



- (51) **International Patent Classification:**
C12N 15/82 (2006.01)
- (21) **International Application Number:**
PCT/EP2013/067519
- (22) **International Filing Date:**
23 August 2013 (23.08.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
1215046.2 23 August 2012 (23.08.2012) GB
- (71) **Applicant:** AARHUS UNIVERSITET [DK/DK]; Technology Transfer Office, Finlandsgade 29, Bygning 5361, DK-8200 Aarhus N (DK).
- (72) **Inventors:** **STUDER, Bruno**; ETH Zürich, Institut f. Agrarwissenschaften, CH-8092 Zurich (CH). **ASP, Torben**; Department of Molecular Biology and Genetics, Faculty of Science and Technology, Aarhus University, Forsøgsvej 1, DK-4200 Slagelse (DK).
- (74) **Agents:** **VAN DER HOFF, Hilary** et al.; Mewburn Ellis LLP, 33 Gutter Lane, London Greater London EC2V 8AS (GB).

(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, 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, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** Z LOCUS SELF-INCOMPATIBILITY ALLELES IN POACEAE

(57) **Abstract:** Z locus genes governing self-incompatibility in *lolium perenne* and other grass plants. Use of Z locus genes for predicting SI phenotype and identifying plants for breeding. Methods of modulating Z locus genes to alter SI phenotype. Methods of self-fertilising plants by modifying Z locus genes or gene products. Kinase inhibitors for inhibiting SI in grass plants, especially ryegrass.



Z Locus Self-Incompatibility Alleles in Poaceae

Field of the Invention

The invention relates to genes that determine self-incompatibility (SI) in plants, and to methods of altering the SI phenotype of plants by modulating expression of the genes or
5 changing the nucleic acid sequence at the gene locus in the plant. The invention further relates to plant breeding methods including steps of controlling SI, and to plants in which the SI phenotype is altered.

Background

Self-incompatibility (SI) is the genetically determined inability of a fertile hermaphrodite
10 seed plant to produce zygotes after self-pollination. It is an important genetic mechanism of fertile plants to prevent inbreeding after self-pollination. The resulting outcrossing is of major significance for evolutionary, diversification and domestication processes in plant species (Pandey, 1977). SI is distributed across half of the flowering plant families (East, 1940) and two major classes of SI systems have been described: gametophytic SI (GSI), where the SI
15 phenotype of the pollen is determined by its own gametophytic haploid genotype, and sporophytic SI (SSI), where the SI phenotype of the pollen is determined by the diploid genotype of the anther (the sporophyte). Both systems appear to have evolved separately (de Nettancourt, 1977).

The best characterized SI systems are those that are controlled by a single genetic locus,
20 the S-locus (Yang et al., 2008). In the single S-locus SSI system of *Brassica* spp., both pollen and stigma components for the S locus have been identified; the S-locus cysteine-rich protein gene (SCR) as the male determinant (Schopfer et al., 1999) and the S-locus receptor protein kinase gene (SRK) as the female determinant (Takasaki et al., 2000). In single S-locus GSI, the S-RNase system described in members of the Solanaceae, Rosaceae and Scrophulariaceae
25 (Cheng et al., 2006; Li et al., 1994; Murfett et al., 1994) involves cytotoxicity of stigma S protein S-RNase, which is crucial for the rejection of incompatible pollen (Lee et al., 1994). The pollen S protein has been identified to be encoded by an S-locus F-box gene (SLF) (Entani et al., 2003; Ushijima et al., 2003), which was confirmed by a transformation experiment in *Petunia inflata* (Sijacic et al., 2004). A mechanistically distinct single S-locus GSI system has been found in
30 Papaveraceae (Franklin-Tong and Franklin, 1992), where SI is mediated by a complex Ca^{2+} -dependent signalling network through interaction of a small pistil S-protein and a highly polymorphic transmembrane receptor PrpS in the pollen (Wheeler et al., 2009), resulting in

programmed cell death (Bosch and Franklin-Tong, 2007; de Graaf, 2006; Snowman et al., 2002; Thomas and Franklin-Tong, 2004).

The effectiveness of SI promotes and maintains high levels of heterozygosity in natural populations, thereby contributing to adaptive success, but also limits efficient production of inbred lines, a basic prerequisite for hybrid breeding schemes. Breeding for hybrid varieties is one of the most significant achievements for feed and food production. To date, many major crops are predominately produced as hybrid varieties. For example, hybrid production of rice (*Oryza sativa* L.), the world's most important staple food, has increased from 2.1 million ha in 1977 to 15.3 million ha in 1997, along with a 20 to 30% yield advantage over the best inbred rice varieties available (Li and Yuan, 2000; Wang et al., 2005). A more recent example of a steeper yield increase after moving from population towards hybrid breeding is rye (*Secale cereale* L.), where first hybrid varieties were released in the 1980s in Germany (Geiger and Miedaner, 2009). However, the best example for the impact of the transition from population to hybrid breeding is maize (*Zea mays* L.). While average yield increases have been limited by population improvement schemes, grain yield has been more than quadrupled since the introduction of hybrid breeding in the late 1920s (Duvick, 2005; Lamkey and Edwards, 1999).

