot into young plant cells or plant embryos (US 05100792; EP 00444882B1; EP 00434616B1). Some genetic material will stay in the cells and transform them. This method may also be used to transform plant plastids. Further, electroporation
15 technique may be used to transform plants according to the invention. During electroporation, transient holes are prepared in cell membranes using electric shock allowing DNA to enter the cell. Another transformation technique within the invention may be viral transformation (transduction). Here, the desired genetic material is packed into a suitable plant virus and allow this modified virus to infect the plant. If the genetic material is
20 DNA, it can recombine with the chromosomes to produce transformed cells. However, genomes of most plant viruses consist of single stranded RNA which replicates in the cytoplasm of infected cell.

The transformation or genetically engineering of the plant or plant cell with a nucleotide sequence or the vector according to the invention can be carried out by standard methods,
25 as for instance described in Sambrook and Russell (2001), Molecular Cloning: A Laboratory Manual, CSH Press, Cold Spring Harbor, NY, USA. Moreover, the transgenic plant cell of the present invention is cultured in nutrient media meeting the requirements of the particular cell used, in particular in respect of the pH value, temperature, salt concentration, aeration, antibiotics, vitamins, trace elements etc.

30 The term “vector” or “vector molecule”, as used herein, may comprise an expression cassette which may comprise expression control sequences operably linked to said polynucleotide. The vector(s) may be in the form of a plasmid, and can be used alone or in combination with other plasmids, to provide transformed plants, using transformation

methods as described below to incorporate transgenes into the genetic material of the plant(s). Further vectors may comprise cosmids, viruses, bacteriophages and other vectors commonly used in genetic engineering.

A "transgene" refers to a gene that has been introduced into the genome by transformation
5 and is stably maintained. Transgenes may include, for example, genes that are either heterologous or homologous to the genes of a particular plant to be transformed. Additionally, transgenes may comprise native genes inserted into a non-native organism, or chimeric genes.

"Coding sequence" refers to a DNA or RNA sequence that codes for a specific amino acid
10 sequence and excludes the non-coding sequences. It may constitute an "uninterrupted coding sequence", *i.e.*, lacking an intron, such as in a cDNA or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a sequence of RNA which is contained in the primary transcript but which is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.
15 The term "expression cassette", as used herein, may be made up of one or more nucleotide sequences of the present invention in operable linkage with regulatory nucleotide sequences controlling their expression. As known in the art, an expression cassette may consist of a promoter sequence (promoter), an open reading frame or a functional part thereof, a 3' untranslated region and a terminator sequence (terminator).
20 The cassette may be part of a vector molecule as described herein above. Different expression cassettes can be transformed into plant or plant cells as long as the correct regulatory sequences are used.

As used herein, the term "promoter" refers to a nucleotide sequence, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing
25 the recognition for RNA polymerase and other factors required for proper transcription. A "promoter" includes a minimal promoter that is a short DNA sequence comprised of a TATA box and other sequences that serve to specify the site of transcription initiation, to which regulatory elements are added for control of expression. Further, a "promoter" also refers to a nucleotide sequence that includes a minimal promoter plus regulatory elements
30 that is capable of controlling the expression of a coding sequence or functional RNA. This type of promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a

promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either upstream or downstream from the promoter. Both enhancers and other upstream promoter elements bind sequence-specific DNA-binding proteins that mediate their effects. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even be comprised of synthetic DNA segments. A promoter may also contain DNA sequences that are involved in the binding of protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions.

The "initiation site" is the position surrounding the first nucleotide that is part of the transcribed sequence, which is also defined as position +1. With respect to this site all other sequences of the gene and its controlling regions are numbered. Downstream sequences (*i.e.*, further protein encoding sequences in the 3' direction) are denominated positive, while upstream sequences (mostly of the controlling regions in the 5' direction) are denominated negative.

Promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation are referred to as "minimal or core promoters". In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription. A “minimal or core promoter” thus consists only of all basal elements needed for transcription initiation, *e.g.* a TATA box and/or an initiator.

According to the invention, the term “promoter” is a regulatory region of nucleic acid (e.g. DNA) driving the transcription of a gene. Promoters are typically located near the genes they regulate, on the same strand and upstream (towards the 5' region of the sense strand). Several types of promoters are well known in the transformation arts, as are other regulatory elements that can be used alone or in combination with other promoters. A "plant promoter" is a promoter capable of initiating transcription in plant cells. Examples of promoters under developmental control include "Tissue-specific promoter" relating to regulated promoters that are not expressed in all plant cells but only in one or more cell types in specific organs (such as leaves or seeds), specific tissues (such as embryo or cotyledon), or specific cell types (such as leaf parenchyma or seed storage cells). These also include promoters that are temporally regulated, such as in early or late embryogenesis, during fruit ripening in developing seeds or fruit, in fully differentiated leaf, or at the onset of senescence.

A "cell type"-specific promoter or also called “inducible promoter” primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves by an external stimulus, such as a chemical, light, hormone, stress, or a pathogen.

5 "Constitutive expression" refers to expression using a constitutive or regulated promoter. "Conditional" and "regulated expression" refer to expression controlled by a regulated promoter.

"Constitutive promoter" refers to a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all 10 developmental stages of the plant. Each of the transcription-activating elements do not exhibit an absolute tissue-specificity, but mediate transcriptional activation in most plant parts at a level of ≥1% of the level reached in the part of the plant in which transcription is most active.

"Regulated promoter" refers to promoters that direct gene expression not constitutively, 15 but in a temporally- and/or spatially-regulated manner, and includes both tissue-specific and inducible promoters. It includes natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. New promoters of 20 various types useful in plant cells are constantly being discovered, numerous examples may be found in the compilation by Okamuro *et al.* (1989). Typical regulated promoters useful in plants include but are not limited to safener-inducible promoters, promoters derived from the tetracycline-inducible system, promoters derived from salicylate-inducible systems, promoters derived from alcohol-inducible systems, promoters derived from 25 glucocorticoid-inducible system, promoters derived from pathogen-inducible systems, and promoters derived from ecdysome-inducible systems.

"Operably-linked" refers to the association of nucleic acid sequences on single nucleic acid fragment so that the function of one is affected by the other. For example, a regulatory DNA sequence is said to be "operably linked to" or "associated with" a DNA sequence that 30 codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (*i.e.*, that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

In the above context, the term “terminator” relates to a regulatory region of a DNA sequence that marks the end of gene on genomic DNA for transcription and, thus, initiates the termination of transcription of the respective gene.

"Expression", as used herein, refers to the transcription and/or translation of an endogenous gene, ORF or portion thereof, or a transgene in plants. For example, in the case of antisense constructs, expression may refer to the transcription of the antisense DNA only. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein.

"Overexpression" relates to the level of expression in transgenic cells or organisms that exceeds levels of expression in normal or untransformed (non transgenic) cells or organisms.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of protein from an endogenous gene or a transgene.

"Gene silencing" refers to homology-dependent suppression of viral genes, transgenes, or endogenous nuclear genes. Gene silencing may be transcriptional, when the suppression is due to decreased transcription of the affected genes, or post-transcriptional, when the suppression is due to increased turnover (degradation) of RNA species homologous to the affected genes (English *et al.*, 1996). Gene silencing includes virus-induced gene silencing (Ruiz *et al.*, 1998).

As used herein, the term “BAC(s)” stands for bacterial artificial chromosome and defines a DNA construct, based on a functional fertility plasmid used for transforming and cloning in bacteria, such as *E. coli.* BACs may be used to sequence the genome of organisms such as plants. A short piece of the organism's DNA is amplified as an insert in BACs, and then sequenced. Finally, the sequenced parts are rearranged *in silico*, resulting in the genomic sequence of the organism.

As used herein, “increased seed vigour” refers to the ability of a seed to germinate rapidly, in a uniform manner, and to achieve a high percentage of germination; to produce seedlings that emerge rapidly from the soil and have a high percentage of emergence; to produce seedlings that grow rapidly and demonstrate superior tolerance to various stresses including but not limited to cold. Seed vigour is the quantification of any of the above-mentioned parameters and any combination of these.

“Seed germination” is defined as emergence of the radicle from the seed coat.

"Speed of germination" hereinafter refers to the average time observed between seed imbibition and emergence of the radicle from the seed coat.

Along the application, speed of germination can eventually be measured by calculating the time period which is needed to observe 50% of seed germination (results expressed as a
5 T50 measurement).

"Increased speed of germination" is to be understood as a significant observable difference between the germination of the seed comprising a polynucleotide according to embodiment 5 versus a seed that does not comprise one of the said polynucleotides. Typically, an increased speed of germination means that the seed comprising a
10 polynucleotide according to embodiment 5 germinates earlier than a seed that does not comprise SEQ ID NO 3, SEQ ID NO: 4 or SEQ ID NO: 6.

"Modified speed of germination" is to be understood as a significant observable difference between the germination of the seed comprising a polynucleotide according to embodiments 1 to 7 versus a seed that does not comprise one of the said polynucleotides.
15 Typically, a modified speed of germination means that the seed comprising a polynucleotide according to embodiments 1 to 7 germinates differently than a seed that does not comprise SEQ ID NO 1, 2, 3, 4, 5 or 6.

As used herein, uniformity of seed germination can be expressed as the time between 10% germination and 90% germination. The shorter this period of time is, the better the
20 uniformity of seed germination is.

"Adverse external environmental conditions" are conditions which inhibit or postpone the germination of the seed or the emergence of the seedling. In the context of the present invention, cold is one factor, among others, that can be considered as being adverse to normal germination conditions.

25 "Emergence (of a seedling)" is meant to refer to growth of the plant that is observable.

"Increased seedling emergence" is to be understood as a significant observable difference between the emergence of a seedling from a seed comprising a polynucleotide according to embodiment 5 versus a seed that does not comprise SEQ ID NO 3, 4 or 6.

The invention will be further apparent from the following non-limiting examples in conjunction with the associated sequence listings as described below:

SEQ ID NO 1: Polynucleotide sequence corresponding to the A12DHd allele of *BolC.VG1.a* gene (*Brassica oleracea* ortholog of *At3g01060* gene).

5 SEQ ID NO 2: Polynucleotide sequence corresponding to the A12DHd allele of *BolC.VG2.a* gene (*Brassica oleracea* ortholog of *At3g01150* gene).

SEQ ID NO 3: Polynucleotide sequence corresponding to the GD33DHd allele of *BolC.VG1.a* gene (*Brassica oleracea* ortholog of *At3g01060* gene).

SEQ ID NO 4: Polynucleotide sequence corresponding to the GD33DHd allele of
10 *BolC.VG2.a* gene (*Brassica oleracea* ortholog of *At3g01150* gene).

SEQ ID NO 5: Polynucleotide sequence corresponding to the truncated A12DHd allele of *BolC.VG2.b* gene (*Brassica oleracea* ortholog of *At3g01150* gene).

SEQ ID NO 6: Polynucleotide sequence corresponding to the truncated GD33DHd allele of *BolC.VG2.b* gene (*Brassica oleracea* ortholog of *At3g01150* gene).

15

Brief description of the figures

20 Fig. 1 is a graphic representation of the cumulative germination curves of seeds from the substitution line SL101 (○) and parent lines (A12DHd (■) and GDDH33 (●)) at 15°C on water. Vertical lines are standard errors.

Fig. 2 is a graphic representation of the cumulative germination curves of SL101 (◇) and the recurrent A12DHd (■) parent and reciprocal F_1 backcross lines (A12DHd x SL101 (△)
25 and SL101 x A12DHd (○).Vertical lines are standard errors.

Fig. 3 is a graphic representation of the cumulative seedling emergence in the field from seeds of the substitution line SL101 (▲) and parental lines A12DHd (●) and GDD33H (■).

Fig. 4 is a diagram illustrating endogenous concentrations of ABA during germination in seeds of the substitution line SL101 (white column) and parent lines, GDDH33 (black
5 column) and A12DHd (grey column).

Fig. 5 is a diagram illustrating the speed of seed germination in 3, for gene *At3g01060* (black columns), and 2, for gene *At3g01150* (grey columns), KO lines. The wild type control line (white column) is col0.

Fig. 6 is a diagram illustrating the speed of seed germination in KO lines 102 (light grey
10 line) and 18 (low medium grey line) (for gene *At3g01060*) and in KO lines 15 (high medium grey line) and 3 (dark grey line) (for gene *Atg01150*) compared to wild type control line Col0 (black line), following seed production in the glasshouse (more stressful conditions).

Example 1

15

Material and methods

Seed production and comparison of lines

Seed samples were obtained from Birmingham University, UK, for a range of *Brassica*
20 *oleracea* chromosome substitution lines derived from the doubled haploid parent lines A12DHd (var. *alboglabra*) and GDDH33 (var. *italica*; Rae *et al.*, 1999). Bulk seeds were then produced and collected from 10 individual replicate plants of the substitution lines and the GDDH33 parent and 20 plants of the recurrent A12DHd parent as substitution lines are compared to the latter in germination experiments. Plants were laid out in a randomized
25 block with 10 replicates in a glasshouse at 16-18°C during the 16 h day and at 10-15°C at night as described by Bettey *et al.* (2000). Supplementary lighting (400 W high pressure sodium lamps; Osram Ltd, St Helens, UK) was supplied when light intensity fell below 300 w m^2 during the 16 h day. Plants were self-pollinated by enclosing the inflorescences in perforated polyethylene bags containing blowflies before the flowers opened. The
30 seedpods were allowed to dry completely on the plant within the enclosing bags before

harvest. The seeds were cleaned, equilibrated at 15 % rh and 15°C, and then stored at -20°C before germination experiments were carried out. Cumulative germination on moist filter paper was recorded at 15°C on 4 replicates of 15 seeds collected from each of the 10 replicate plants (or 20, A12DHd) described above. Previous work had shown this to be
5 sufficient seeds (Bettey *et al.*, 2000). Frequent counts were made to allow an accurate calculation of the time to 50% germination from these measurements. Percentage germination was high in all seed lots.

In later experiments, F1 seeds from reciprocal backcrosses were produced in the same manner as described above. Bud pollination was performed to make the cross resulting in
10 the F1. Seeds from the parent lines were produced at the same time for comparison to minimize the influence of environmental differences during seed production. In addition, on a number of occasions at different times of the year seeds were produced from replicate plants of both the parent A12DHd and substitution line SL101 in glasshouses as described above. Although glasshouse heating, venting and lighting settings remained the same as
15 those described above ambient temperature differed and was recorded.

Germination assays

Three biological replicates of 50 seeds from substitution and parent *Brassica oleracea* lines or 3 to 15 biological replicates of 50 seeds from *Arabidopsis* wild type and mutant
20 lines were placed to germinate on 2 layers of filter paper (Whatman International Ltd., UK) kept moist throughout treatment with water.

In all germination experiments, seeds on filter paper were held in clear polystyrene boxes laid out in randomized blocks and kept in the dark. No evidence of fungal infection was observed and so seeds were not sterilized to avoid influencing their germination.
25 Germination (radicle emergence) was recorded at intervals to construct cumulative germination curves.

Field emergence

As part of a larger unpublished comparison of seedling emergence from *B. oleracea*
30 genotypes, 100 seeds of the parent lines GDDH33, A12DHd and substitution line SL101 were sown on 31 May, in 4 replicate 1m rows arranged in a randomized block. Seeds were sown by hand in a 15mm deep furrow, covered with sieved soil (sieve hole size <4mm) and the surface rolled once with a Stanhay seed drill press wheel. The soil was a sandy loam and irrigation was applied to maintain soil moisture throughout seed germination and

seedling emergence. The latter was recorded at regular intervals until no more seedlings emerged.

Hormone analysis

5 Samples of seeds from substitution and parent lines were taken as non-imbibed dry seeds and seeds imbibed for 24 and 48 hours at 15°C on moist filter paper. Each sample contained 1 g of dry seed measured before imbibition. The samples were placed immediately in liquid nitrogen, freeze dried and then placed in a domestic freezer at -20°C until extraction. The samples were in 10 ml of cold (4°C) 80% methanol (containing 20 mg 10 l-l of BHT), then 500 ng ABA and 100 ng each of GA standards were added. The samples were stirred overnight in a cold cabinet and then centrifuged. The supernatant was decanted and the pellet re-suspended in a further 10 ml of 80 % methanol. The samples were stirred for 4 h, centrifuged and the supernatants combined. The supernatants were evaporated to aqueous (*c.* 3-4 ml) and 10 ml of 0.5 M pH 8.2 phosphate buffer was added, 15 samples were partitioned with 2 x 15 ml of dichloromethane, and the dichloromethane discarded. The aqueous fraction was adjusted to pH 3 with 1 M phosphoric acid partitioned with 3 x 15 ml of ethylacetate.

The combined ethylacetate fractions were washed with 2 x 3.5 ml pH 3.0 water and evaporated to dryness, re-dissolved in 5 ml water and the pH adjusted to 8. The solution 20 was then loaded onto QMA cartridges which were subsequently washed with 5 ml of 15 % methanol pH 8.0. GAs and ABA were eluted from the QMA cartridges directly onto C18 21 cartridges with 0.2 M formic acid in 5 % methanol. C18 cartridges were then washed with 5 ml of 20 % methanol and samples recovered in 5 ml of 80 % methanol and evaporated to dryness. Samples were then dissolved in methanol and methylated with excess ethereal 25 diazomethane. Following evaporation to dryness samples were re-dissolved in dry ethylacetate and passed through amino cartridges. The resulting samples were analyzed directly for ABA content by GC-MS, then evaporated to dryness and re-dissolved in BSTFA prior to analysis for GA content by GC-MS.