Similar outcomes could be expected in other grass family species, for example perennial ryegrass (*Lolium perenne* L.). Due to SI, perennial ryegrass is currently improved as populations and synthetic varieties, only partially exploiting the genetically available heterosis. In contrast, forage grass varieties based on hybrid breeding schemes have the potential to outperform current populations and synthetic varieties through targeted exploitation of heterosis. Initial studies in perennial ryegrass found substantial levels of heterosis and hybrid performance for biomass yield (Posselt, 2010).

Besides the potential to maximize seed and biomass yield and to increase resistance/tolerance to biotic/abiotic stresses, hybrid varieties are genetically more homogeneous than populations and synthetic varieties. As a consequence, more uniform product qualities can be obtained. Additional benefits such as higher nutrient use efficiency or better root growth can be expected. Moreover, hybrids provide a simple means to protect intellectual property rights of breeders and guarantee a return on investment, as new seeds cannot be propagated from hybrids without a significant loss in performance and thus must be purchased for each planting.

Allogamous Poaceae species such as perennial ryegrass exhibit a GSI system which is controlled by at least two multiallelic and independent loci, S and Z (Lundqvist, 1954). GSI has been reported in both diploid and polyploid species within the tribes Triticeae, Poeae, and Paniceae, and seems to be monophyletic (Yang et al., 2008). The incompatibility response occurs when both the S and Z alleles of the haploid pollen grain are matched by identical alleles in the diploid pistil. The genetic positions of S and Z have been defined by linked markers but,

despite intense research efforts in the last decades, the genes determining the initial recognition mechanism are yet to be identified.

The S-locus has been mapped to linkage group (LG) 1 and the Z-locus to LG 2, in accordance with the Triticeae consensus map (Thorogood et al., 2002). These regions show
5 synteny to regions of rice chromosomes 5 and 4, respectively (Yang et al., 2008). More detailed microsynteny for the Z locus region with regions in rice, *Brachypodium distachyon* (L.) Beauv.) and sorghum (*Sorghum bicolor* (L.) Moench.) – all self-compatible species – has been demonstrated (Shinozuka et al., 2009). Recently, an additional SI-related locus F that showed genetic interaction with S was identified on LG 3 (Thorogood et al., 2002).

10 A putative S gene Bm2 was identified from Blue canary grass (*Phalaris coerulea* Desf.) (Li et al., 1994), but the expression of the Bm2 gene homolog was barely detectable in other SI grass species such as rye, bulbous barley (*Hordeum bulbosum* L.) and perennial ryegrass (Li et al., 1997). Later studies revealed that Bm2 encodes a thioredoxin-like protein and is located around 1cM from the S-locus (Baumann et al., 2000).

15 To isolate genes controlling SI in perennial ryegrass, cDNA-amplified fragment length polymorphism analysis and suppression subtractive hybridization were used to identify genes differentially expressed in self-incompatible and self-compatible pollen-stigma interactions (Van Daele et al., 2008b; Yang et al., 2009). Some differentially expressed fragments were homologous to genes involved in other SI systems, such as protein kinases, actins, a GTP-
20 binding protein and ubiquitin-related proteins (Van Daele et al., 2008a). In the most recent study, an additional candidate gene for Z containing a conserved domain of unknown function (DUF247) was found by sequencing of BAC clones covering the Z region (Shinozuka et al., 2009).

25 Hackauf and Wehling (2005) described a co-segregating marker for the Z locus in rye with sequence similarity to an ubiquitin-specific protease in a testcross population with a progeny of 204 individuals.

In bulbous barley, candidate genes for S were recently reported (HSLF1 and HSLF2), one showing specific expression in the pistil, the other increasing expression during the maturation of anthers (Kakeda, 2009).

30 However, despite SI having been recognised in grass family plants for over fifty years, and the long-standing problems that SI creates for controlled breeding and hybrid production in allogamous grass family species, S and Z have yet to be definitively characterised on either the pollen or stigma side in any member of the grass family (Yang et al. 2008; Klaas et al., 2011).

Summary of the Invention

35 We now report the identification of the Z SI locus genes in perennial ryegrass. Contrary to earlier publications that suggested other candidate Z locus genes, we found that the Z SI

locus is encoded by two glycerol kinase-like genes, LpGK1 and LpGK2, next to each other. Furthermore, we identified a variable region (VR) in each of the two genes. The VR distinguishes Z alleles and is predictive for Z locus incompatibility. The present invention constitutes the basis for the identification of SI alleles at the Z locus and, thus, the basis for utilising SI to control pollination in hybrid breeding schemes of Poaceae species, addressing many of the problems discussed above.