30 *Marker analysis*

Primer pairs were designed to 30 *Arabidopsis* gene models that were spread at intervals across the *SOG1* region using Primer 3 software (http://gene.pbi.nrc.ca/cgibin/primer/primer3_www.cgi) and gene data from Tair (http://www.arabidopsis.org/) to give PCR products from 200 to 700 bp. The PCR mix used

was standard but a touch-down program was used. This consisted of cycling parameters as follows: 94°C for 5mins; then annealing at 65°C to 55°C for 10 cycles dropping a degree each cycle with 30 s extension at 72°C and 30 s denaturation at 94°C over the 10 cycles; followed by 30 cycles of 94°C 30 s, 55°C 30 s, 72°C 45 s; and a final extension at 72°C for 5 15 min. Primer sequences for the gene models that gave polymorphic results were selected as markers (Table.1).

Data analysis

All analyses were performed using the statistical package Genstat 5 (Payne *et al.* 1993), 10 and where appropriate data were subjected to analyses of variance.

Results

QTL for speed of germination (SOG1) on linkage group C1 confirmed and fine-mapped

15 Analysis of variance of germination data comparing substitution and parent lines showed that the GDDH33 parent germinated significantly ($P<0.001$) faster than the A12DHd parent confirming that the positive speed of germination alleles are provided by GDDH33, as shown by Bettey *et al.* (2000) (fig. 1). There were 4 substitution lines which spanned the *SOG1* QTL (SL101, SL111, SL118, SL119) and all of these had significantly ($P<0.005$) 20 faster germination than the A12DHd parent. The substitution line SL101 had the smallest introgressed region (1-9 cM; Rae *et al.*, 1999) that enhanced speed of germination compared to A12DHd and accounted for much of the difference in speed of germination between the parent lines (fig. 1) and was therefore selected for further study of *SOG1.*

25

The SOG1 fast germinating phenotype is not influenced by the maternal genotype

Speed of germination is determined by the embryo, but can also be significantly influenced by the tissues that surround it which are maternal in origin. Reciprocal backcrosses between SL101 and A12DHd and between GDDH33 and A12DHd were carried out to 30 determine the maternal and zygotic genetic components at the *SOG1* locus. Germination was recorded from the F1 seeds of each cross and from seeds of the selfed parent lines produced at the same time. There was no significant difference in the speed of germination of SL101 and the F1 from the reciprocal backcrosses with A12DHd (SL101 as

mother plant and pollen from A12DHd and *vice versa*), but germination of seeds from all three were significantly (P<0.01) faster than that from seeds of A12DHd (fig. 2). This shows the faster germinating GDDH33 allele to be dominant with no genetic maternal influence on inheritance of the trait confirming that it is embryo based.

5

Differences in speed of germination lead to differences in the timing of seedling emergence in the field

The data above show that the speed of germination of GDDH33 and SL101 seeds was significantly greater than that of A12DHd under constant temperature conditions. In the

10 field this resulted in significantly earlier seedling emergence from GDDH33 and SL101 than from A12DHd (fig. 3).

Endogenous ABA concentration differs between genotypes at maturity

The endogenous concentration of ABA in dry and imbibing seeds of the three genotypes

15 was measured using GCMS. There was no significant difference in the endogenous concentration of ABA in the dry seed of SL101 and GD33 or during imbibition to 48 hours just prior to radicle emergence. ABA concentration in these two seed lots remained the same over the first 24 hours and then declined to half that by 48 hours. In contrast, the endogenous concentration of ABA in seeds of A12DHd was initially three fold higher than

20 that in SL101 and GDDH33 and then declined progressively over the 48 hour period of imbibition, but remained significantly above that of the other two seed lots (fig. 4). Interestingly, if ABA continues to decline at the same speed in A12DHd it would reach the same level after 72 hours, the point of germination, as seen in the other two lines immediately before their germination.

25 The results for GDDH33, SL101 and the A12DHD lines presented here show that a clear genetically determined relationship between higher endogenous ABA content, and lower speed of germination and *vice versa.*

A quantitative genetic analysis of the speed of germination trait in Brassica oleracea has

30 *been carried out.*

By fine mapping the QTL with the previously described substitution line SL101, we finally identified two genes underlying the QTL (*Speed Of Germination* (*SOG1*)). This line SL101 has a short introgression at the telomeric end of C1 from the fast germinating parent (GD33, Broccoli) in the background of the slow germinating parent (A12, Kale) in *Brassica*

oleracea. By the way of using markers across this introgressed region of SL101, 30 recombinations along this area were identified within 1,300 lines, thanks to a BAC tilling path strategy. This strategy allows gathering the lines into 5 distinct groups and the 2 parent lines. Seed germination was then evaluated throughout these 7 groups and statistical analysis revealed that faster germination was associated with 2 markers at the telomeric end of C1. Co-localization of these markers with a single BAC on the telomeric end of C1 was further assessed by Fluorescence In-Situ Hybridization. The BAC was sequenced and 12 full-length genes were found to be present within.

Putative orthologs of these *Brassica* genes have been identified in *Arabidopsis*, at the top arm of chromosome 3, where a *SOG1* QTL has been located (Clerkx *et al.* 2004). Based on the good genetic colinearity between the telomeric end of C1 in *Brassica oleracea* and the top arm of chromosome 3 in *Arabidopsis* (see above and below), it is reasonable to think that the genes that were identified in *Brassica* and in *Arabidopsis* do share common function in seed germination. In order to strengthen our hypothesis, we have identified *Arabidopsis* knockout (KO) mutant lines in putative ortholog genes of those that have been discovered in *Brassica*. Two of these KO lines were found to have a significant germination phenotype (faster germination compared to control Col0 line) suggesting that these genes act as negative regulators of germination (fig. 5). It is interesting to note that at least two different KO lines were used to assess germination phenotype and that these distinct KO lines showed similar results regarding to the speed of germination. When assessed following seed production in more stressful conditions (fig. 6), the results are very comparable.

This functional study has thus confirmed the role of *Atg01060* and *Atg01150* genes in the regulation of seed vigour, in particular in the regulation of the speed of germination. The *Brassica* orthologs of these genes, *BolC.VG1.a* and *BolC.VG1.b*, which have been identified within the *SOG1* QTL in *Brassica oleracea*, can therefore indeed be considered as tools allowing the modulation of the seed vigour, more particularly the modulation of the speed of seed germination in brassicaceae.

The linkage between B. oleracea linkage group C1 and the top arm of chromosome three of Arabidopsis is confirmed

Studies on SL101 above have shown that a single introgressed region at the telomeric end of linkage group C1 (LGC1) contains the QTL for *SOG1*. In the current work we aimed to

establish colinearity between this region in *B. oleracea* and the *Arabidopsis* genome to enable comparison with the extensive QTL analysis carried out on this model species. Previously a number of linkages have been shown between the *Brassica* genome and *Arabidopsis* (Cogan *et al.*, 2004; Parkin *et al.*, 2005).

5 In particular, the linkage between the LGC1 and *Arabidopsis* (Cogan *et al.*, 2004) was utilized to assist in the development of further informative markers. Using this approach, primer pairs were designed to 30 *Arabidopsis* gene models, which were spread across this region. These primers were tested to determine if they amplified a *B. oleracea* product and then if there were any polymorphisms between SL101 and the parental lines A12DHd and 10 GD33DHd (table 1). A banding pattern that is the same in SL101 and GD33DHd, but different from that in A12DHd indicatess its presence at this locus, and therefore its usefulness as a marker for *SOG1*. Primers for three gene models were identified as informative markers (*At3g01190*, *At3g07130*, *At3g02420, table 1*) that anchored the linkage between the top arm of Arabidopsis chromosome three and the *SOG1* region of *B.* 15 *oleracea* LGC1. This confirmation of colinearity justifies comparison of *SOG1* with QTL for seed performance located to this region of the *Arabidopsis* genome.

Gene model	Primers
At3g01190	F: TTCTTCCACGACTGCTTCG
	R: CTAACAAAACTGATCCGTCAC
At3g02420	F: GTTGCGTTGCCATCTGCAG
	R: CAGGCTGAGATAGCCATTGG
At3g07130	F: CTACTAACCATGGAGTTACC
	R: AACGCTGGTGGGATTCAC

Table 1. Primers for selected markers used to anchor SL101 introgression to *Arabidopsis*

20

Altogether, these results open up the possibility of using either *BolC.VG1.a* A12/GD33 alleles or *BolC.VG2.a* A12/GD33 alleles or *BolC.VG2.b* A12/GD33 alleles according to any of embodiments 1 to 23, in particular to engineer plants in which the seed vigour has been modulated, more particularly plants in which speed of germination has been modulated.

25

Example 2. Experiments highlighting additional phenotypes linked to seed vigour.

These experiments underlines the differences in germination characteristics between seeds with (SL101) or without (A12) the GD33 allele.

5

Seeds of the lines A12 and SL101 originated in the UK and were replicated in 2009 in South Africa by a commercial seed producer. Determination of the germination characteristics of the seeds took place in early 2010 in Enkhuizen, Netherlands.

10 *Seed performance under standard commercial conditions*

Two replicates of 100 seeds were sown in standard trays filled with soil as in a normal commercial practice. Trays were placed in a germination chamber at 18°C in the dark for three days. Trays were then transferred to a greenhouse with an average temperature of 20°C. At 10 days after sowing the number of normally developed seedlings and the 15 number of non-emerged seeds were counted.

Performance of SL101 under these conditions was considerably better than for A12 (table 2).

	normal seedlings	non germinated seeds
A12	19	80
SL101	91	7

20

Table 2. Percentage normal seedlings and percentage non germinated seeds as measured under practical conditions of seedling production

Temperature sensitivity of germination on paper

25 Two replicates of 100 seeds of each A12 and SL101 were sown for each combination of conditions on wet filter paper in transparent plastic germination boxes at temperatures of 10, 20 or 30°C. Boxes were either kept in the dark at the mentioned temperatures or were placed under fluorescent light.

Germination, measured as radical protrusion, was counted daily. Germination was counted 30 until 10 days after the start of the measurement when no additional germination was observed.

The data show that under each condition the percentage germination was higher for SL101 compared to A12 (table 3).

		temperature		
		10	20	30
light	A12	47	94	59
	SL101	94	100	99
dark	A12	65	24	2
	SL101	95	78	50

5

Table 3. Final germination percentage of A12 and SL101 at different temperatures in the light or in the dark

Speed of germination on paper

10 Speed of germination was determined under conditions of 20°C in the light, a condition where both lines had their maximum germination. Two replicates of 50 seeds per line were placed on paper in plastic germination boxes under the mentioned conditions. The t50 (time until 50% of the germinating seeds have germinated) was determined (table 4).
As shown in the table the time until 50% germination was considerably shorter for SL101
15 compared to A12.

line	t50 (h)
A12	59.0
SL101	47.5

Table 4. Time until 50% germination at 20°C in the light for A12 and SL101

20

Sensitivity to low temperatures during germination on paper

Two replicates of 50 seeds per treatment were incubated at 5°C or 10°C in the light as described before for a germination test on paper. Incubation was in water or in a 1mM solution of GA3. After 10 days seeds that had germinated were removed. The boxes with
25 the remaining non-germinated seeds were then placed at 20°C in the light. After a further 5 days the number of germinated seeds was counted. As a comparison seeds of both A12

and SL101 were germinated at 20°C in the light without a pretreatment. Germination percentage of these seeds was counted after 10 days.
Especially at 5°C there was a large difference between the two lines (table 5). The pretreatment in water largely prevented for A12 germination at 20°C in the light later on.
5 This was not the case for pretreatment in 1mM GA3. SL101 showed only a minor reduction in germination after the pretreatment at 5°C in water.

	A12	SL101
no pretreatment	99	100
10d 5C in water	15	84
10d 5C in 1mM GA3	100	100
10d 10C in water	82	99
10d 10C in 1mM GA3	100	100

10 Table 5. Percentage of germination of seeds of A12 and SL101 with or without a pretreatment at 5 or 10°C in the light with an incubation medium of either water or a 1mM aqueous solution of GA3

15 Example 3 Effect of genotype on hybrid seed performance

Hybrid seeds were produced under commercial seed production conditions in South Africa by pollinating flowers of two male sterile lines with pollen from 7 other lines.
Female line 1 contained the GD33 allele, female line 2 contained the A12 allele.
20 The sensitivity to low temperature during germination was tested for these seeds. Seeds were incubated on moistened filter paper at 5°C in white fluorescent light for 10 days. Percent germination was recorded after 10 days and non germinated seeds were transferred to 20°C in light. After five days the percentage germinated seeds was recorded.
25 Seeds produced on the female line containing the GD33 allele showed considerably higher germination at 5°C and germinated more than 90% after transfer to 20°C. Seeds produced on the female without the GD33 allele showed only low germination percentages and did not recover after transfer to 20°C.

	female 1 (GD33)			female 2 (A12)	
	10d 5C	+5d 20C		10d 5C	+5d 20C
male1	35	96		0	6
male2	28	96		0	10
male3	72	94		19	44
male4	9	96		0	4
male5	43	100		0	8
male6	28	99		0	4
male7	69	97		35	58
average	41	97		8	19

Table 6. Percentage of germination of hybrid seeds produced using a GD33 or A12 female line with a pretreatment for 10 days at 5°C in the light followed by incubation of non
5 germinated seeds for 5 days at 20°C.

Seeds from the same batches of hybrid seed were tested in a germination test on paper at 25°C in white fluorescent light. From daily counts of germination the time until 50% germination was calculated (t50). The final germination percentage was determined after 5
10 days. All hybrids had at least 85% germination, the majority more than 95%. Seeds produced on the female line with the GD33 allele showed on average around 25% faster germination, shown by the lower t50 of 1.33 days versus 1.80 days for seeds produced on the female with the A12 allele.

SUBSTITUTE SHEET (RULE 26)

	female 1 (GD33)		female 2 (A12)	
	germ%	t50 (d)	germ%	t50 (d)
male 1	100	1.08	99	1.73
male 2	97	1.95	98	1.85
male 3	98	1.85	100	1.90
male 4	100	1.15	95	1.84
male 5	99	1.07	98	1.68
male 6	99	1.14	85	1.84
male 7	95	1.05	98	1.79
average	98	1.33	96	1.80

Table 7. Percentage germination and time taken until 50% germination (t50) of hybrid seed produced using a GD33 or A12 female line in the light at 25°C.

5

References

Bettey M., Finch-Savage W.E., King G.J., Lynn J.R. 2000. Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in *Brassica oleracea* L. New Phytol
10 148: 277-286.

Clercks E.J.M., El-Lithy M.E., Vierling E., Ruys G.J., Blankestijn-De Vries H., Groot S.P.C., Vreugdenhil D., Koornneef M. 2004. Analysis of natural allelic variation of *Arabidopsis* seed germination and seed longevity traits between the accessions Landsberg errecta and Shakdara, using a new recombinant inbred line population. Plant Physiol 135: 432-443.

15 Finch-Savage W.E. 1995. Influence of seed quality on crop establishment, growth and yield. In AS Basra, ed. Seed quality: Basic mechanisms and agricultural implications. Haworth Press, Inc, New York, pp 361-384.

Finch-Savage W.E., Côme D., Ly J.R., Corbineau F. 2005. Sensitivity of *Brassica oleracea* seed germination to hypoxia: a QTL analysis. Plant Sci 169: 753-759.

20 Finch-Savage W.E, and Leubner-Metzger G. 2006. Seed dormancy and the control of germination. New Phytol. 171: 501-523.

Finkelstein R.R., Gampala S.S., and Rock C.D. 2002. Abscisic acid signaling in seeds and seedlings. Plant Cell 14, S15-45.

Foolad M.R., Lin G.Y., Chen F.Q. 1999. Comparison of QTLs for seed germination under non-stress, cold stress and salt stress in tomato. Plant Breed 118: 167-173.

5 Groot S.P.C., van der Geest A.H.M., Tesnier K.J.Y., Alonso-Blanco C., Bentsink L., Donkers H., Koornneef M., Vreugdenhil D., Bino R.J. 2000. Molecular genetic analysis of *Arabidopsis* seed quality. In M Black, J Vasquez-Ramos, eds, Seed Biology: Advances and applications, CAB International., London, pp 123-132.

Hilhorst H.W.M. and Toorop P.E. 1997. Review on dormancy, germinability and 10 germination in crop and weed seeds. Advances Agron 61: 115-165.

Holdsworth M.J., Bentsink L., Soppe W.J.J. 2008a. Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. New Phytol 179: 33-54.

Holdsworth M.J., Finch-Savage W.E., Grappin P., Job J. 2008b. Post-genomics dissection 15 of seed dormancy and germination. Trends Plant Sci 13: 7-13.

Ruan S.L., Duan X.M. and Hu W.M. 2000. Occurrence of seed vivipary in hybrid rape (*Brassica napus* L.) and its effect on seed quality. Journal of Zhejiang University (Agriculture and Life Sciences) 2000 Vol. 26 No. 5 pp. 573-578.

Sequence alignments

DNA sequence alignment for the Brassica oleracea VG2 genes. Alignments were performed using the ClustalW web based program (http://www.genome.jp/tools/clustalw/) and the annotation is drawn using the web base program BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html). A12_BOLC.VG2.A is an A12 full length copy of the gene located in the SOG1 region of linkage group C1. A12_BOLC.VG2.B is a truncated copy of a similar gene located within 50Kb of the full length gene. A similar annotation has been used to describe the same region in the GD33 genomic background.

```
A12_BOLC.VG2.A     1 GTAGGCCAAGCCAAGCATATGGATCACTTAGTTCAGAGAAGTTATGGCAAGAAATAAGCC
GD33_BOLC.VG2.A    1 GTAGGCCAAGCCAAGCATATGGATCACTTAGTTCAGAGAAGTTATGGCAAGAAATAAGCC
A12_BOLC.VG2.B     1 -ACGTCTATGACCGGACTA--GTCTCCTTCGTGAAGACCGGTCACTGAAACGCCAATGAT
GD33_BOLC.VG2.B    1 -ACGTCTATGACCGGACTA--GTCTCCTTCGTGAAGACCGGTCACTGAAACGCCAATGAT

A12_BOLC.VG2.A    61 ATGCAGATTAGAAAAAGAGTGCAACAAAC-TAAGTACAGTACTAAAAGACTAAGTTAATA
GD33_BOLC.VG2.A   61 ATGCAGATTAGAAAAAGAGTGCAACAAAC-TAAGTACAGTACTAAAAGACTAAGTTAATA
A12_BOLC.VG2.B    58 GTTCGCCATAATCTCCCATTTCTTTGAAGGTATTTGAACGCTTTGCTGCGTTGATTGA-G
GD33_BOLC.VG2.B   58 GTTCGCCATAATCTCCCATTTCTTTGAAGGTATTTGAACGCTTTGCTGCGTTGATTGA-G

A12_BOLC.VG2.A   120 ATACAATTAAACCAAAATTAATGGATGAGTACCTTCACAAAACATTAGTGA-ATTTTCAG
GD33_BOLC.VG2.A  120 ATACAATTAAACCAAAATTAATGGATGAGTACCTTCACAAAACATTAGTGA-ATTTTCAG
A12_BOLC.VG2.B   117 GCGGATCTTAAGCAGCTCGAGTGGATCTA-ATCAGCTGAGAGAATCGGGAGACGATTAG
GD33_BOLC.VG2.B  117 GCGGATCTTAAGCAGCTCGAGTGGATCTA-ATCAGCTGAGAGAATCGGGAGACGATTAG

A12_BOLC.VG2.A   179 ACCCTCTGGACCAATACAAGACTAACGATGTCACACTTAACTGACAGTTAGTTCATGTAG
GD33_BOLC.VG2.A  179 ACCCTCTGGACCAATACAAGACTAACGATGTCACACTTAACTGACAGTTAGTTCATGTAG
A12_BOLC.VG2.B   176 ATCATGCGG-----TCACAGATTCGAGA-GGAAACCCTAATCGCCA-TCGCCGCCAGACA
GD33_BOLC.VG2.B  176 ATCATGCGG-----TCACAGATTCGAGA-GGAAACCCTAATCGCCA-TCGCCGCCAGACA

A12_BOLC.VG2.A   239 GGCCAATAAACAGCTCAAATTCGTTAGCTTTCAAGTTCCAATTTTCTGCGTTGTGTTTTG
GD33_BOLC.VG2.A  239 GGCCAATAAACAGCTCAAATTCGTTAGCTTTCAAGTTCCAATTTTCTGCGTTGTGTTTTG
A12_BOLC.VG2.B   229 GTCTCATCTCCAGTGTCA-----TTCATATGCAAATCTCCTGGATTTG-GTTGTTTTTTT
GD33_BOLC.VG2.B  229 GTCTCATCTCCAGTGTCA-----TTCATATGCAAATCTCCTGGATTTG-GTTGTTTTTTT

A12_BOLC.VG2.A   299 CTTTTTGAGGGTTGATTTGGAATTTCTCCAAAGAGGTCACATAATTTAAAAGAAATTGAT
GD33_BOLC.VG2.A  299 CTTTTTGAGGGTTGATTTGGAATTTCTCCAAAGAGGTCACATAATTTAAAAGAAATTGAT
A12_BOLC.VG2.B   283 CTTTTTCAGGTTTGATTTGGAATTTCTCTAAAGAGGTCACATTAATTAAAATAAACTAGA
GD33_BOLC.VG2.B  283 CTTTTTCAGGTTTGATTTGGAATTTCTCTAAAGAGGTCACATTAATTAAAATAAACTAGA

A12_BOLC.VG2.A   359 CAAAATTATATTGGACAAG--GCAAGCAAATCCATTAAGAAGATCACAACAAAAGAAAAA
GD33_BOLC.VG2.A  359 CAAAATTATATTGGACAAG--GCAAGCAAATCCATTAAGAAGATCACAACAAAAGAAAAA
A12_BOLC.VG2.B   343 CCCTTATCCGCGCGCCAGC--GCAGATATGAATTTTTAGTTTTTAATTATTTATTTTATT
GD33_BOLC.VG2.B  343 CCCTTATCCGCGCGCCAGC--GCAGATATGAATTTTTAGTTTTTAATTATTTATTTTATT

A12_BOLC.VG2.A   417 TTAAAAACCAACAAAAGAAATTTCCCATTTCACGAACGTGATGAACAAGATCACTACGTA
GD33_BOLC.VG2.A  417 TAAAAAACCAACAAAAGAAATTTCCCATTTCACGAACGTGATGAACAAGATCACTACGTA
A12_BOLC.VG2.B   401 CAATCATGTATTTCTAATATTCGTTCATATTATATTCGTTAAGTA-AATATTTTTTTGTA
```

```
   GD33_BOLC.VG2.B    401  CAATGATGTATTTGTAATATTCGTTCATATTATATTCGTTAAGTA-AATATTTTTTTGTA

   A12_BOLC.VG2.A     477  TACTCCACCAATGCATGACACTACTGAAAACATACA--CAAACAAGCTCGTGATAGTTGT
   GD33_BOLC.VG2.A    477  TACTCCACCAATGCATGACACTACTGAAAACATACA--CAAACAAGCTCGTGATAGTTGT
 5 A12_BOLC.VG2.B     460  T-CTTAAACTATCTATTTTTTTAC-GAATGTGTGATATCATATAAAAATATTA-AAAAAT
   GD33_BOLC.VG2.B    460  T-CTTAAACTATCTATTTTTTTAC-GAATGTGTGATATCATATAAAAATATTA-AAAAAT

   A12_BOLC.VG2.A     535  GACCACCTCACACATGAAAAAACTCCATGAATGTCAAAAGTTTGCACGATTATGTAAACA
   GD33_BOLC.VG2.A    535  GACCACCTCACACATGAAAAAACTCCATGAATGTCAAAAGTTTGCACGATTATGTAAACA
10 A12_BOLC.VG2.B     517  GAGTATACACTTAATAACATAATTTTAAAAACACAAATTTTTCTTTCGTGTGCGATATCA
   GD33_BOLC.VG2.B    517  GAGTATACACTTAATAACATAATTTTAAAAACACAAATTTTTCTTTCGTGTGCGATATCA

   A12_BOLC.VG2.A     595  CTTAATCACTTATTCATTGAAGGATAATGTTGCACAGAATATTATCATTGTTTCTAATGG
   GD33_BOLC.VG2.A    595  CTTAATCACTTATTCATTGAAGGATAATGTTGCACAGAATATTATCATTGTTTCTAATGG
15 A12_BOLC.VG2.B     577  TATAAATAATGATCCGCCCAG----------TTAACAAAA-ATCATGATTATTTTATGTG-
   GD33_BOLC.VG2.B    577  TATAAATAATGATCCGCCCAG----------TTAACAAAA-ATCATGATTATTTTATGTG-

20 A12_BOLC.VG2.A     655  TAAGCATAAGCAAACATTGTAAAAAGAAGTCATGATCATGCTATATCTTATAGGAGTTTA
   GD33_BOLC.VG2.A    655  TAAGCATAAGCAAACATTGTAAAAAGAAGTCATGATCATGCTGTATCTTATAGGAGTTTA
   A12_BOLC.VG2.B     626  TAATTTTCTTTAT---TTTGACTATTTCCTTAAAACTATATTAAATTTTACAT-ATTTTA
   GD33_BOLC.VG2.B    626  TAATTTTCTTTAT---TTTGACTATTTCCTTAAAACTATATTAAATTTTACAT-ATTTTA

25
   A12_BOLC.VG2.A     715  ATTTATTCCAAGTTCACATTTAGTGATGCTCTAATCTTCAA--CAATAAAACATAACCAA
   GD33_BOLC.VG2.A    715  ATTTATTCCAAGTTCACATTTAGTGATGCTCTAATCTTCAA--CAATAAAACATAACCAA
   A12_BOLC.VG2.B     682  ATTAGAATATTTTTAATATCT--TTACCTTTTTATTTGAAATGCAACTCAATATTTTTTT
   GD33_BOLC.VG2.B    682  ATTAGAATATTTTTAATATCT--TTACCTTTTTATTTGAAATGCAACTCAATATTTTTTT
30
   A12_BOLC.VG2.A     773  CTGATAGTCAGAGTAACCAATTGAATGCCAGTTTTCAAAAGAAAT--AATACATATCTAA
   GD33_BOLC.VG2.A    773  CTGATAGTCAGAGTAACCAATTGAATGCCAGTTTTCAAAAGAAAT--AATACATATCTAA
   A12_BOLC.VG2.B     740  TTAACAATTATAACAAAATATTTAAAAAATATTTTTAGAATTTGTTTGAAAAATATGAAA
   GD33_BOLC.VG2.B    740  T-AACAATTATAACAAAATATTTAAAAAATATTTTTAGAATTTGTTTGAAAAATATGAAA
35
   A12_BOLC.VG2.A     831  CTGATAGTCAGTCAAATCAACCAAATAAGTAGAAAGAACCAAACAAACACAACTAGGGCT
   GD33_BOLC.VG2.A    831  CTGATAGTCAGTCAAATCAACCAAATAAGTAGAAAGAACCAAACAAACACAACTAGGG-T
   A12_BOLC.VG2.B     800  ATTACATT---TTAAATTAA---AATTATCCTAAAATAT--GATAAGTTTTGATGTAAAC
   GD33_BOLC.VG2.B    799  ATTACATT---TTAAATTAA---AATTATCCTAAAATAT--GATAAGTTTTGATGTAAAC
40
   A12_BOLC.VG2.A     891  GGACACTTCGATTATTTTGTAGGTTCGGTTTGGATTCGGTTCGGTTTGGTTAATTTATTC
   GD33_BOLC.VG2.A    890  GGACACTTCGATTATTTTGTAGGTTCGGTTTGGATTCGGTTCGGTTTGGTTAATTTATTC
   A12_BOLC.VG2.B     852  GAAAACTCAAATTATGTCAAAAAT---AATGATATTCAATG-ATAATAACAATTTTAATT
   GD33_BOLC.VG2.B    851  GAAAACTCAAATTATGTCAAAAAT---AATGATATTCAATG-ATAATAACAATTTTAATT
45
   A12_BOLC.VG2.A     951  TAGAATTTTTTAAACTGAAGTAAACCATAAGTTTCAGTCTGATTTGTATTCGGCATGGTT
   GD33_BOLC.VG2.A    950  TAGAATTTTTTTAACTGAAGTAAACCATAAGTTTCAGTCTGATTTGTATTCGGCATGGTT
   A12_BOLC.VG2.B     908  GATTATTTTTTAAA---AAATA---CATTTACAAAAATATTGTTTGAAAGAAAATTCATT
   GD33_BOLC.VG2.B    907  GATTATTTTTTAAA---AAATA---CATTTACAAAAATATTGTTTGAAAGAAAATTCATT
50
   A12_BOLC.VG2.A    1011  TGTGTTTCAGTTCAGTTTAATCGGATTCCCTTAATTTAATTCTTTTAAGAAAAATCATGT
   GD33_BOLC.VG2.A   1010  TGTGTTTCAGTTCAGTTTAATCGGATTCCCTTAATTTAATTCTTTTAAGAAAAATCATGT
   A12_BOLC.VG2.B     962  TATCTTTCAAAT--ATAAAATGAAAATATTATAATTAAATAATAAAAAATATAAATAAAA
   GD33_BOLC.VG2.B    961  TATCTTTCAAAT--ATAAAATGAAAATATTATAATTAAATAATAAAAAATATAAATAAAA
55
   A12_BOLC.VG2.A    1071  TTTAAAACACGGTCATGAAATCATCATGTTGGTGACTATGAAAAAATAAATTATATAATT
   GD33_BOLC.VG2.A   1070  TTTAAAACATGGTCATGAAATCATCATGTTGGTTACTATGAAAAAATAAATTATATAATT
   A12_BOLC.VG2.B    1020  ACCATAAATTTAGCAAATGACAACTGAGTTA--TATTATGCTAAAAAAAAATTTCTAACA
   GD33_BOLC.VG2.B   1019  ACCATAAATTTAGCAAATGACAACTGAGTTA--TATTATGCTAAAAAAAAATTTCTAACA
60
   A12_BOLC.VG2.A    1131  AAATCTAATAT-ATAACTGAAACAAAAATCCTAAAAAACCACAAAAAAATTTATAAGAAC
```

```
   GD33_BOLC.VG2.A  1130 AAATCTAATAT-ATAACTGAAACAAAAATCCTAAAAAACCACAAAAAATTTTATAAGAAC
   A12_BOLC.VG2.B   1078 AAATATCAAATGACAAATGAGTTATGAGT----AAATAATACCATATAATTTTTAAAAAC
   GD33_BOLC.VG2.B  1077 AAATATCAAATGACAAATGAGTTATGAGT----AAATAATACCATATAATTTTTAAAAAC

 5 A12_BOLC.VG2.A   1190 ACAATCAAACCAAAGT-TAACTAAAGTAAATAGAAAATTAAAAAAAAAA-CAATGTCAAT
   GD33_BOLC.VG2.A  1189 ACAATCAAACCAAAGT-TAACTAAAGTAAATAGAAAATTAAAAAAAAAAACAATGTCAAT
   A12_BOLC.VG2.B   1134 ATAGCCATTCTTAAATATTTTTTGAATAAATAATTATTTTAAATTTAATCAACTAAAAAT
   GD33_BOLC.VG2.B  1133 ATAGCCATTCTTAAATATTTTTTGAATAAATAATTATTTTAAATTTAATCAACTAAAAAT

10

   A12_BOLC.VG2.A   1248 AAAGTTAATACTGACTTTAGTGTTAGGATTTAGTATAAAAATGTTTAATTCAAATAATAA
   GD33_BOLC.VG2.A  1248 AAAGTTAATACTGACTTTAGTGTTAGGATTTAGTATAAAAATGTTTAATTCAAATAATAA
   A12_BOLC.VG2.B   1194 AATATCTGTACAATTGTGTGAGTCAAATTCTAGTTTTA------TTGATGCATGTTATAT
15 GD33_BOLC.VG2.B  1193 AATATCTGTACAATTGTGTGAGTCAAATTCTAGTTTTA------TTGATGCATGTTATAT

   A12_BOLC.VG2.A   1308 CAAAGACACATTGGTTTTTCGGTGCAATTTGATTCGGTTCGATTCGGTTCGGTCTGATTT
   GD33_BOLC.VG2.A  1308 CAAAAACACATTGGTTTTTCGGTGCAATTTG-TTCGGTTTGATTCGGTTCGGTCTGATTT
   A12_BOLC.VG2.B   1248 ATTAGTGCTGCATGTTTTAGGTAACATTTTAATGATGCACGTTTTCAAAAATCTCCATTA
20 GD33_BOLC.VG2.B  1247 ATTAGTGCTGCATGTTTTAGGTAACATTTTAATGATGCACGTTTTCAAAAATCTCCATTA

   A12_BOLC.VG2.A   1368 TTTTGCTTGTCTGAACACAACGACGGCCGACAGATAACTTTTTAAACGTTACGTGAAAAT
   GD33_BOLC.VG2.A  1367 TTTTGCTTCTCTGAACACAACGACGGCCGACAGATAACTTTTTAAACGTTACGTGAAAAT
   A12_BOLC.VG2.B   1308 TTAAAATTA--TGCACGTAATGATAACTCATGAGTT-TTTTTTGTTGATTAAATGACATT
25 GD33_BOLC.VG2.B  1307 TTAAAATTA--TGCACGTAATGATAACTCATGAGTT-TTTTTTGTTGATTAAATGACATT

   A12_BOLC.VG2.A   1428 GTATTGCACTATTGGGCCTTAGTAAAGGCTAATGAACACATTAAATGGGCTTATAGACGC
   GD33_BOLC.VG2.A  1427 GTATTGCACTATTGGGCCTTAGTAAAGGCTAATGAACACATTAAATGGGCTTATAGACGC
   A12_BOLC.VG2.B   1365 ATGTCCAAATTT---ATTTAAGGATATTTCATTAAGGTAAGTGAAAAAGACAAACAATAA
30 GD33_BOLC.VG2.B  1364 ATGTCCAAATTT---ATTTAAGGATATTTCATTAAGGTAAGTGAAAAAGACAAACAATAA

   A12_BOLC.VG2.A   1488 AATTGCATGGTATCTATCGGG-AAGCATAAATCGTCAAGTAATTCAATTTAGGGGACTCA
   GD33_BOLC.VG2.A  1487 AATTGCATGGTATCTATCGGG-AAGCATAAATCGTCAAGTAATTCAATTTAGGGGACTCA
   A12_BOLC.VG2.B   1422 AATAGGAAGTTTAAAATCATTTAAGCATATTTCAT----TAATGTAACTGAAAAGACA-A
35 GD33_BOLC.VG2.B  1421 AATAGGAAGTTTAAAATCATTTAAGCATATTTCAT----TAATGTAACTGAAAAGACA-A

   A12_BOLC.VG2.A   1547 TTTGCAATTTGTGCAAAGATTTTAGCTGTACAGAGTAGGGCTGAACATAAAATCCGTAAC
   GD33_BOLC.VG2.A  1546 TTTGCAATTTGTGCAAAGATTTTAGCTGTACAGAGTAGGGCTGGGCATAAAATCCGTAAC
   A12_BOLC.VG2.B   1477 GTAATAAATTAGGCAG---TTTTATTCGTTT----TATGATATTTCATAAA---TGCAAC
40 GD33_BOLC.VG2.B  1476 GTAATAAATTAGGCAG---TTTTATTCGTTT----TATGATATTTCATAAA---TGCAAC

   A12_BOLC.VG2.A   1607 CCGAAATCCGAACCGAACCGAAAAACCCGATCCGTACCCGATCCAAAATGTAAAAAATAC
   GD33_BOLC.VG2.A  1606 CCGAAATCCGAACCGAACCGAAAAACCCGATCCGTACCCGATCCAAAATGTAAAAAATAC
   A12_BOLC.VG2.B   1527 TGGAAAAGACAAACAAA--AAAAAATAGGGAGTTTAATTAAT--GAAATA-AGAACATTT
45 GD33_BOLC.VG2.B  1526 TGGAAAAGACAAACAAA--AAAAAATAGGGAGTTTAATTAAT--GAAATA-AGAACATTT

   A12_BOLC.VG2.A   1667 CTCAATGAATATTGTAGGGTGTTATAAAATATATCTGAACCCGAAGTGTTATTAACCGAA
   GD33_BOLC.VG2.A  1666 CTCAATGGATATTGTAGGGTGTTATAAAATATATCTGAACCCGAAGTGTTATTAACCGAA
   A12_BOLC.VG2.B   1582 TCAAATGAGTACTTC----TCTTTTAATA-ATAT---------AGATTTCATTAAT-GTA
50 GD33_BOLC.VG2.B  1581 TCAAATGAGTACTTC----TCTTTTAATA-ATAT---------AGATTTCATTAAT-GTA

   A12_BOLC.VG2.A   1727 CCCGAACGGATAATCCGAAAACCCGAAAAACCAAAAAATTCGAAAAAATATCCGAAAAAA
   GD33_BOLC.VG2.A  1726 CCCGAACGGATAATCCGAAAACCCGAAAAACCAAAAAACTCGAAAAAATATCCGAAAAAA
   A12_BOLC.VG2.B   1627 ACTGAAAAGACAAGTAATAAATTAGCCAGTTTTATTCGTTTTAGTATATTTC---ATAAA
55 GD33_BOLC.VG2.B  1626 ACTGAAAAGACAAGTAATAAATTAGCCAGTTTTATTCGTTTTAGCATATTTC---ATAAA

   A12_BOLC.VG2.A   1787 CTGATCCGAATGCGTAAATTAATATAAAATATAAATATTTGAAACATACA-TTGGATATG
   GD33_BOLC.VG2.A  1786 CTGATCCGAATGCGTAAATTAATATAAAATATAAATATTTGAAACATACA-TTGGATATG
   A12_BOLC.VG2.B   1684 TGCAACTGAAAAAGACAAGAAAAA-AAAATA-AAGAATTTAATTAATGAA-GTAAGAATA
60 GD33_BOLC.VG2.B  1683 TGCAACTGAAAAAGACAAGAAAAA-AAAATA-AAGAATTTAATTAATGAA-GTAAGAATA
```

```
A12_BOLC.VG2.A    1846 ATATCTAACAA-TAAGTATTTAAAATTTAAATAAATACTTTAAATACTCCATTATATACA
GD33_BOLC.VG2.A   1845 ATATCTAACAA-TAAGTATTTAAAATTTAAATAACTACTTTAAATACTCTATTATATACA
A12_BOLC.VG2.B    1741 TTTTCAAATGAGTATTTCTCTTTTAATAATATAGATTGATCAAAAATTATATTGGATA-A
GD33_BOLC.VG2.B   1740 TTTTCAAATGAGTACTTCTCTTTTAATAATATAGATTGATCAAAAATTATATTGGATA-A
5
A12_BOLC.VG2.A    1905 AA--GAAGTATATATTTTTTATGTTTTACTTTTGAATTTTAGATTTTACTATCGATATAT
GD33_BOLC.VG2.A   1904 AA--GAAGTATATATTTTTTATGTTTTACTTTTGAATTTTAGATTTTACTATCTATATAT
A12_BOLC.VG2.B    1800 GG--CAAGCA-AAATCCATTGAGAAGATCACAACAACAAAAGAAAGGAACAAACACACAA
GD33_BOLC.VG2.B   1799 GG--CAAGCA-AAATCCATTGAGAAGATCACAACAACAAAAGAAAGGAACAAACACACAA
10
A12_BOLC.VG2.A    1963 CCGAGCCGATCCGATA--CAATCCGAATCCGAATGTTATATGGCTACTTTGGATATATCT
GD33_BOLC.VG2.A   1962 CCGAGCCGATCCGATA--CAATCCGAATCCGAATGTTATATGGCTATTTTGGATATATCT
A12_BOLC.VG2.B    1857 CCACGGCGA-CAGATAACTAAAATGTACTGGACTATTGGGCCCTTAATATAGGCTAAT--
GD33_BOLC.VG2.B   1856 CCACGGCGA-CAGATAACTAAAATGTACTGGACTATTGGGCCCTTAATATAGGCTAAT--
15
A12_BOLC.VG2.A    2021 GAACCGATCCGAAACCGAAGTTGTTATATCCGAATCTCATCCGTACTTGTAAATTTACTA
GD33_BOLC.VG2.A   2020 GAACCGATCCGAAACCGAAGTTGTTATATCCGAATCTCATCCGTACTGGTAAATTTACTA
A12_BOLC.VG2.B    1914 GAACAGATT---AAATGGGCTAATTATAGACGTCATTGCATGGTATCCATCTATCTGGAA
GD33_BOLC.VG2.B   1913 GAACAGATT---AAATGGGCTAATTATAGACGTCATTGCATGGTATCCATCTATCTGGAA
20
A12_BOLC.VG2.A    2081 GA-ATGAGACATAGGGGATGATACAAAATAGAACCGAAATCCGAAAAAACCGATCCGAAC
GD33_BOLC.VG2.A   2080 GA-ATGAGACACGGGGGATGATACAAAATAGAACCGAAATCCGAAAAAACCGATCCGAAC
A12_BOLC.VG2.B    1971 GC-ATAGTAATTTG----TTTTAGGGGACTGGTTTGCAATTTTTACAGAT--ATTTTAGC
GD33_BOLC.VG2.B   1970 GC-ATAGTAATTTG----TTTTAGGGGACTGGTTTGCAATTTTTACAGAT--ATTTTAGC
25
A12_BOLC.VG2.A    2140 CCGAACGCCGAGGCCTAGTACAGAGTTCCGATCAGAGATACAGAGAATCCTTTGAGCGTC
GD33_BOLC.VG2.A   2139 C-GAACGCCCAGGCCTAGTACAGAGTTCCGATCAGAGATACAGAGAATCCTTTGAGCGTC
A12_BOLC.VG2.B    2024 T-----------GTACAGTACAGAG----------------AGAGAATCCTTTGAGCGTC
GD33_BOLC.VG2.B   2023 T-----------GTACAGTACAGAG----------------AGAGAATCCTTTGAGCGTC
30
A12_BOLC.VG2.A    2200 G---GCCTTGCTTTGTGCGTAGGGATCATCT-CCGACTAATTGTTGTAATTATAAAGGGC
GD33_BOLC.VG2.A   2198 G---GCCTTGCTTTGTGCGTAGGGATCATCT-CCGACTAATTGTTGTAATTATAAAGGGC
A12_BOLC.VG2.B    2057 GTCGGCCTTGCTTTGTGTGTAGGGATCTCCGACTGACTAATTGTTGTAACTACAAAAGGG
GD33_BOLC.VG2.B   2056 GTCGGCCTTGCTTTGTGTGTAGGGATCTCCGACTGACTAATTGTTGTAACTACAAAAGGG
35
A12_BOLC.VG2.A    2256 GGCC----GCGCCTTTTTAGCGGAGAGATGGCGAGCTCA---------CAGTTCAGGTAT
GD33_BOLC.VG2.A   2254 GGCC----GCGCCTTTTTAGCGGAGAGATGGCGAGCTCA---------CAGTTCAGGTAT
A12_BOLC.VG2.B    2117 CGCCCCCCGCGCCTTTTTACT----AGATGGCGAGCTCAT------CACAGTTTAGGTAT
GD33_BOLC.VG2.B   2116 CGCCCCCCGCGCCTTTTTACT----AGATGGCGAGCTCAT------CACAGTTTAGGTAT
40
A12_BOLC.VG2.A    2303 ACGCAGACGCCGTCGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
GD33_BOLC.VG2.A   2301 ACGCAGACGCCGTCGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
A12_BOLC.VG2.B    2167 ACCCAGACGCCGTCGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
GD33_BOLC.VG2.B   2166 ACCCAGACGCCGTCGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
45
A12_BOLC.VG2.A    2363 GAGCTCATCGACCTATGCAAACGATTCGGCAAGATCGTCAATACGAAGACCAATGTCGGC
GD33_BOLC.VG2.A   2361 GAGCTCATCGACCTATGCAAACGATTCGGCAAGATCGTCAATACGAAGACCAATGTCGGC
A12_BOLC.VG2.B    2227 GAGCTCATCGACCTATGCAAACGATTCGGCAAGATCGTCAATACGAAGACCAATGTCGGC
GD33_BOLC.VG2.B   2226 GAGCTCATCGACCTATGCAAACGATTCGGCAAGATCGTCAATACGAAGACCAATGTCGGC
50
A12_BOLC.VG2.A    2423 GCCAATCGCAACCAAGCCTTTGTCGAATTCGTAA---CAACT--------TTTTATTCTC
GD33_BOLC.VG2.A   2421 GCCAATCGCAACCAAGCCTTTGTCGAATTCGTAA---CAACT--------TTTTATTCTC
A12_BOLC.VG2.B    2287 GCCAATCGCAACCAAGCCTTTGTCGAATTCGTAA---GAACT--------TTTTATTCTC
GD33_BOLC.VG2.B   2286 GCCAATCGCAACCAAGCCTTTGTCGAATTCGTAA---CAACT--------TTTTATTCTC
55
A12_BOLC.VG2.A    2472 TTGGATCATCAGATTGTTTCT---------------------------------------
GD33_BOLC.VG2.A   2470 TTCGATCATCAGATTGTTTCTGAAG--------CTTTG-TGATTTCTATGATGAACTCTT
A12_BOLC.VG2.B    2336 TTGGATCATCAGATTGTTGCTTCCC--------CCCCAATGGATTTGGTTAGAAATTGAA
GD33_BOLC.VG2.B   2335 TTGGATCATCAGATTGTTGCTTCCC--------CCCCAATGGATTTGGTTAGAAATTGAA
60
A12_BOLC.VG2.A    2493 ------------------TTTTTTGGTATCTCGTTGTTGT-----TTTGGTAT-----T
```

```
GD33_BOLC.VG2.A  2521 -----------------TTTTTTTTGGTATCTCGTTGTTGT-----TTTGGTAT-----T
A12_BOLC.VG2.B   2388 -----------------TAAAAAAAGGACTCGAACTCTTTT-----TTTGGTATCTCGTT
GD33_BOLC.VG2.B  2387 -----------------TAAAAAAAGGACTCGAACTCTTTT-----TTTGGTATCTCGTT

5 A12_BOLC.VG2.A  2524 GTT-------GTTTTGAAGATGAAACTGTATACTTT------------------GATTCA
GD33_BOLC.VG2.A  2554 GTT-------GTTTTGAAGATGAAACTGTATACTTT------------------GATTCA
A12_BOLC.VG2.B   2426 GTT-------GTTTTGAAGATGAAACTGTATACTTT------------------GATTCA
GD33_BOLC.VG2.B  2425 GTT-------GTTTTGAAGATGAAACTGTATACTTT------------------GATTCA

10 A12_BOLC.VG2.A  2559 TATT---------------------CGCAGGGTGACGTGAATCAGGCAATATCAATGGTT
GD33_BOLC.VG2.A  2589 TATT---------------------CGCAGGGTGACGTGAATCAGGCAATATCAATGGTT
A12_BOLC.VG2.B   2461 TATT---------------------CGCAGGGTGACGTGAATCAGGCAATATCAATGGTT
GD33_BOLC.VG2.B  2460 TATT---------------------CGCAGGGTGACGTGAATCAGGCAATATCAATGGTT

15
```