Z locus genes and encoded sequences

A first aspect of the invention is the isolated nucleotide sequence of a Z locus Poaceae gene. In perennial ryegrass, as well as in a number of other allogamous Poaceae species, the Z locus comprises a pair of Z locus genes, which we designate LpGK1 and LpGK2 respectively. Orthologues of gene LpGK2 are found in the majority of, possibly all, Poaceae, as well as in other plants such as Arabidopsis. The shorter gene, LpGK1, is found in fewer species and its presence is linked with SI. Thus, allogamous Poaceae species typically contain a pair of genes at the Z locus. LpGK1 and LpGK2 appear to be paralogues, both being glycerol kinase-like genes. The pair of genes may be arranged in tandem on the genomic DNA at the Z locus, with the coding sequence of each gene expressed from its corresponding promoter:

LpGK2 promoter – LpGK2 coding sequence – LpGK1 promoter – LpGK1 coding sequence

This arrangement is seen in perennial ryegrass and a number of other allogamous Poaceae, with the longer and more conserved LpGK2 gene located upstream of LpGK1. For convenience, the names LpGK1 and LpGK2 may be used to differentiate the two paralogues. LpGK1 may also be referred to as the downstream or short Z locus gene, while LpGK2 may be referred to as the upstream or long Z locus gene. The number of introns may also be used to differentiate the two genes, since LpGK1 was found to have two introns while LpGK2 was found to have three.

The invention includes Z locus alleles of LpGK1 and LpGK2. A nucleic acid according to the invention may comprise a nucleotide sequence of a perennial ryegrass Z locus gene, or a Z locus gene from another member of the grass family, Poaceae. Unless the context dictates otherwise, a Z locus gene referred to herein may be an LpGK1 gene or an LpGK2 gene. Examples of Z locus genes from perennial ryegrass and from other Poaceae are set out in this specification, and include the following:

(i) LpGK1 allele of perennial ryegrass Z locus haplotype P205. Genomic DNA including this allele and corresponding regulatory elements is shown in SEQ ID NO: 1. cDNA is shown in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 8 and SEQ ID NO: 23. The encoded amino acid sequence is shown in SEQ ID NOS: 3 and 5.

(ii) LpGK1 allele of perennial ryegrass Z locus haplotype B724. cDNA is shown in SEQ ID NO: 6 and SEQ ID NO: 9. The encoded amino acid sequence is shown in SEQ ID NO: 7.

(iii) LpGK2 allele of perennial ryegrass Z locus haplotype P226. cDNA is shown in
5 SEQ ID NO: 10. The encoded amino acid sequence is shown in SEQ ID NO: 11.

(iv) LpGK2 allele of perennial ryegrass Z locus haplotype P205. Genomic DNA including this allele and corresponding regulatory elements is shown in SEQ ID NO: 1. cDNA is shown in SEQ ID NO: 12. The encoded amino acid sequence is shown in SEQ ID NO: 13.

(v) LpGK2 allele of barley (*Hordeum vulgare* L.). cDNA is shown in SEQ ID NO: 14.

10 (vi) LpGK2 allele of Brachypodium. cDNA is shown in SEQ ID NO: 15. Amino acid sequence is shown in SEQ ID NO: 19.

(vii) LpGK2 allele of rice. cDNA is shown in SEQ ID NO: 16. Amino acid sequence is shown in SEQ ID NO: 20.

(viii) LpGK2 allele of sorghum. cDNA is shown in SEQ ID NO: 17. Amino acid
15 sequence is shown in SEQ ID NO: 22.

(ix) LpGK2 allele of maize. cDNA is shown in SEQ ID NO: 18. Amino acid sequence is shown in SEQ ID NO: 21.

(x) LpGK1 allele of haplotype P226. Coding DNA is shown in SEQ ID NO: 24. Amino acid sequence is shown in Figures 11 and 12 (SEQ ID NO: 25 and SEQ ID NO: 26).

20 (xi) LpGK2 allele of haplotype S089. cDNA is shown in SEQ ID NO: 27. Amino acid sequence is shown in SEQ ID NO: 28.

(xii) LpGK2 allele of haplotype S065. cDNA is shown in SEQ ID NO: 29. Amino acid sequence is shown in SEQ ID NO: 30.

25 (xiii) LpGK2 allele of haplotype S027. cDNA is shown in SEQ ID NO: 31. Amino acid sequence is shown in SEQ ID NO: 32.

(xiv) LpGK2 allele of haplotype S021. cDNA is shown in SEQ ID NO: 33. Amino acid sequence is shown in SEQ ID NO: 34.