SUBSTITUTE SHEET (RULE 26)

```
  A12_BOLC.VG2.A    2598 TCTTACTATGCTTCGTCTTCAGAGCCGGCTCAGATTCGAGGGAAGACTGTTTATATTCAG
  GD33_BOLC.VG2.A   2628 TCTTATTATGCTTCGTCTTCAGAGCCGGCTCAGATTCGAGGGAAGACTGTTTATATTCAG
  A12_BOLC.VG2.B    2500 TCTTACTATGCTTCGTCTTCAGAGCCGGCTCAGATTCGAGGGAAGACTGTTTATATTCAG
5 GD33_BOLC.VG2.B   2499 TCTTACTATGCTTCGTCTTCAGAGCCGGCTCAGATTCGAGGGAAGACTGTTTATATTCAG

  A12_BOLC.VG2.A    2658 TACTCTAATCGCCATGAGATTGTCAACAATCAGAGTCCTGGAGAGGTCCCTGGCAATGTC
  GD33_BOLC.VG2.A   2688 TACTCTAATCGGCATGAGATTGTCAACAATCAGAGTCCTGGAGAGGTCCCTGGCAATGTC
  A12_BOLC.VG2.B    2560 TACTCTAATCGCCATGAGATTGTCAACAATCAGAGTCCTGGAGAGGTCCCTGGCAATGTC
10 GD33_BOLC.VG2.B  2559 TACTCTAATCGCCATGAGATTGTCAACAATCAGAGTCCTGGAGAGGTCCCTGGCAATGTC

  A12_BOLC.VG2.A    2718 CTCTTGGTCACCTTTGAAGGAGTCCAATCTCACCATGTCTGCATCGATGTCATCCATCTG
  GD33_BOLC.VG2.A   2748 CTCTTGGTCACCTTTGAAGGAGTCCAATCTCACCATGTCTGCATCGATGTCATCCATCTG
  A12_BOLC.VG2.B    2620 CTATTGGTCACCTTTGAAGGAGTCCAATCTCACCATGTCTGCATCGATGTCATCCATCTG
15 GD33_BOLC.VG2.B  2619 CTATTGGTCACCTTTGAAGGAGTCCAATCTCACCATGTCTGCATCGATGTCATCCATCTG

  A12_BOLC.VG2.A    2778 GTATGTGAATATTCAGCTTACCTTCCACTATTGTTTCTT-GTTATGTTAAGTGATTTTTT
  GD33_BOLC.VG2.A   2808 GTATGTGAATATTCAGCTTACCTTCCACTATTGTTTCTT-GTTATGTTAAGTGATTTTTT
  A12_BOLC.VG2.B    2680 GTATGTGAATATTCAGCTTACCTTCCACTATTGTTTCTT-GTTAT----AGTGATTTTTT
20 GD33_BOLC.VG2.B  2679 GTATGTGAATATTCAGCTTACCTTCCACTATTGTTTCTT-GTTAT----AGTGATTTTTT

  A12_BOLC.VG2.A    2837 TCG-----------TTTCTTCG----AGTAGATTCTAATCTATGAAAATATTTCAACTTG
  GD33_BOLC.VG2.A   2867 TCG-----------TTTCTTCG----AGTAGATTCTAATCTATGAAAATATTTCAACTTG
  A12_BOLC.VG2.B    2735 TCG-----------TTTCTTCG----AGTAGATTCTAATCTATGAAAATATTTCAACTTG
25 GD33_BOLC.VG2.B  2734 TCG-----------TTTCTTCG----AGTAGATTCTAATCTATGAAAATATTTCAACTTG

  A12_BOLC.VG2.A    2882 TTGTTATTAGGCAAACTT-------CTTTTG--AGTGTATTTTTTTCC------------
  GD33_BOLC.VG2.A   2912 TTGTTATTAGGCAAACTT-------CTTTTG--AGTGTATTTTTTTCC------------
  A12_BOLC.VG2.B    2780 TTGTTATTAGGCAAACTT-------CTTTTG--AGTGTATTTTTTTCC------------
30 GD33_BOLC.VG2.B  2779 TTGTTATTAGGCAAACTT-------CTTTTG--AGTGTATTTTTTTCC------------

  A12_BOLC.VG2.A    2921 --------ACTTATTGTTAGACATAC-------AGTATGTC--------ACATACTATTG
  GD33_BOLC.VG2.A   2951 --------AGTTATTGTTAGACATAC-------AGTATGTC--------ACATACTATTG
  A12_BOLC.VG2.B    2819 --------AGTTATTGTTAGACATAC-------AGTATGTC--------ACATACTATTG
35 GD33_BOLC.VG2.B  2818 --------AGTTATTGTTAGACATAC-------AGTATGTC--------ACATACTATTG

  A12_BOLC.VG2.A    2958 TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCTCTT
  GD33_BOLC.VG2.A   2988 TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCTCTT
  A12_BOLC.VG2.B    2856 TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCTCTT
40 GD33_BOLC.VG2.B  2855 TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCTCTT

  A12_BOLC.VG2.A    3018 TATTATAATCTTTCTGGAGAGAGTTTGGAAAATAGTTTCATGTTCGTCATTCTCAT----
  GD33_BOLC.VG2.A   3048 TATTATAATCTTTCTGGAGAGAGTTTGGAAAATAGTTTCATGTTCGTCATTCTCAT----
  A12_BOLC.VG2.B    2916 TATTATAATCTTTCTGGAGAGAGTTTGGAAAATAGTTTCATGTTCGTCTTTCTCAT----
45 GD33_BOLC.VG2.B  2915 TATTATAATCTTTCTGGAGAGAGTTTGGAAAATAGTTTCATGTTCGTCTTTCTCAT----

  A12_BOLC.VG2.A    3074 ----------------------GGACAGGTGTCACTCTGCACTTATCTAGTCATTCTCTT
  GD33_BOLC.VG2.A   3104 ----------------------GGACAGGTGTCACTCTGCACTTATCTAGTCATTCTCTT
  A12_BOLC.VG2.B    2972 ----------------------GGACAGGTGTCACTCTGCACTTATCTAGTCATTCTCTT
50 GD33_BOLC.VG2.B  2971 ----------------------GGACAGGTGTCACTCTGCACTTATCTAGTCATTCTCTT

  A12_BOLC.VG2.A    3112 TTTTAGTCTCCTAA----------TTTGAGTTTATTTCGATTGATTTGCTTCCCTTAGTT
  GD33_BOLC.VG2.A   3142 TTTTAGTCTCCTAACCCGGCAACATTTGAGTTTATTTCGATTGATTTGCTTCCCTTAGTT
  A12_BOLC.VG2.B    3010 TTTTAGTCTCCTAA----------TTTGAGTTTATTTCGATTGATTTGCTTCCCTTAGTT
55 GD33_BOLC.VG2.B  3009 TTTTAGTCTCCTAA----------TTTGAGTTTATTTCGATTGATTTGCTTCCCTTAGTT

  A12_BOLC.VG2.A    3162 ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT
  GD33_BOLC.VG2.A   3202 ATTATCAATTTACTCCACTGTATTATATGGACATGGCTCATATCTAGTCCAAACTTTTGT
  A12_BOLC.VG2.B    3060 ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT
60 GD33_BOLC.VG2.B  3059 ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT
```

```
A12_BOLC.VG2.A    3222 TTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAGCTGCTG
GD33_BOLC.VG2.A   3262 TTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAGCTGCTG
A12_BOLC.VG2.B    3120 TTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAGCTGCTG
GD33_BOLC.VG2.B   3119 TTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAGCTGCTG
5
A12_BOLC.VG2.A    3282 GTTTCCAGGTTAGAGATATCGAGTTTTGTTTTCAAGGTCTTGCAT--------A--AATC
GD33_BOLC.VG2.A   3322 GTTTCCAGGTTAGAGATATCGAGTTTTGTTTTCAAGGTCTTGCCT--------A--AATC
A12_BOLC.VG2.B    3180 GTTTCCAGGTTAGAGATATCGAGTTTTGTTTTCAAGGTCTTGCCTTTTTCATAA--AATC
GD33_BOLC.VG2.B   3179 GTTTCCAGGTTAGAGATATCGAGTTTTGTTTTCAAGGTCTTGCCTTTTTCATAA--AATC
10
A12_BOLC.VG2.A    3332 TAATGATGTTTAAACTCATCGTCTGCCGTCATAACCTGCTCAGGCACTTGTTCAGTTTAC
GD33_BOLC.VG2.A   3372 TAATGATGTTTAAACTCATCGTCTGCCGTCATAACCTGCTCAGGCACTTGTTCAGTTTAC
A12_BOLC.VG2.B    3238 TAATGATGTTTAAACTCATCATCTGCCGTCATAACCTGCTCAGGCACTTGTTCAGTTTAC
GD33_BOLC.VG2.B   3237 TAATGATGTTTAAACTCATCATCTGCCGTCATAACCTGCTCAGGCACTTGTTCAGTTTAC
15
A12_BOLC.VG2.A    3392 TGATGTGGACACTGCCTTAGCGGCAAGGACTGCGCTGGATGGTAGAAGTATACCCAAGTA
GD33_BOLC.VG2.A   3432 TGATGTGGACACTGCCTTAGCGGCAAGGACTGCGCTGGATGGTAGAAGTATACCCAAGTA
A12_BOLC.VG2.B    3298 TGATGTGGACACTGCCTTAGCGGCAAGGACTGCGCTGGATGGTAGAAGTATACCCACGTA
GD33_BOLC.VG2.B   3297 TGATGTGGACACTGCCTTAGCGGCAAGGACTGCGCTGGATGGTAGAAGTATACCCACGTA
20
A12_BOLC.VG2.A    3452 TGCTCAAATCCTTCATTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
GD33_BOLC.VG2.A   3492 TGCTCAAATCCTTCATTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
A12_BOLC.VG2.B    3358 TGCTCAAATCCTTCATTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
GD33_BOLC.VG2.B   3357 TGCTCAAATCCTTCATTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
25
A12_BOLC.VG2.A    3512 CTTTTCTTTTGGTAAATTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
GD33_BOLC.VG2.A   3552 CTTTTCTTTTGGTAAATTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
A12_BOLC.VG2.B    3418 CTTTTCTTTTGGTAAATTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
GD33_BOLC.VG2.B   3417 CTTTTCTTTTGGTAAATTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
30
A12_BOLC.VG2.A    3572 AATGTCTTACTCAGCTCATACTGATCTAAATATCAAATTTCAGTCCCACCGCAGCAGGTA
GD33_BOLC.VG2.A   3612 AATGTCTTACTCAGCTCATACTGATCTAAATATCAAATTTCAGTCCCACCGCAGCAGGTA
A12_BOLC.VG2.B    3478 AATGTCTTACTCAGCCCACACTGATCTAAATATCAAATTTCAGTCCCACCGCAGCAGGTA
GD33_BOLC.VG2.B   3477 AATGTCTTACTCAGCCCACACTGATCTAAATATCAAATTTCAGTCCCACCGCAGCAGGTA
35
A12_BOLC.VG2.A    3632 GAGTTTTGAGTCCTCGCAAATGTGCCTCTCCATTGTTATACTGCTCTTACTGTTGGAAGC
GD33_BOLC.VG2.A   3672 GAGTTTTGAGTCCTCGCAAATGTGCCTCTCCATTGTTATACTGCTCTTACTGTTGGAAGC
A12_BOLC.VG2.B    3538 GAGTTTTGAGTCCTCGCAAATGTGCCATTCCATTGTTTTACTGCTCCTACTGTTGGAAGC
GD33_BOLC.VG2.B   3537 GAGTTTTGAGTCCTCGCAAATGTGCCATTCCATTGTTTTACTGCTCCTACTGTTGGAAGC
40
A12_BOLC.VG2.A    3692 CTTATGGTTGAATAGTTACTTCATGTGTTT-AATTCTCTTACGCACTCAGGGACTACACA
GD33_BOLC.VG2.A   3732 CTTATGGTTGAATAGTTACTTCATGTGTTT-AATTCTCTTACGCACTCAGGGACTACACA
A12_BOLC.VG2.B    3598 CTTATGGTTGAATAGTTGATTCATGTGTTTTGATTTTCTGACGTACTCAGGGACTACACA
GD33_BOLC.VG2.B   3597 CTTATGGTTGAATAGTTGATTCATGTGTTTTGATTTTCTGACGTACTCAGGGACTACACA
45
A12_BOLC.VG2.A    3751 AATCCATACCTTCCGGTGAATCAAACCGCTATGGACGGTTCTATGCAGGTATATTTTTCC
GD33_BOLC.VG2.A   3791 AATCCATACCTTCCGGTGAATCAAACCGCTATGGACGGTTCTATGCAGGTATATTTTTCC
A12_BOLC.VG2.B    3658 AATCCATATCTTCCGGTGAATCAAACTGCTATGGATGGTTCTATGCAGGTAAATATTTCC
GD33_BOLC.VG2.B   3657 AATCCATATCTTCCGGTGAATCAAACTGCTATGGATGGTTCTATGCAGGTAAATATTTCC
50
A12_BOLC.VG2.A    3811 TTTGATCTAATCTTATCTCCAACAGCTTATATTATTCTTACACTGAATGCGTTTGTATTA
GD33_BOLC.VG2.A   3851 TTTGATCTAATCTTATCTCCAACAGCTTATATTATTCTTACACTGAATGCGTTTGTATTA
A12_BOLC.VG2.B    3718 TCTT-TCTAGTCT-ATCTCCAACAGCTTAGACAATTCTTACATTGAATGCGTT-------
GD33_BOLC.VG2.B   3717 TCTT-TCTAGTCT-ATCTCCAACAGCTTAGACAATTCTTACATTGAATGCGTT-------
55
A12_BOLC.VG2.A    3871 CGTTACTGATCGTCTTTGTCCATCACTGCGTTCAGCCTGCTTTGGGTGCTGATGGAAAGA
GD33_BOLC.VG2.A   3911 CGTTACTGATCGTCTTTGTCCATCACTGCGTTCAGCCTGCTTTGGGTGCTGATGGAAAGA
A12_BOLC.VG2.B    3769 ----ACTGATTGTCTTTGAC-GTCATTGCATTCAGCCTGCTTTGGGTGCTGATGGAAAGA
GD33_BOLC.VG2.B   3768 ----ACTGATTGTCTTTGAC-GTCATTGCATTCAGCCTGCTTTGGGTGCTGATGGAAAGA
60
A12_BOLC.VG2.A    3931 GGGTTGAAACTCAGAGCAACGTCCTGCTTGCTTTGATTGAGAATATGCAGTACGCTGTCA
```

```
   GD33_BOLC.VG2.A  3971 GGGTTGAAACTCAGAGCAACGTCCTGCTTGCTTTGATTGAGAATATGCAGTACGCTGTCA
   A12_BOLC.VG2.B   3824 GGGTAGAAACTCAGAGCAACGTCCTACTTGCTTTGATTGAGAATATGCAATACGCTGTCA
   GD33_BOLC.VG2.B  3823 GGGTAGAAACTCAGAGCAACGTCCTACTTGCTTTGATTGAGAATATGCAATACGCTGTCA

 5 A12_BOLC.VG2.A   3991 CCGTGGATGTTCTTCACACGGTGAGGAACAACAAT-ATGCCTGTTTCTTCATCTCATCAT
   GD33_BOLC.VG2.A  4031 CCGTGGATGTTCTTCACACGGTGAGGAACAACAAT-ATGCCTGTTTCTTCATCTCATCAT
   A12_BOLC.VG2.B   3884 C----------------ACGGTGAGGAACACCATT-ATGCACGTTTCTGCATCTCAGCAT
   GD33_BOLC.VG2.B  3883 C----------------ACGGTGAGGAACACCATT-ATGCACGTTTCTGCATCTCAGCAT

10
   A12_BOLC.VG2.A   4050 GCTTTCATATCTCAAAATTAACTATT-CCTTTGCTTCTCATAGGTGTTTTCCGCTTATGG
   GD33_BOLC.VG2.A  4090 GCTTTCATATCTCAAAATTAACTATT-CCTTTGCTTCTCATAGGTGTTTTCCGCTTATGG
   A12_BOLC.VG2.B   3927 CCTTTCATTTCTCAAAATTAATTATTTCCTTTGCTTGTCATAGGTGTTTTCCGCTTATGG
   GD33_BOLC.VG2.B  3926 CCTTTCATTTCTCAAAATTAATTATTTCCTTTGCTTGTCATAGGTGTTTTCCGCTTATGG
15

   A12_BOLC.VG2.A   4109 AACTGTGCAGAAGATTGCAATATTTGAGAAAAATGGTTCAACGCAAGCCTTAATTCAATA
   GD33_BOLC.VG2.A  4149 AACTGTGCAGAAGATTGCAATATTTGAGAAAAATGGTTCAACGCAAGCCTTAATTCAATA
   A12_BOLC.VG2.B   3987 GACTGTGCAGAAGATTGCAATATTTGAGAAAAATGGTTCAACGCAAGCCTTAATTCAATA
20 GD33_BOLC.VG2.B  3986 GACTGTGCAGAAGATTGCAATATTTGAGAAAAATGGTTCAACGCAAGCCTTAATTCAATA

   A12_BOLC.VG2.A   4169 CTCTGGTACATGACCTTGATAATCTGAATACATATATGATTCTACTATACTTCTTTTGGT
   GD33_BOLC.VG2.A  4209 CTCTGGTACATGACCTTGATAATCTGAATACATATATGATTCTACTATACTTCTTTTGGT
   A12_BOLC.VG2.B   4047 CTCTGGTACATGACCTTGATGATCTGAATACATATATTATTATACTATACTTCTTTCGGT
25 GD33_BOLC.VG2.B  4046 CTCTGGTACATGACCTTGATGATCTGAATACATATATTATTATACTATACTTCTTTCGGT

   A12_BOLC.VG2.A   4229 TATATGGTGAATCTTCTTATGTGTATGCAAAGAGAAGAGCCTAAGAATTTGAATATAGTC
   GD33_BOLC.VG2.A  4269 TATATGGTGAATCTTCTTATGTGTATGCAAAGAGAAGAGCCTAAGAATTTGAATATAGTC
   A12_BOLC.VG2.B   4107 TATATGGCGAATCATCTTATGTGTATGC-------------TAAGAATTTGAAT--AGTC
30 GD33_BOLC.VG2.B  4106 TATATGGCGAATCATCTTATGTGTATGC-------------TAAGAATTTGAAT--AGTC

   A12_BOLC.VG2.A   4289 ATCGTTTGCTTCATGTTCCGTTTTATTGATTCTTCTGTAGACATTTCAACGGCGACGATG
   GD33_BOLC.VG2.A  4329 ATCGTTTGCTTCATGTTCCGTTTTATTGATTCTTCTGTAGACATTTCAACGGCGACGATG
   A12_BOLC.VG2.B   4152 ATCATTTGCTTCATGTTCCGTTTTCTTGATTCTTCTGTAGACATACCAACGGCGACAATA
35 GD33_BOLC.VG2.B  4151 ATCATTTGCTTCATGTTCCGTTTTCTTGATTCTTCTGTAGACATACCAACGGCGACAATA

   A12_BOLC.VG2.A   4349 GCGAAAGAAGCACTGGAGGGACACTGCATATATGACGG-AGGCTACTGTAAGCTTCGACT
   GD33_BOLC.VG2.A  4389 GCGAAAGAAGCACTGGAGGGACACTGCATATATGACGG-AGGCTACTGTAAGCTTCGACT
   A12_BOLC.VG2.B   4212 GCGAAAGAAGCATTGGAGGGACACTGCATATATGACGGGAGGCTACTGTAAGCTTCGACT
40 GD33_BOLC.VG2.B  4211 GCGAAAGAAGCATTGGAGGGACACTGCATATATGACGGGAGGCTACTGTAAGCTTCGACT

   A12_BOLC.VG2.A   4408 AACTTACTCTCGGCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTT-----
   GD33_BOLC.VG2.A  4448 AACTTACTCTCGGCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTT-----
   A12_BOLC.VG2.B   4272 AACATACTC--GTCATACTGATCTCAATGTAAAGGTACGTAAGATCAGTTGCTTTCTTGA
45 GD33_BOLC.VG2.B  4271 AACATACTC--GTCATACTGATCTCAATGTAAAGGTACGTAAGATCAGTTGCTTTCTTGA

   A12_BOLC.VG2.A   4463 ---ATGTACCCAAACAAAAA-GTGCAGTTAAAGAGAATAAGAAGTGATTTCTGCAAGAAC
   GD33_BOLC.VG2.A  4503 ---ATGTACCCAAACAAAAA-GTGCAGTCAAAGAGAATAATAAGTGATGTCTGCAAGAAC
   A12_BOLC.VG2.B   4330 TTTGTGCACCCAAACAAAAAAGCTTGGGGAAAGAGAATAATAACTTATGTCTGCAAGAAC
50 GD33_BOLC.VG2.B  4329 TTTGTGCACCCAAACAAAAAAGCTTGGGGAAAGAGAATAATAACTTATGTCTGCAAGAAC

   A12_BOLC.VG2.A   4519 CGCATGCT----TTTTTCTCATCAACAATTG--AGTGAAATCTGAATTCGACATTCTAGT
   GD33_BOLC.VG2.A  4559 CGCATGCT----TTTTTCTCATCAACAATTG--AGTGAAATCTGAATTCGACATTCTAGT
   A12_BOLC.VG2.B   4390 CATATGCTATTTTTTTTCTCTTCAACAATTG--TATGAAATCTGAATTCGACTTTCTAGT
55 GD33_BOLC.VG2.B  4389 CATATGCTATTTTTTTTCTCTTCAACAATTG--TATGAAATCTGAATTCGACTTTCTAGT

   A12_BOLC.VG2.A   4573 TGATTCAACTAATGATTTT-CGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC
   GD33_BOLC.VG2.A  4613 TGATTCAACTAATGATTTT-CGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC
   A12_BOLC.VG2.B   4448 TTATTCAACTAGTGATTTTTCGTCTTCTTTCTTCAGGCATTTAGTGACAAAAGCGGAGAC
60 GD33_BOLC.VG2.B  4447 TTATTCAACTAGTGATTTTTCGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCGGAGAC
```

```
A12_BOLC.VG2.A    4632 TACACACTGCCTGATCTAAGCCAACTGGTGGGCCAAAAGGTTCCAGGAGTGGCTGCAGCT
GD33_BOLC.VG2.A   4672 TACACACTGCCTGATCTAAGCCAACTGGTGGGCCAAAAGGTTCCAGGAGTGGCTGCAGCT
A12_BOLC.VG2.B    4508 TACACACTGCCTGATCTAAGCCAACTGGTGGGCCAAAAGGTTCCAGGAGTGGCCGCTGCT
GD33_BOLC.VG2.B   4507 TACACACTGCCTGATCTAAGCCAACTGGTGGGCCAAAAGGTTCCAGGAGTGGCCGCTGCT
5
A12_BOLC.VG2.A    4692 AGTGGGCCAACAGATGGTTGGCCCAATGGGCAGGTGCAGACTCAATACATGGG-------
GD33_BOLC.VG2.A   4732 AGTGGGCCAACAGATGGTTGGCCCAATGGGCAGGTGCAGACTCAATACATGGG-------
A12_BOLC.VG2.B    4568 AGTGGGCCAACAGATTGTTGGCACAATGGGCAGGTGCAGACTCAATACACAAG-------
GD33_BOLC.VG2.B   4567 AGTGGGCCAACAGATTGTTGGCACAATGGGCAGGTGCAGACTCAATACACAAG-------
10
A12_BOLC.VG2.A    4745 --AAGTTCGTATATGTACCCACC--AGCTGATCCCACAGGAGCTTCACCTTCTTCTGGTC
GD33_BOLC.VG2.A   4785 --AAGTTCGTATATGTACCCACC--AGCTGATCCCACAGGAGCTTCACCTTCTTCTGGTC
A12_BOLC.VG2.B    4621 --AAGTTCATATATGTACCCACTGTAAGCAAAGGTACTAGATTTTGATAAAGCGCGAGTT
GD33_BOLC.VG2.B   4620 --AAGTTCATATATGTACCCACTGTAAGCAAAGGTACTAGATTTTGATAAAGCGCGAGTT
15
A12_BOLC.VG2.A    4801 -ATCCTC---CTTATTATGGTTGATCCATGTACTGT---TTTCTACTTTATATAA-ATTT
GD33_BOLC.VG2.A   4841 -ATCCTC---CTTATTATGGTTGATCCATGTACTGT---TTTCTACTTTATATAA-ATTT
A12_BOLC.VG2.B    4679 TATTTTTAAATTTTTTTCAATTGACAAATATTTAGTAAATGTCATATTTTCATAT-ATTT
GD33_BOLC.VG2.B   4678 TATTTTTAATTTTTTTTCAATTGACAAATATTTAGTAAATGTCATATTTTCATAT-ATTT
20
A12_BOLC.VG2.A    4853 CGGATTATGTTGAAGACTAGTTTCCATGTTATAGGTTTCAGGTAAAAGTGTGTTGTTCTT
GD33_BOLC.VG2.A   4893 CGGATTATGTTGAAGACTAGTTTCCATGTTATAGGTTTCAGGTAAAAGTGTGTTGTTCTT
A12_BOLC.VG2.B    4738 -GTGTTTTATTTTATAAAAGACTTAAACTTTTTATCTTTATTTAT---CGTATT-TTATT
GD33_BOLC.VG2.B   4737 -GTGTTTTATTTTATAAAAGACTTAAACTTTTTATCTTTATTTAT---CGTATT-TTATT
25
A12_BOLC.VG2.A    4913 TTATTTGGTTAAATAAACA---TGGGAAAACCAATAATATTATAGTAGTATAATTTATAA
GD33_BOLC.VG2.A   4953 TTATTTGGTTAAATAAACA---TGGGAAAACCAATAATATTAT-----------------
A12_BOLC.VG2.B    4793 TTAAATGACTATTTATGTT---TAAAAAATTAAACTTTATTTCTTTAATGAATTAAGTTG
GD33_BOLC.VG2.B   4792 TTAAATGACTATTTATGTT---TAAAAAATTAAACTTTATTTCTTTAATGAATTAAGTTG
30
A12_BOLC.VG2.A    4970 ACATTGTGTATTTGTTTCTATTTGACTTTGTTATGCAGCTTCCGAAGTGTTAGAGCTT--
GD33_BOLC.VG2.A        ------------------------------------------------------------
A12_BOLC.VG2.B    4850 ATATAACTCTGATAAATTAATTTTATTATGTGGTTAATATTTTAATTAAAAAAAATTATA
GD33_BOLC.VG2.B   4849 ATATAACTCTGATAAATTAATTTTATTATGTGGTTAATATTTTAATTAAAAAAAATTATA
35
A12_BOLC.VG2.A    5028 --CTTTTTCTCGTACTTTAAACTGTTTCATTCCTGGATTTGTTTGAAGATCTGCAGACTT
GD33_BOLC.VG2.A        ------------------------------------------------------------
A12_BOLC.VG2.B    4910 TACTTTTAATAAAGATTTGTATTTTTCAATCAAAAAATTCAATTTTTTTTATGAATGCTT
GD33_BOLC.VG2.B   4909 TACTTTTAATAAAGATTTGTATTTTTCAATCAAAAAATTCAATTTTTTTTATGAATGCTT
40
A12_BOLC.VG2.A    5086 ACACTTTAGTGACTTTTGCTATCC
GD33_BOLC.VG2.A        ------------------------
A12_BOLC.VG2.B    4970 AAATTATATTAAGAAAAGAAAAGA
GD33_BOLC.VG2.B   4969 AAATTATATTAAGAAAAGAAAAGA
45
```