30 Nucleic acids comprising any of the sequences provided, fragments of the sequences, probes or primers based on these sequences, and polypeptides encoded by the nucleic acids, are all aspects of the invention.

A Z locus gene contributes to the compatibility phenotype of a plant in which it is expressed. The compatibility phenotype of the plant refers to its ability to self-fertilise, to fertilise other plants and be fertilised by other plants. A plant may be self-incompatible or self-compatible, and the identity and expression of Z locus genes contribute to this self-incompatible or self-compatible phenotype. A functional Z locus gene is one that is capable of interacting
35 with another Z locus gene to inhibit fertilisation or setting of seed in a plant. This may confer an SI phenotype on the plant. Interaction may take place on a DNA, RNA and/or polypeptide level.

A functional allele of a Z locus gene may express a polypeptide in the pistil and/or the pollen that is capable of interacting with a polypeptide encoded by a corresponding Z locus gene expressed in the pistil and/or pollen. For example, Z locus allele expressed in pollen may encode a polypeptide that is capable of interacting with a polypeptide encoded by a Z locus allele expressed in the pistil. Interaction leads to incompatibility, and where the pollen and pistil are of the same plant, therefore leads to self-incompatibility. For example, interaction may inhibit pollen tube growth into the pistil. Conversely, a Z locus allele expressed in the pistil may encode a polypeptide that is capable of interacting with a polypeptide encoded by a Z locus allele expressed in the pistil, leading to incompatibility, e.g. by inhibiting pollen tube growth into the pistil. As described in more detail elsewhere herein, a Z locus gene may interact with another Z locus gene through one or more conserved regions and/or through the variable regions. For example, at the nucleic acid and/or polypeptide level, the conserved regions and variable regions of two interacting Z locus genes or gene products may come into contact and bind one another. A functional Z locus gene may be one which, on expression in a plant e.g. in stigma or pollen, interacts with another expressed Z locus gene in the same or a different plant, to inhibit fertilisation. As noted, interaction may be between nucleic acid sequences or encoded polypeptides. Interaction may take place between Z locus genes having matching variable regions, e.g. identical Z locus genes. A functional Z locus gene may encode a polypeptide that has kinase activity, e.g. a glycerol kinase.

Combinations of Z alleles present in a particular plant determine whether the plant is self-fertile, by determining whether or not pollen from the plant is able to fertilise the same plant to produce seed, i.e. whether the plant is self-compatible or self-incompatible. Combinations of Z alleles in two different plants of the same species determine whether one plant is able to fertilise the other.

A nucleic acid according to the invention may comprise the nucleotide sequence of two adjacent Z locus genes, LpGK1 and LpGK2, or it may comprise only a single Z locus gene, LpGK1 or LpGK2. Nucleic acid according to the invention may comprise a nucleotide sequence of a Z locus gene from any allogamous Poaceae species, for example a gene sequence shown in the figures or in the accompanying sequence listing, or may comprise a variant, such as a mutant, allele, orthologue or derivative. A variant may retain a functional characteristic of the wild-type sequence, for example so that the compatibility phenotype of a plant containing the variant gene is unchanged. Alternatively a variant gene may have one or more altered functional characteristics and a plant containing the variant gene may have an altered compatibility phenotype. A Z locus gene in accordance with the invention may be a functional Z locus gene. In other embodiments, however, it may be a Z locus gene that is not capable of interacting with another Z locus gene as described. Such non-functional Z locus genes may be linked with self-compatibility phenotypes and with ability of a plant to fertilise and be fertilised by

other plants. A non-functional Z locus gene may be a Z locus gene that encodes a polypeptide that lacks kinase activity.

Nucleic acid according to the invention may comprise a variant nucleotide sequence that is at least 70 % identical to a Z locus nucleotide sequence shown in any of the drawings or in the accompanying sequence listing, e.g. at least 80 % identical, at least 90 % identical, at least 95 % identical, at least 98 % identical or at least 99 % identical. It may encode an amino acid sequence that is encoded by a nucleotide sequence set out in any of the figures or in the accompanying sequence listing. It may encode an amino acid sequence that is at least 70 % identical to an amino acid sequence encoded by such a nucleotide sequence, e.g. at least 80 % identical, at least 90 % identical, at least 95 % identical, at least 98 % identical or at least 99 % identical.

While a polypeptide or nucleic acid sequence may share for example at least 70 % sequence identity overall with a Z locus sequence shown herein, it may share a greater percentage identity in the conserved regions, for example 90%, 95 %, 98 % or 99 % identity in each of the conserved regions.