Protein sequence alignment of VG2 Brassica oleracea orthologues proteins from GD33 and A12. Protein sequences were predicted using the web based bioinformatic program FGENESH (http://mendel.cs.rhul.ac.uk/mendel.php?topic=fgen-file).

```
5  A12_BOLC.VG2.A     1 ---MASSQFRYTQTPSKVVHLRNLPWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE
   GD33_BOLC.VG2.A    1 ---MASSQFRYTQTPSKVVHLRNLPWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE
   A12_BOLC.VG2.B     1 --MASSSQFRYTQTPSKVVHLRNLPWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE
   GD33_BOLC.VG2.B    1 --MASSSQFRYTQTPSKVVHLRNLPWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE

10 A12_BOLC.VG2.A    58 FGDVNQAISMVSYYASSSEPAQIRGKTVYIQYSNRHEIVNNQSPGEVPGNVLLVTFEGVQ
   GD33_BOLC.VG2.A   58 FGDVNQAISMVSYYASSSEPAQIRGKTVYIQYSNRHEIVNNQSPGEVPGNVLLVTFEGVQ
   A12_BOLC.VG2.B    59 FGDVNQAISMVSYYASSSEPAQIRGKTVYIQYSNRHEIVNNQSPGEVPGNVLLVTFEGVQ
   GD33_BOLC.VG2.B   59 FGDVNQAISMVSYYASSSEPAQIRGKTVYIQYSNRHEIVNNQSPGEVPGNVLLVTFEGVQ

15
```

```
  A12_BOLC.VG2.A     118 SHHVCIDVIHLVFSAYGFVHKIATFEKAAGFQALVQFTDVDTALAARTALDGRSIPKYLL
  GD33_BOLC.VG2.A    118 SHHVCIDVIHLVFSAYGFVHKIATFEKAAGFQALVQFTDVDTALAARTALDGRSIPKYLL
  A12_BOLC.VG2.B     119 SHHVCIDVIHLVFSAYGFVHKIATFEKAAGFQALVQFTDVDTALAARTALDGRSIPTYLL
5 GD33_BOLC.VG2.B    119 SHHVCIDVIHLVFSAYGFVHKIATFEKAAGFQALVQFTDVDTALAARTALDGRSIPTYLL

  A12_BOLC.VG2.A     178 PEHVGSCNLRMSYSAHTDLNIKFQSHRSRDYTNPYLPVNQTAMDGSMQPALGADGKRVET
  GD33_BOLC.VG2.A    178 PEHVGSCNLRMSYSAHTDLNIKFQSHRSRDYTNPYLPVNQTAMDGSMQPALGADGKRVET
  A12_BOLC.VG2.B     179 PEHVGSCNLRMSYSAHTDLNIKFQSHRSRDYTNPYLPVNQTAMDGSMQ------------
10 GD33_BOLC.VG2.B   179 PEHVGSCNLRMSYSAHTDLNIKFQSHRSRDYTNPYLPVNQTAMDGSMQ------------

  A12_BOLC.VG2.A     238 QSNVLLALIENMQYAVTVDVLHTVFSAYGTVQKIAIFEKNGSTQALIQYSDISTATMAKE
  GD33_BOLC.VG2.A    238 QSNVLLALIENMQYAVTVDVLHTVFSAYGTVQKIAIFEKNGSTQALIQYSDISTATMAKE
  A12_BOLC.VG2.B     227 -----------------------VFSAYGTVQKIAIFEKNGSTQALIQYSDIPTATTAKE
15 GD33_BOLC.VG2.B   227 -----------------------VFSAYGTVQKIAIFEKNGSTQALIQYSDIPTATTAKE

  A12_BOLC.VG2.A     298 ALEGHCIYDGGYCKLRLTYSRHTDLNVKAFSDKSRDYTLPDLSQLVGQKVPGVAAASGPT
  G GD33_BOLC.VG2.A298 ALEGHCIYDGGYCKLRLTYSRHTDLNVKAFSDKSRDYTLPDLSQLVGQKVPGVAAASGPT
  A12_BOLC.VG2.B     264 ALEGHCIYDG------------------RLLDKSGDYTLPDLSQLVGQKVPGVAAASGPT
20 GD33_BOLC.VG2.B   264 ALEGHCIYDG------------------RLLDKSGDYTLPDLSQLVGQKVPGVAAASGPT

  A12_BOLC.VG2.A     358 DGWP--NGQVQTQYMGSSYMYPPADPTGASPSSGHPPYYG
  GD33_BOLC.VG2.A    358 DGWP--NGQVQTQYMGSSYMYPPADPTGASPSSGHPPYYG
  A12_BOLC.VG2.B     306 DCWH--NGQVQTQYTRSSYMYPL-----------------
25 GD33_BOLC.VG2.B   306 DCWH--NGQVQTQYTRSSYMYPL-----------------
```

DNA sequence alignment for the VG1 Brassica oleracea genes

```
A12     1 AAAAAGCAGGCTAAGAGTGTAAGCGACCCATTTCACATCGGATCTAAAAATCATATAATC
GD33    1 AAAAAGCAGGCTAAGAGTGTAAGCGACCCATTTCACATTGGATCTAAAAATCAAATAATC

A12    61 TAGAGTCTGACAAGAATCAACGTTGATAGAGAGTGCGAGAGAATAGTACAGAGCCAGTCG
GD33   61 TAGAGTGTGACAAGAATCAACGTTGATAGAGAGTGAGAGAGAATAGTACAGAGCCAGTCG

A12   121 TCGAGGTAGAAATGTTTGTCCTGAGCCCAAATGGTATAGTTGATATTGGGATTCCAGAAC
GD33  121 TCGAGGTAGAAATGTTTGTCCTGAGCCCAAATGGTATAGTTGATATTGGGATTCCAGAAC

A12   181 TTGTTATCACCCACTATCCATCTCTTAGCGGCCACCTCAGTCACCGGAGCTGCCGCAAGA
GD33  181 TTGTTATCACCCACTATCCATCTCTTAGCGGCCACCTCAGTCACCGGAGCTGCCGCAAGA

A12   241 AAAGCCAGAACCATTGCCGCAGCCATCAACGCAAGTCTAGCCATAGATCACTCTAGAAAG
GD33  241 AAAGCCAGAACCACTGCCGCAGCCATCAACGCAAGTCTAGCCATAGATCACTCTAGAAAG

A12   301 AAAAAGAAGCTCAGTGCTTAATGACTCTTAGTCTTATCTCTCAAGAAATAGCAGTTTTAT
GD33  301 AAAAAGAAGCTCAGTGCTTGATGACTCTTAGTCTTATCTCTCAAGAAATAGCAGTTTTAT

A12   361 GAATGAATGTATATGAGCGAGTGGGGCTGTTTCCAAGTCAAATCAGACATACAAATTTCT
GD33  361 GAGTGA--------------GTGGGGCTGTTTCCAAGTCAAATCAGACATACAAATTTCT

A12   421 CGTTTTACGGGGAAAAGATCTCTCAAGATCCGTAATTTCTTCAACTAACGTTTGGAATTT
GD33  407 CGTTTTACGGGGAAAAGATCTCTCAAGATCCGTAATTTCTTCAACTAACGTTTGGAATCT

A12   481 TATTACTACTTTAATCAACCTCGATGTTGACGATGAAGTACACGTTAATTAAAGGTACGC
GD33  467 TATTACTACTTTAATCAACCTCGATGTTGACGATGACGTACACGTTAATTAAAGGTACGC

A12   541 GTTATATAGTTAAAGGAGTAATTAGATTTTAATTGCTGGCATTATTATTATTTTTCTCTG
GD33  527 GTTATATAGTTAAGGGAGTAATTAGATTTTAACTGCTGGCATTATTATTATTTTTCTCTG

A12   601 GT----ATATT-------TTTAAAGGAATATTTTCTCACCTTTTAGGCGGGAATTTATTT
GD33  587 GTTATTACATTATTCTGTTTTGTAGTAATATTTTCTCACCTTTTAGGCGGGAATTTATTT

A12   650 CCATTTTCTTCTTTCAAGGGTTCAGAGATTCAATTCCTTTATTAACATACGAAATGAGTA
GD33  647 CCATTTTCTTCTTTCAAGGGTTCAGAGATTCAATTCCTTTATTAACATACGAAATGAGTA

A12   710 TATAGTATCTTTTGAATATTTAAATAATTAAAATAAAAATAAATTACCATAACAGTGATA
GD33  707 TATAGTATCTTTTAAATATTTAAATAATTAAAATAAAAATAAATTACTATAACAGTGATA

A12   770 CATATCAAAATTTCCAAAACTAATGTGAGTACTTATATTAGTTGTACTTCTTAGTAAAAT
GD33  767 CATATCAAAATTTCCAAAACTAATGTGAGTACTTATATTAGTTGTACTTCTTAGTAAAAT

A12   830 AACGAGATATGTTATGGCTCGTCGCATACAACAAAGCAGAGCATATGTTCCCGCAATAAC
GD33  827 AACGAGATATGTTATGGCTCGTCGCATACAGCAAAGCAGAGCATATGTTCCCGCAATAAC
```

```
    A12   890 ATGCACACGCAACTGATAGCACGTCGTAACCATTTCCGCGAACCAACCAATGGATATTGT
    GD33  887 ATGCACACGCAACTGATAGCACGTCGTAACCATTTCCGCGAACCAACCAATGGATATTGT
5
    A12   950 AATTCGATAAGACGACGTCGTGTTGTTGACCACGAATATTTCTTAAATTTCCTTTTATCT
    GD33  947 AATTCGATAAGACGACGTCGTGTTGTTGACCACGAATATTTCTTAAATTTCCTTTTATCT

10  A12  1010 TTTGTTTTTGACGTCATCACATTGGTTTATGTCCACACGATATCAGTCATAAAATCCACG
    GD33 1007 TTTGTTTTTGACGTCATGACATTGGTTTATGTCCACACGATATCAGTCATAAAATCCACG

    A12  1070 TAGGGAACGACTATCCATTACACCCATCCACGTGTTCTCATAACAGCTTCTAGTTCCATT
15  GD33 1067 TAGG-AACGACTATCCATTACACCCATCCACGTGTTCTCATAACAGCTTCTAGTTCCATT

    A12  1130 AGCTGAACGAACAATAGAAACAGAGCATCGAACTGAGAAAGAAGAAGAAGAAGGGAAGCA
    GD33 1126 AGCTGAACGAACAATAGAAACAGAGCATCGAACTGAGAAAGAAGAAGAAGAAGGGAAGCA
20
    A12  1190 AGAGAAATGGCGGCAGCGGCCATGGCCGTTCATCTCCCGAAACATTCATCGTTTCTTCCT
    GD33 1186 AGAGAAATGGCGGCAGCGGCCATGGCCGTTCATCTCCCGAAACATTCATCGTTTCTTACG
25
    A12  1250 AATCCCAAGCTTCCATTGAATCAAAACTCTAACTTCCTCGGTGTGTCCTTGAAGATCGGT
    GD33 1246 AAACCCAAGCATCCATCCGATCAAAACTCTAACTTTCTCGGTGGGTCCTTAAAGTTCGGT

30  A12  1310 CGACCCATGTCCGTTAACCGGAAAATGAAAGGTCCGGTTACGGTTTCAGCTGCTTCGACC
    GD33 1306 CGATCCATGTCTCTAAACCGGAAAACTAAAGGTCCAGTTATGGTTTCAGCTGCTTCGACC

    A12  1370 GTCGAAGGCGATCGAAGCAAACAGTTTTACATAAACTTCACTGGATTCCCATTTCCTCTT
35  GD33 1366 GTCGAAGGCGATCGAAGCAAACAGTTTTACATAAACTTCACTGGATTCCCATTTCCTCTT

    A12  1430 GGTCCTTTCCTTAACCGGCGCACCATCAGAACCGAGGTTAGGCTTTCTCCATCCCTCTTG
    GD33 1426 GGTCCTTTCCTTAACCGGCGCACCATCAGAACCGAGGTTAGGCTTTCTCCATCCCTCTTG
40
    A12  1490 AGTTTTGATTTGAACTCGTTGAGAATCCCATATGGAT---GTTATGCAGGCGGTTAAAGG
    GD33 1486 AGTTTTGATTTGAACTCGTTGAGAATCCCATATGGATAATGTTATGCAGGCGGTTAAAGG
45
    A12  1547 AAGCATATGGATGTTTGAACAAGAACAAGCTTTAGGTTTCAGCAGTGTCTCTACCAATAT
    GD33 1546 AAGCATATGGATGTTTGAACAAGAACAAGCTTTAGGTTTCAGCAGTGTCTCAACGAATAT

50  A12  1607 AAGAATGACTGTCATCAGACTCAAATCCGGTGGCTTATGGGTTCATGCCCCTATTGCTCC
    GD33 1606 AAGAATGACTGTCATCAAACTCAAATCCGGTGGCTTATGGGTTCATGCCCCTATTGCTCC

    A12  1667 CACCAAAGAGTGTATTCAGGTTCCCCCTTTTTATTCATTATCCTTGATAAAGTATCATCC
55  GD33 1666 CACCAAAGAGTGTATTCAGGTTCCCTCTTTT-ATTCATTATC--------AGTATCATCC

    A12  1727 TTTCTTATCATTTGATAA--TAACCAATGTCTAAACTCTTTTTTGGGTTGTTTTGCAAAC
    GD33 1717 TTTCTTATCATTTGATAAATTAACCAATGTCTAAACTCTATTTTGGGTTGTTTTGCTTAC
60
```