In some embodiments, a conserved region of the Z locus gene is retained without mutation, so that the gene comprises the conserved region of a wild-type Z locus allele, such as any of the alleles of SEQ ID NOS: 1-2, 4, 6, 8-10, 12, 14-18, 23-24, 27, 29, 31 and 33. Alternatively, conserved regions may be retained with only minor variation. For example, a Z allele may comprise a coding sequence in which the conserved regions are both at least 90 %, 95 %, 98 % or 99 % identical to the corresponding conserved regions of a wild-type Z locus allele, for example a Z allele of any of SEQ ID NOS: 1-2, 4, 6, 8-10, 12, 14-18, 23-24, 27, 29, 31 or 33. It may encode an amino acid sequence in which the conserved regions are at least 90 %, 95 %, 98 % or 99 % identical to the conserved regions of an amino acid sequence encoded by a wild-type Z locus allele, for example a Z allele shown in any of SEQ ID NOS: 3, 5, 7, 11, 13, 19-22, 25-26, 28, 30, 32 and 34.

In some embodiments, sequence variation is restricted or mainly restricted to the VR. The VR of a Z locus allele may differ from VR sequences shown in the figures or in the accompanying sequence listing, by containing one more nucleotide insertions, deletions or substitutions. The VR may optionally be deleted, or replaced with a nucleotide sequence that is less than 90 %, less than 80%, less than 70 % or less than 50 % identical with a wild-type VR such as the VR of an allele sequence shown in any of SEQ ID NOS: 1-2, 4, 6, 8-10, 12, 14-18, 23-24, 27, 29, 31 and 33. Nucleic acid according to the invention may encode an amino acid sequence comprising a VR that is deleted or is less than 90 %, less than 80 %, less than 70 % or less than 50 % identical with a wild-type VR amino acid sequence such as a VR encoded by an allele sequence shown in any of SEQ ID NOS: 3, 5, 7, 11, 13, 19-22, 25-26, 28, 30, 32 or 34.

In other embodiments, a Z locus allele comprises a VR that is substantially unchanged from wild-type. It may comprise the VR of a wild type Z locus allele such as an allele shown in any of SEQ ID NOS: 1-2, 4, 6, 8-10, 12, 14-18, 23-24, 27, 29, 31 or 33, or it may comprise a VR that is at least 90 % or at least 95 % identical with a VR of an allele sequence shown in any of SEQ ID NOS: 1-2, 4, 6, 8-10, 12, 14-18, 23-24, 27, 29, 31 or 33. The nucleic acid may encode an amino acid sequence comprising a VR that is at least 90 % or at least 95 % identical with a VR amino acid sequence encoded by a wild-type Z locus allele such as an allele sequence shown in any of SEQ ID NOS: 3, 5, 7, 11, 13, 19-22, 25-26, 28, 30, 32 or 34. For example, the VR nucleotide sequence or amino acid sequence may comprise only one or two substitutions, insertions or deletions of codons or residues respectively.

A Z locus nucleic acid may comprise a nucleotide sequence that comprises a VR flanked by upstream and downstream conserved regions, wherein each of the VR and conserved regions shares at least 90 %, 95 %, 98 % or 99 % identity with the VR and conserved regions respectively of a Z locus sequence shown in any of SEQ ID NOS: 1-2, 4, 6, 8-10, 12, 14-18, 23-24, 27, 29, 31 or 33. A Z locus nucleotide sequence may comprise the upstream conserved region, VR and downstream conserved region of any Z locus sequence shown herein. Polypeptides encoded by such nucleic acids are also included in the invention.

For example, nucleic acid according to the invention may comprise a nucleotide sequence that shares at least 70 % sequence identity overall with a Z locus sequence shown herein, and shares at least 90%, 95 %, 98 % or 99 % identity in each of the conserved regions. It may additionally share at least 90 % identity in the variable region.

As noted, a Z locus gene may be an LpGK1 gene or an LpGK2 gene. Where the gene is an LpGK1, stated % identities at the nucleic acid and amino acid level may be considered with reference to an LpGK1 sequence shown herein, e.g. P205 LpGK1. Where the gene is an LpGK2, the stated % identities may be considered with reference to an LpGK2 sequence shown herein, e.g. P226. Alternatively, for convenience, the P226 haplotype may be used as reference sequence for any Z locus allele.

Conserved and variable regions of particular Z locus alleles are shown in the drawings and in the accompanying sequence listing. A conserved region is located on each side of the VR. For example, with reference to the amino acid sequence encoded by the Z locus P226 haplotype LpGK2 allele, the VR is amino acids 189 to 194 (underlined in Figure 6 and 7 on cDNA and amino acid sequence, respectively). The VR is flanked by an upstream conserved region which is residues 170-188 and a downstream conserved region which is residues 195-211 (bold in Figure 6 and 7 on cDNA and amino acid sequence, respectively).

The conserved regions and VR of a Z locus gene can be identified by aligning the conserved regions of a sequence with the conserved regions of a reference sequence such as one shown herein. When the two sequences are aligned, the sequences will usually be strongly

aligned in one or both conserved regions, and the VR can be identified as the stretch of amino acids between the upstream and downstream conserved regions. The corresponding location of the VR and flanking conserved regions of other Z locus alleles can be determined by alignment, as illustrated in the sequence alignments in the drawings. The reference sequence
 5 used for alignment can be either an LpGK2 allele or an LpGK1 allele, depending on whether the sequence of interest is LpGK2 or LpGK1, since the best fit will be obtained aligning LpGK1 with LpGK1 and aligning LpGK2 with LpGK2. For example, for an LpGK1 allele, the reference sequence may be the LpGK1 P205 haplotype nucleic acid sequence shown in Figure 10 (SEQ ID NO: 23), the full length cDNA sequence for LpGK1 P205 (SEQ ID NO: 2 or SEQ ID NO: 4) or
 10 the encoded amino acid sequence of the P205 haplotype shown in Figure 2 or Figure 11 (SEQ ID NO: 3 or SEQ ID NO: 5). For an LpGK2 allele, the reference sequence may be the P226 haplotype nucleic acid sequence shown in Figure 6 (SEQ ID NO: 10) or the encoded amino acid sequence of the P226 haplotype shown in Figure 7 (SEQ ID NO: 11). Alternatively, the reference sequence may be the S089 LpGK2 haplotype nucleic acid sequence shown in SEQ
 15 ID NO: 27 or its encoded amino acid sequence SEQ ID NO: 28. Preferably, however, the LpGK2 P226 haplotype is used as the reference sequence for any alignment or sequence comparison, and any LpGK1 or LpGK2 Z locus gene or encoded amino acid sequence may be aligned with the P226 haplotype sequence.

Figure 7 illustrates alignment with the P226 haplotype LpGK2 allele and illustrates the
 20 conserved and variable regions. Alleles may be of variable length and therefore have a different residue numbering, but the corresponding regions and residue numbering can be identified by aligning the sequences. The VR of a nucleotide sequence is the region encoding the VR in the amino acid sequence, which can be determined for any Z locus allele, e.g. an LpGK2 allele, by alignment with the P226 LpGK2 allele. A conserved region of a nucleotide
 25 sequence is the sequence of residues encoding the upstream or downstream conserved region, which again can be defined by reference to the corresponding sequence of the Z locus P226 haplotype. A VR of a Z locus gene may thus be a sequence encoding amino acids 189 to 194 with reference to the residue numbering of the Z locus P226 haplotype LpGK2 allele sequence, and the upstream and downstream conserved regions may be a sequence encoding residues
 30 170-188 and 195-211 respectively, with reference to the residue numbering of the Z locus P226 haplotype LpGK2. The VRs and conserved regions of several Z locus genes are illustrated in the Figure 7 alignment with the Z locus P226 LpGK2 allele.

Figure 10 illustrates an alignment of LpGK1 alleles, and the conserved and variable regions are indicated. With reference to the numbering of the P205 LpGK1, the upstream
 35 conserved region is nucleotides 40-86. The downstream conserved region is nucleotides 136-189. The VR is the region between them, nucleotides 87-135. Figure 11 shows the corresponding amino acid sequences. With reference to the numbering of the P205 LpGK1, the

upstream conserved region is residues 14-32. The downstream conserved region is residues 27-63. The VR is the region between them, residues 33-46.

Figure 13a illustrates an alignment of the cDNA sequences of LpGK2 alleles S089, S027, S021 and S065. With reference to the numbering of the S089 LpGK2, the upstream conserved region is nucleotides 92-148. The downstream conserved region is nucleotides 191-241. The VR is the region between them, nucleotides 149-190.

Figure 13b illustrates an alignment of the amino acid sequences of the same LpGK2 alleles. With reference to the numbering of the S089 LpGK2 allele, the upstream conserved region is residues 31-49. The downstream conserved region is residues 64-80. The VR is the region between them, residues 50-63.

Polynucleotides which are not 100% identical to the sequences shown herein but fall within the scope of the invention can be obtained in a number of ways, for example by mutation or by isolation from other plants of the same or different species of Poaceae.

The nucleic acid typically comprises a coding region, which is expressed as an amino acid sequence. The coding region can be operably linked to one or more transcriptional control elements. The coding region may have the nucleotide sequence of a cDNA shown in one of the figures or the accompanying sequence listing or a variant of that sequence, as noted above. A transcriptional control element may be heterologous to the coding region, i.e. not the naturally occurring control element in the Z locus gene, and may be from a different species compared with the coding region. The transcriptional control element may be a promoter, for example an inducible promoter. Accordingly, nucleic acid of the invention may comprise one or more transcriptional control elements such as an inducible promoter, and a coding region of a Z locus gene, where expression from the gene is under control of the inducible promoter and/or other element or elements.