SUBSTITUTE SHEET (RULE 26)

```
   A12  1785 TTTTTCTTTGAGGCAGCTTATTGAGGAGTTGGGAGCTCCGGTTGAGTACATTGTCCTGCC
   GD33 1777 TTTTTCTTTGGGGCAGCTTATTGAGGAGTTGGGAGCTCCGGTTGAGTACATCGTCCTACC

 5 A12  1845 AACTTTTGCTTACGAGCACAAGATCTTCGTCGGTCCCTTCTCTAGAAAGTTCCCCAAGGC
   GD33 1837 AACCTTCGCTTACGAGCACAAGATCTTCGTCGGTCCCTTCTCTAGAAAGTTCCCCAAGGC

   A12  1905 TCAAGTATGGGTGGCGCCAAGACAATGGAGCTGGCCACTGAACTTACCACTCGAGTTTTT
10 GD33 1897 TCAAGTATGGGTGGCGCCAAGACAATGGAGCTGGCCACTGAACTTACCACTCGAGTTTTT

   A12  1965 CGGTATCTTTCGCGCTAAAACCATTAAAGACGGAGACTTATCTACCCCGTGGGCTGATGA
   GD33 1957 CGGTATCTTTCGCGCTAAAACCATTAAAGACGGTGACTTGTCTACCCCGTGGGCTGATGA
15
   A12  2025 GATCGAGCAGAAAGTCTTAAGCTCTCCTGAAGTCGGTACGTTCCTTCACTTTCTCAATCT
   GD33 2017 GATCGAGCAGAAAGTCTTAAGCTCTCCTGAAGTCGGTACGTTCCTTCACTTTCTCAATCT

20
   A12  2085 TGTTGGCCCTCCTTAGCCAAGCCTCAGTGAAACATTCGCCTAAACCCGCACTCTCCTCCT
   GD33 2077 TGTTGGCCCTCCTTAGCCAAGCCTCA-TAAAACATTCGCTTCAACCC---CTCACC----

25 A12  2145 CTTCCCTCCAAAATATTTGAAAAGAGTTATATATACATAGCCTCCCAAGTTATTAAAAAA
   GD33 2129 ---CCACCCAAATTTTTTGAAAAAAGTTACATATACATAGCC-CCAAATTTATTAAAAAA

   A12  2205 AACGTTTTGTCTCCCAAACTATTGAAATCTTTATCAAATTGTCTATAGCCTCCAAAATCC
30 GD33 2185 ATC----------------TATTGAAATCTTTATTAAATTATCTACGTCCCCTAGAATCT

   A12  2265 GAGGGTCGGTCCTGTACTTATGAAATAA-TAATATACAGGAATAGGACCGTATGTGGAAG
   GD33 2229 GAGGGTCGGCCCTGTACTTATGGAATAAATAATATACAGGAATTGGACCGTATGTGGAAG
35
   A12  2324 TAGCGTTCTACCATAAGCGTTCAAGAACTCTATTAGTCACTGATGCTGTGATCTTCGTCC
   GD33 2289 TAGCGTTCTACCATAAGCGTTCAAGAACTCTATTAGTCACCGACGCTGTGATCTTCGTCC

40
   A12  2384 CACGGAAGCCGCCATCGAGTATCAGCAGCGAGTCCTTGCTGGCCTCTGCTAAGAACGGAC
   GD33 2349 CAAAGAAGCCGCCATCGAGTATCAGCAGCGAGTCTTTGCTGGCTTCTGCTAAGAACGGAC

45 A12  2444 TGGCTGTGAAGATACTTAGCAAAGGCAAACAAGTACCTGATGACCCTGTCGTTGATACCC
   GD33 2409 TGGCTGTGAAGATACTTAGCAAAGGCAAACAAGTACCTAATGACCCTGTCGTTGATACCC

   A12  2504 CAAACACCCGCCAAAAAGGTCAGCCATCAGCTTCTTGAAATTCAAATTCATCAAACAAAA
50 GD33 2469 CAAACACCCGCCAAAAAGGTCAGCCATCAGCTTCTTGAAATTCAAATTCATCAAACAAA-

   A12  2564 GATCTCGATGGGTGTCTTTTAACGTGCAGGATGGGAAAGAATGGTGCTGCAAATCCTGTT
   GD33 2528 GATCTCAATGGGTGTCTTTTAACGTGCAGGATGGGAAAGAATGGTGCTGCAAATCCTGTT
55
   A12  2624 TCTTGGTCCGTCTAATCTCTTGGAGCCAAACGCGAGCTTCGCACAAATGTCACAGAAGCT
   GD33 2588 TCTCGGCCCATCTAATCTCTTGGAGCCAAACGCGAGCTTTGCGCAAATGTCACAGAAGCT

60
   A12  2684 GATCGTTTCTCCCATTGTCAAGACTCTGGTCTTTAGCAAAGTCCCTGAGAAGGTGAGGGA
```

```
GD33 2648 GATAGTTTCTCCCATTGTCAAGACTCTGGTCTTTAGCAAAGTCCCTGAGAAGGTGAGGGA

   A12 2744 CTGGATCGATGAGATAGCGAGTGACTGGAGATTCAAGAGGATAATCCCAGCTCATTTCGA
5  GD33 2708 CTGGATCGATGAGATAGCGAGTGACTGGAGGTTCAAGAGGATAATCCCAGCTCATTTCGA

   A12 2804 GGCTCCGGTAAACGCAGGGAGGTCAGAATTTCTAGCTGCGTTTGGGTTTCTTGATGATCT
   GD33 2768 GGCTCCGATAAACGCGGGGAGGTCAGAGTTTCTAGCTGCGTTCGGGTTTCTTGATGATCT
10
   A12 2864 TCTAGGGGAAAGATATGTGAACCGTCCTCCTTCGCTCTCTGTTCTCTTCACTTCGCTGAT
   GD33 2828 TCTAGGGGAAAGATATGTGAACCGTCCTCCTTCGCTCTCTGTTCTCTTTACTTCGCTGAT

15
   A12 2924 GGGTAAAGCAGCGAGCTATTTTCCTCCGGATGATATGAGGACTCTCTCTTCTCTTGATCA
   GD33 2888 GGGTAAAGCTGCCAGCTATTTTCCTCCGGATGATATGAGAACTCTCTCTTCTCTTGATCA

20 A12 2984 GTTCTTAGTCTCTGTTGGTGCTGTTAAGAAGACCGTCTCTGGTAGAAAACGAAGATGACG
   GD33 2948 GTTCTTAGTCTCTGTTGGTGCTGTTAAGAAGACCGTCTCTGGTAGAAAACGAAGATGACG

   A12 3044 GAACAAACCCAGCTTTCT
25 GD33 3008 GAACAAACCCAGCTTTCT
```

Protein sequence alignment of VG1 Brassica oleracea orthologues proteins from A12 and GD33. Protein sequences were predicted using the web based bioinformatic program FGENESH.

```
A12      1  MAAAAMAVHLPKHSSFLPNPKLPLNQNSNFLGVSLKIGRPMSVNRKMKGPVTVSAAST--
GD33     1  MAAAAMAVHLPKHSSFLTKPKHPSDQNSNFLGGSLKFGRSMSINRKTKGPVMVSAAST--

A12     59  ------VEGDRSKQFYINFTGFPFPLGPFLNRRTIRTEAVKGSIWMFEQEQALGFSSVST
GD33    59  ------VEGDRSKQFYINFTGFPFPLGPFLNRRTIRTEAVKGSIWMFEQEQALGFSSVST

A12    113  NIRMTVIRLKSGGLWVHAPIAPTKECIQLIEELGAPVEYIVLPTFAYEHKIFVGPFSRKF
GD33   113  NIRMTVIKLKSGGLWVHAPIAPTKECIQLIEELGAPVEYIVLPTFAYEHKIFVGPFSRKF

A12    173  PKAQVWVAPRQWSWPLNLPLEFFGIFRAKTIKDGDLSTPWADEIEQKVLSSPEVGIGPYV
GD33   173  PKAQVWVAPRQWSWPLNLPLEFFGIFRAKTIKDGDLSTPWADEIEQKVLSSPEVGIGPYV

A12    233  EVAFYHKRSRTLLVTDAVIFVPRKPPSSISSESLLASAKNGLAVKILSKGKQVPDDPVVD
GD33   233  EVAFYHKRSRTLLVTDAVIFVPKKPPSSISSESLLASAKNGLAVKILSKGKQVPNDPVVD

A12    293  TPNTRQKGWERMVLQILFLGPSNLLEPNASFAQMSQKLIVSPIVKTLVFSKVPEKVRDWI
GD33   293  TPNTRQKGWERMVLQILFLGPSNLLEPNASFAQMSQKLIVSPIVKTLVFSKVPEKVRDWI

A12    353  DEIASDWRFKRIIPAHFEAPVNAGRSEFLAAFGFLDDLLGERYVNRPPSLSVLFTSLMGK
GD33   353  DEIASDWRFKRIIPAHFEAPINAGRSEFLAAFGFLDDLLGERYVNRPPSLSVLFTSLMGK

A12    413  AASYFPPDDMRTLSSLDQFLVSVGAVKKTVSGRKRR
GD33   413  AASYFPPDDMRTLSSLDQFLVSVGAVKKTVSGRKRR
```

CLAIMS

1) A polynucleotide, particularly an isolated polynucleotide, comprising a nucleic acid molecule selected from the group consisting of

5 a. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 1 (A12 version of BolC.VG1.a);

b. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 2 (A12 version of BolC.VG2.a);

c. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID
10 NO: 3 (GD33 version of BolC.VG1.a);

d. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 4 (GD33 version of BolC.VG2.a);

e. a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) – d);

15 f. a nucleic acid molecule comprising a nucleotide sequence that deviates from the nucleotide sequence defined in any of a) – e) by the degeneracy of the genetic code;

wherein said nucleic acid molecule as defined in any of a) – f) upon expression in a plant or plant part, leads to a modified seed vigour.

20 2) A polynucleotide, particularly an isolated polynucleotide, according to claim 1, comprising a nucleic acid molecule selected from the group further consisting of

a. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 5 (truncated A12 allele of BolC.VG2.b);

b. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID
25 NO: 6 (truncated GD33 allele of BolC.VG2.b);

c. a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) – b);

d. a nucleic acid molecule comprising a nucleotide sequence that deviates from the nucleotide sequence defined in any of a) – c) by the degeneracy of the genetic code;

SUBSTITUTE SHEET (RULE 26)

wherein said nucleic acid molecule as defined in any of a) – d) upon expression in a plant or plant part, leads to a modified seed vigour.

3) A polynucleotide, particularly an isolated polynucleotide, comprising a nucleic acid
5 molecule selected from the group consisting of

a. nucleic acid molecule comprising a nucleotide sequence that has at least 60% sequence identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

10 b. nucleic acid molecule comprising a nucleotide sequence that has at least 80% sequence identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

c. nucleic acid molecule comprising a nucleotide sequence that has at least 90%
15 sequence identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

d. nucleic acid molecule comprising a nucleotide sequence that has at least 95% sequence identity to any of the sequences as depicted in the group comprising : SEQ
20 ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6;

e. nucleic acid molecule comprising a nucleotide sequence that has at least 98% sequence identity to any of the sequences as depicted in the group comprising : SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID
25 NO:6;

f. nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) – e);

g. nucleic acid molecule comprising a nucleotide sequence that deviates from the nucleotide sequence defined in any of a) – f) by the degeneracy of the genetic code;

30 wherein said nucleic acid molecule as defined in any of a) – g) upon expression in a plant or plant part, leads to a modified seed vigour.

4) A polynucleotide according to claims 1 to 3 wherein the modified seed vigour phenotype is characterized by a further phenotype selected in the group comprising: modified speed of germination, modified speed of seedling emergence, modified uniformity of seed germination, modified uniformity of seedling emergence, modified percentage of seed germination, modified tolerance of the seed vis-a-vis external environmental and/or maternal conditions, modified sensitivity to ABA or modified content of ABA.

5) A polynucleotide, particularly an isolated polynucleotide, comprising a nucleic acid molecule selected in the group comprising:

a. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 3;
b. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 4;
c. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 6 ;
d. a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) – c) ;
e. a nucleic acid molecule comprising a nucleotide sequence that deviates from the nucleotide sequence defined in any of a) – d) by the degeneracy of the genetic code;

wherein said nucleic acid molecule as defined in any of a) – e) upon expression in a plant or plant part, leads to an increased seed vigour.

6) A polynucleotide according to claim 5, wherein the increased seed vigour phenotype is characterized by a further phenotype selected in the group comprising: increased speed of germination, increased speed of seedling emergence, increased uniformity of seed germination, increased uniformity of seedling emergence, increased percentage of seed germination, increased tolerance of the seed vis-a-vis

external environmental and/or maternal conditions, decreased sensitivity to ABA or decreased content of ABA.

7) A polynucleotide, particularly an isolated polynucleotide, comprising a nucleic acid
5 molecule selected in the group comprising:

a. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 1;

b. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 2;

10c. a nucleic acid molecule comprising a nucleotide sequence as depicted in SEQ ID NO: 5;

d. a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes to the nucleic acid molecule of any of a) to c) ;

e. a nucleic acid molecule comprising a nucleotide sequence that deviates from the
15 nucleotide sequence defined in any of a) – d) by the degeneracy of the genetic code;

wherein said nucleic acid molecule as defined in any of a) – e) upon expression in a plant or plant part, leads to a decreased seed vigour.

8) An expression cassette comprising a polynucleotide of any of the preceding
20 claims.

9) A vector molecule comprising the expression cassette according to claim 8.

10) Use of a polynucleotide according to claims 1 to 3 for modifying seed vigour.