A further aspect of the invention is nucleic acid comprising a transcriptional control sequence for a Z locus Poaceae gene, optionally isolated from its native Z locus gene.

Nucleic acid according to the invention may be conveniently contained in a vector.

Sequences complementary to Z locus gene sequences shown in the figures or in accompanying sequence listing are also part of the invention. Complementary nucleic acid may comprise a nucleotide sequence that is complementary to all or part of the Z locus gene sequence, for example to all or part of the coding region. Fragments of complementary sequences can be used for antisense or RNAi inhibition of Z locus gene expression.

Polypeptide gene products of the Z locus are also part of the invention. A polypeptide may comprise an amino acid sequence shown in any of the figures or in the accompanying sequence listing, or an amino acid sequence that is encoded by a nucleotide sequence set out in any of the figures or the sequence listing or described above. It may encode an amino acid sequence that is at least 70 % identical to such an amino acid sequence, e.g. at least 80 %

identical, at least 90 % identical, at least 95 % identical, at least 98 % identical or at least 99 % identical. Example amino acid sequences are shown in SEQ ID NOS: 3, 5, 7, 11, 13, 19-22, 25-26, 28, 30, 32 and 34.

Identification of further Z locus sequences and orthologues

5 Other Z locus alleles, for example other allelic forms of the same species, or alleles of other Poaceae species, can be obtained by screening cDNA or genomic DNA libraries using probes or primers that hybridise to the sequences provided here, probing for or amplifying the further alleles from nucleic acid preparations of the same or other Poaceae species, or screening sequence databases using the sequences provided here.

10 Accordingly, a further aspect of the invention relates to use of a Z locus gene from perennial ryegrass, or a fragment of such a gene, for identifying or cloning a corresponding Z locus gene from another grass family plant.

A method of identifying or cloning a Z locus gene in a plant may employ an oligonucleotide or oligonucleotides comprising or consisting of a sequence or sequences from a
15 conserved region of a Z locus gene.

The method may comprise:

providing a preparation of nucleic acid from the plant;

providing a probe that hybridises to a the Poaceae Z locus, preferably to a conserved
region;

20 contacting the nucleic acid in the preparation with the probe under conditions for hybridisation; and

identifying a Z locus gene if present by its hybridisation with the probe.

Alternatively, a method of identifying a Z locus gene in a plant of the grass family may comprise

providing a preparation of nucleic acid from the plant;

25 providing a pair of primers for hybridisation to the Poaceae Z locus, suitable for amplifying a region of the locus, for example for amplifying one or more conserved regions and/or variable region;

contacting the nucleic acid in the preparation with the primers under conditions for hybridisation and amplification; and

30 determining the presence or absence of an amplification product, where the presence of an amplification product indicates the presence of a Z locus gene in the plant.

The method of identifying a Z locus gene may further comprise determining the sequence of the Z locus gene in the plant.

35 Z locus genes can also be identified using bioinformatic techniques. Many plant sequences are now available in public databases, and further sequences can be obtained by whole genome sequencing or targeted genome sequencing. Another method of identifying a Z

locus gene in a plant is therefore to screen a sequence database, or screen all or part of a plant genome sequence, for a Z locus gene sequence. This may be preceded by a step of sequencing all or part of a plant genome, e.g. after obtaining a preparation of genomic DNA from the plant.

5 Methods of screening for Z locus genes may comprise identifying a Z locus gene as described herein, by shared sequence identity with any of the Z locus gene sequences set out in the drawings or in the accompanying sequence listing. As described in more detail elsewhere herein, the Z locus gene may share a certain percentage identity with one or more of such sequences across its full length, and/or across the VR and/or conserved regions. Methods
10 may further comprise isolating nucleic acid comprising the Z locus gene from the plant.

Primers and probes

A further aspect of the invention is a primer or probe comprising an oligonucleotide sequence that specifically hybridises to a Poaceae Z locus gene. The oligonucleotide sequence may hybridise to a conserved region of Z locus alleles, or to a variable region, or to a junction
15 between a conserved region and a variable region. Such probes and primers may be used in methods of identifying and cloning Z locus genes from Poaceae, or in methods of haplotyping and predicting SI phenotypes, as described.

Genotyping and haplotyping

Further aspects of the invention relate to determining the identity of Z locus alleles in a
20 plant, which is of particular value for predicting compatibility phenotype in a plant, including predicting self-incompatibility or self-compatibility phenotype in a plant and predicting compatibility between a first and second plant.