25

11) A method for modifying the seed vigour comprising introgressing through crossing or by plant transformation techniques to and expressing in a plant or plant

part a polynucleotide, an expression cassette or a vector molecule of any of the claims 1 to 10.

12) A method for producing seed with modified seed vigour comprising:

5 a. obtaining a first plant verified to contain the polynucleotide of any of the claims 1 to 7;
b. crossing said first plant with a second plant verified to lack the said polynucleotide ; and
c. identifying a plant seed resulting from the cross exhibiting a modified seed vigour as
10 compared to seeds delivered by the second plant.

13) A plant or plant part which contains within its genome an introgression comprising the polynucleotide, the expression cassette or the vector molecule of any of claims 1 to 9 and exhibits a modification of seed vigour as compared to a plant or
15 plant part that does not comprise the said polynucleotide, expression cassette or vector molecule.

14) A plant or plant part which contains within its genome an introgression comprising a polynucleotide according to claim 5 and exhibits an increased seed
20 vigour as compared to a seed delivered by a plant or plant part that does not comprise the said polynucleotide.

15) A plant or plant part which contains within its genome an introgression comprising a polynucleotide according to claim 7 and exhibits a decreased seed
25 vigour as compared to a seed delivered by a plant or plant part that does not comprise the said polynucleotide.

16) A method for selecting plant or plant part with modified seed vigour, comprising the detection in the plant or plant part to be tested of the presence or absence of a
30 polynucleotide according to any of claims 1 to 7.

17) A method for selecting plant or plant parts with modified seed vigour, comprising contacting candidate plant or plant part with a selection tool selected from the group comprising the polynucleotides of any of claims 1 to 7.

5 18) A plant or plant part according to any of the preceding claims that are cultivated plant or cultivated plant part and are selected in the group comprising *Brassica oleracea, Brassica napus, Brassica rapa, Brassica campestris, Brassica juncea, Brassica nigra, Brassica pekinensis, Brassica chinensis, Brassica rosularis, Eruca vesicaria, Eruca sativa, Raphanus sativus, Lepidium sativum, Nasturtium officinale,*
10 *Wasabia japonica.*

19) A plant or plant part according to claims 13 to 15 or claim 18, wherein said plant is a hybrid plant.

15 20) A plant or plant part according to claim 19 obtainable from seed deposited at NCIMB under deposit number NCIMB 41951, or progeny thereof.

21) A non biological method for obtaining plant or plant part with modified seed vigour, comprising introducing a polynucleotide according to any of claims 1 to 7 into
20 the genome of said plant or plant part.

22) A method according to claim 16 comprising (a) obtaining a first plant verified to contain the polynucleotide of any of claims 1 to 7; (b) crossing said first plant with a second plant verified to lack the said polynucleotide; and (c) identifying a plant
25 resulting from the cross exhibiting modified seed vigour, and containing the said polynucleotide.

23) A method according to claim 22, wherein presence of the polynucleotide is verified by use of a molecular marker, particularly by a molecular marker physically

located in a position that is within or outside the genetic locus containing the polynucleotide.

24) A method according to claim 22, wherein presence of the polynucleotide is
5 verified by use of at least two molecular markers, particularly by at least two molecular markers physically located in a position that is flanking the genetic locus containing the polynucleotide.

25) A seed comprising a polynucleotide according to claims 1 to 7, wherein the said
10 seed is coated with any type of coating.

<u>Figures</u>

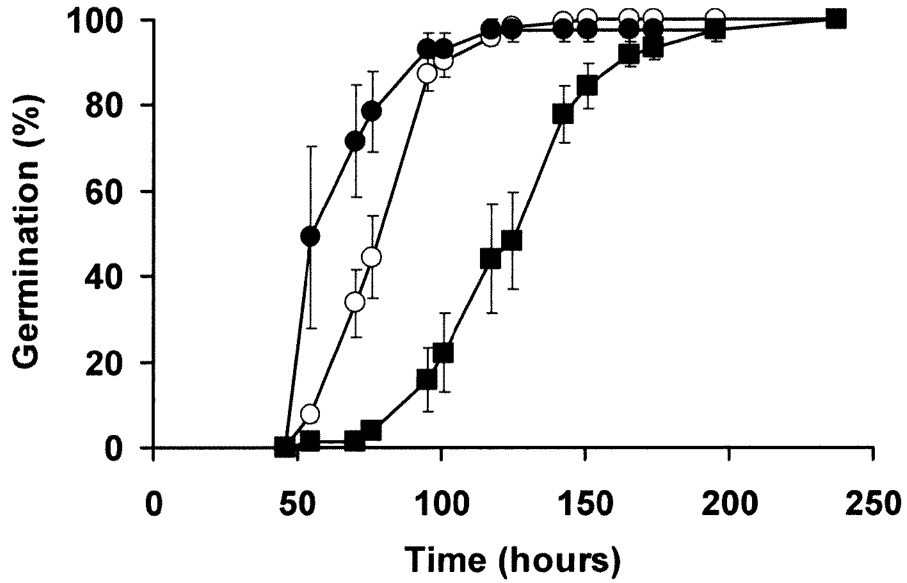
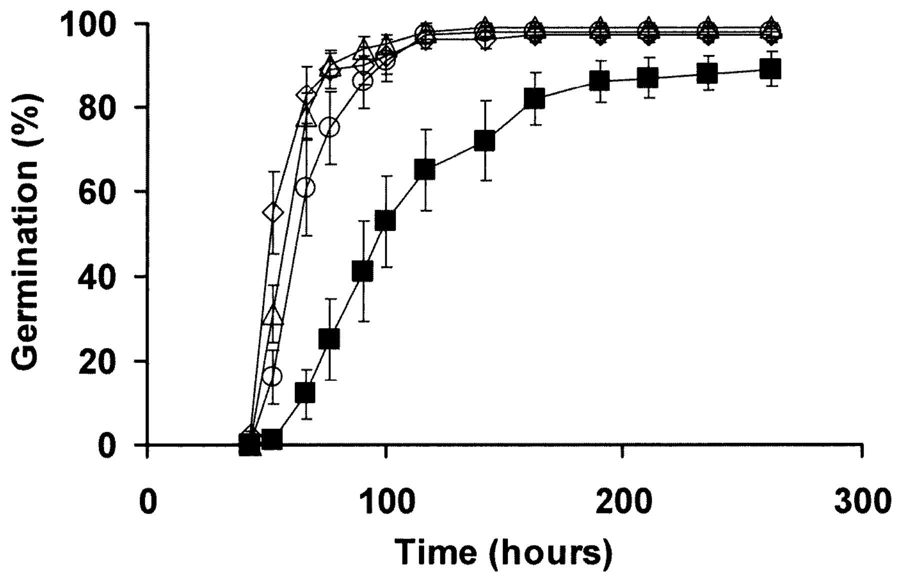


Fig. 1



5

Fig. 2

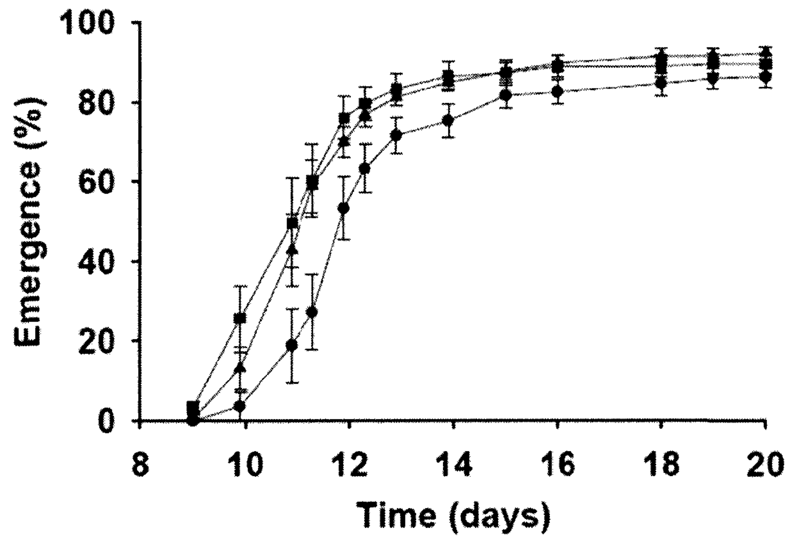
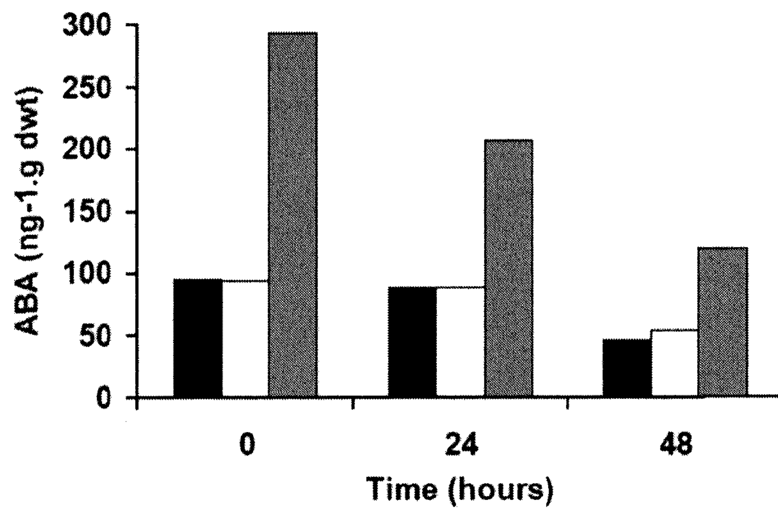


Fig. 3

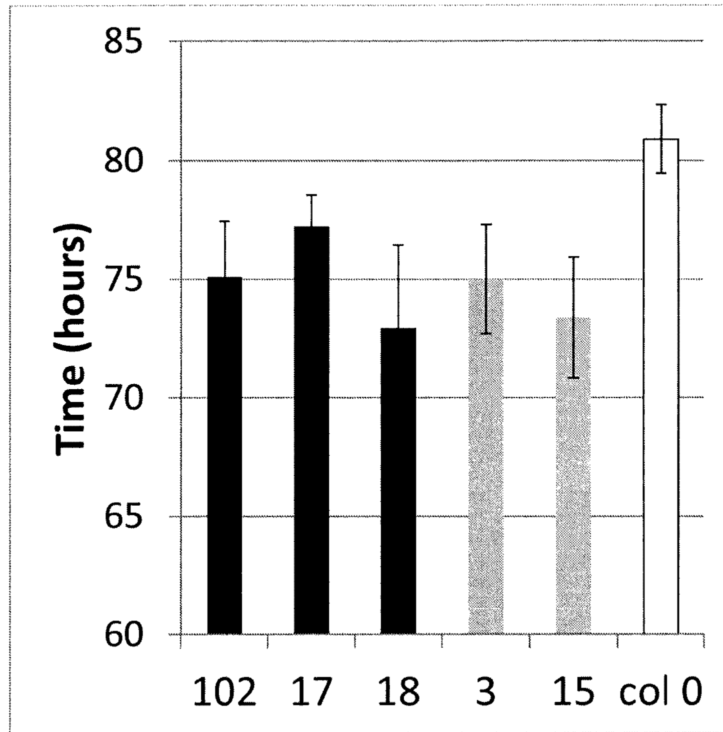
5



10

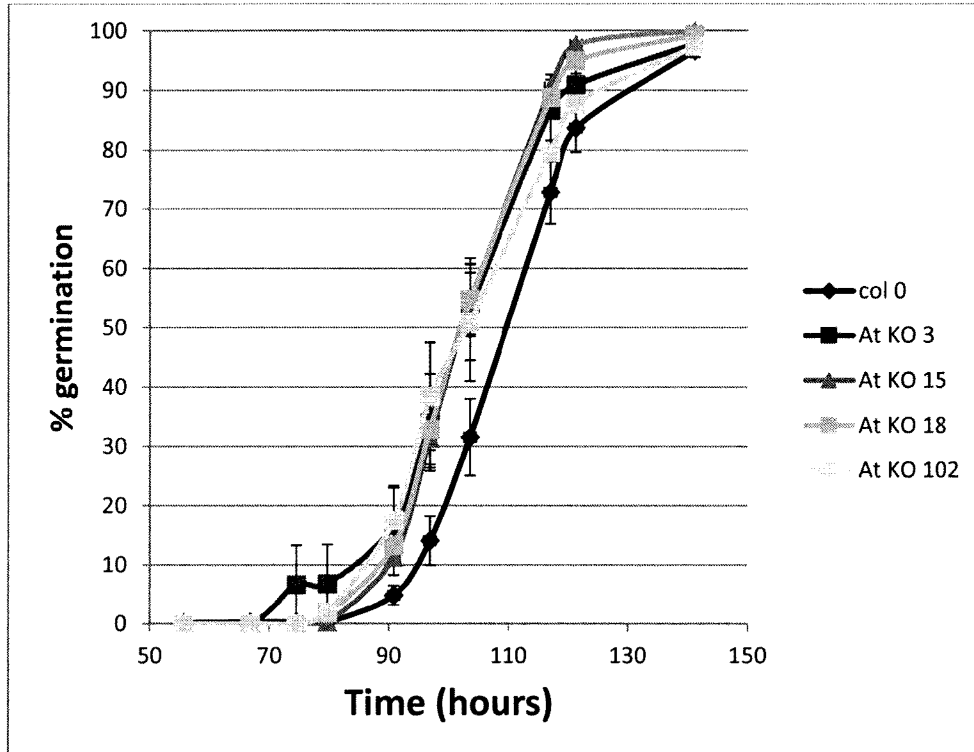
Fig. 4

SUBSTITUTE SHEET (RULE 26)



5

Fig. 5



5 Fig. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/053845

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/415 C12N15/82
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, BIOSIS, CHEM ABS Data, EMBASE, PAJ, WPI Data, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 9 June 2009 (2009-06-09), "Brassica rapa subsp. pekinensis clone KBrB043E02, complete sequence.", XP002695968, retrieved from EBI accession no. EM_HTG:FP340379 Database accession no. FP340379	3,4,7-25
Y	the whole document	1
Y	WO 2007/078286 A2 (CERES INC [US]; ALEXANDROV NICKOLAI [US]; BROVER VYACHESLAV [US]; MASC) 12 July 2007 (2007-07-12) claims 1-25	1,3,4, 7-25

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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	Date of mailing of the international search report
24 April 2013	24/07/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Marchesini, Patrizia

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2011/019391 A1 (ABBOTT & COBB INC [US]; LONG BRYANT JEROME [US]) 17 February 2011 (2011-02-17) claims 1-133 -----	1,3,4, 7-25
Y	FINCH-SAVAGE W E ET AL: "Towards a genetic understanding of seed vigour in small-seeded crops using natural variation in Brassica oleracea", PLANT SCIENCE, ELSEVIER IRELAND LTD, IE, vol. 179, no. 6, 1 December 2010 (2010-12-01), pages 582-589, XP027449543, ISSN: 0168-9452, DOI: 10.1016/J.PLANTSCI.2010.06.005 [retrieved on 2010-07-01] the whole document -----	1,3,4, 7-25

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2013/053845

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.:
because they relate to subject matter not required to be searched by this Authority, namely:

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:

see additional sheet

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.:

see additional sheet(s)

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.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1, 3, 4, 7-25(all partially)

   Relates to an isolated polynucleotide depicted by SEQ ID No. 1. Also relates to vectors, host cells, plants and plant parts comprising the polynucleotide, as well as methods for use.

   ---

2. claims: 1, 3, 4, 7-25(all partially)

   Relates to an isolated polynucleotide depicted by SEQ ID No. 2. Also relates to vectors, host cells, plants and plant parts comprising the polynucleotide, as well as methods for use.

   ---

3. claims: 1, 3-6, 8-25(all partially)

   Relates to an isolated polynucleotide depicted by SEQ ID No. 3. Also relates to vectors, host cells, plants and plant parts comprising the polynucleotide, as well as methods for use.

   ---

4. claims: 1, 3-6, 8-25(all partially)

   Relates to an isolated polynucleotide depicted by SEQ ID No. 4. Also relates to vectors, host cells, plants and plant parts comprising the polynucleotide, as well as methods for use.

   ---

5. claims: 2-4, 7-25(all partially)

   Relates to an isolated polynucleotide depicted by SEQ ID No.5. Also relates to vectors, host cells, plants and plant parts comprising the polynucleotide, as well as methods for use.

   ---

6. claims: 2-6, 8-25(all partially)

   Relates to an isolated polynucleotide depicted by SEQ ID No. 6. Also relates to vectors, host cells, plants and plant parts comprising the polynucleotide, as well as methods for use.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2013/053845

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007078286 A2	12-07-2007	AU 2005339695 A1	12-07-2007
		BR PI0520822 A2	23-06-2009
		CA 2632947 A1	12-07-2007
		CN 101370938 A	18-02-2009
		EP 1974038 A2	01-10-2008
		JP 2009521922 A	11-06-2009
		WO 2007078286 A2	12-07-2007
WO 2011019391 A1	17-02-2011	AU 2010282964 A1	09-02-2012
		CA 2769183 A1	17-02-2011
		EP 2464213 A1	20-06-2012
		JP 2012516136 A	19-07-2012
		WO 2011019391 A1	17-02-2011