SI alleles can be identified by means of diagnostic DNA markers at the Z locus. A probe or primer that hybridises to the Poaceae Z locus can be used to determine the identity of alleles
25 present at the Z locus of a plant of the grass family, for example to determine the Z locus haplotype of the plant. The probe or primer may comprise an oligonucleotide sequence that hybridises to the conserved or VR of a Z locus allele.

One aspect of the invention is a method of determining the identity of Z locus alleles in a plant of the grass family, comprising

30 providing a preparation of nucleic acid from the plant;
 providing a probe or primer that hybridises to the Z locus;
 annealing the probe or primer to the Z locus in the nucleic acid from the plant; and
 identifying the Z locus allele present at the Z locus by detecting hybridisation and/or by obtaining an amplification product.

35 The method may comprise

providing a probe that hybridises to a variable region of a Z locus allele at the Poaceae Z locus;

contacting the nucleic acid in the preparation with the probe under conditions for hybridisation; and

identifying the allele if present by its hybridisation with the probe.

Alternatively the method may comprise

providing a pair of primers for hybridisation to the Poaceae Z locus, e.g. to specifically amplify nucleic acid containing the VR of a Z locus allele;

contacting the nucleic acid in the preparation with the primers under conditions for hybridisation and amplification; and

determining the presence or absence of an amplification product, where the presence of an amplification product indicates the presence of the Z locus allele.

A method of the invention may further comprise determining the haplotype of the plant by identifying the combination of alleles paired at the Z locus in the plant.

We describe an example of a marker assay based on high resolution melting of polymerase chain reaction (PCR)-amplified DNA fragments that can be used to visualise and predict haplotype combinations for the Z SI locus.

Prediction of compatibility or SI phenotype

Genotype and haplotype information provided by methods described above allows the compatibility phenotype of a plant to be predicted, for example predicting whether a plant is self-incompatible or self-compatible. By knowing which Z locus alleles are present in the plant, pollen haplotypes can be predicted, and compatibility of pollen haplotype with stigma genotype can be determined by comparing whether particular combinations of Z alleles are incompatible or compatible, i.e. whether the Z alleles are interacting or non-interacting.

One further aspect of the invention is a method of predicting the ability of a plant of the grass family to self-fertilise, comprising

determining the identity of Z locus genes in the plant;

and

determining whether an incompatible combination of Z locus genes is present.

Where one or more Z locus genes are absent or mutated to be non-functional, this may for example result in there being no incompatible combination of Z locus genes, in which case a compatibility phenotype is predicted.

Another aspect of the invention is a method of predicting the ability of a first plant to fertilise a second plant, comprising

determining the identity of Z locus genes in the first plant and in the second plant;

determining possible combinations of pollen Z locus genes in the first plant and stigma Z locus genes in the second plant; and

identifying whether each possible combination of pollen and stigma Z locus genes includes an incompatible combination of Z locus genes.

5 An incompatible combination of Z locus genes leads to incompatibility, whereas if no incompatible combination of Z locus genes is present then this indicates that the first plant may be able to fertilise the second plant.

An incompatible combination of Z locus genes may be a combination of Z locus genes with matching VRs. Where the Z locus genes in pollen have the same VRs as the Z locus
10 genes in the stigma, the combination is incompatible. Z locus genes with matching VRs interact, leading to incompatibility. Methods may therefore comprise determining whether pollen produced by a first plant would contain a Z locus allele having a VR that matches a Z locus allele VR in a second plant. Matching VRs indicates an incompatible combination, indicating that the pollen of the first plant is unable to fertilise the second plant.

15 Combination of genes including similar VRs initiate a SI reaction. For example, a diploid plant (referred to as plant 1 in Figure 8) is producing haploid pollen consisting of haplotype 1 or haplotype 2. As both haplotypes contain the same amino acid sequence motif (VDGNGQ, SEQ ID NO: 35) at the VR of LpGK1 and LpGK2, and the same amino acid sequence motif (VDGNGQ) is also present in the diploid pistil tissue of a second plant (thereafter referred to as
20 plant 2), pollen from plant 1 is 100% incompatible on plant 2. Similarly, haploid pollen from plant 2 can either consist of haplotype 3 or haplotype 4, both containing different amino acid sequence motifs (VDGNGQ and IGGK...LLGQ, respectively). As the amino acid sequence motif (VDGNGQ) of haplotype 3 is present in the diploid pistil tissue of plant 1, pollen containing haplotype 3 will be incompatible on plant 1. However, pollen from plant 2 that contain the amino
25 acid sequence motif of haplotype 4 (IGGK...LLGQ, SEQ ID NOS: 36 and 37) will be able to pollinate plant 1. Thus, pollen from plant 2 is 50% incompatible on plant 1.

Altering SI phenotype

The Z locus, including Z locus genes and transcriptional control elements, can be targeted or modified to alter the compatibility phenotype of a plant, for example to control self-
30 compatibility, including inhibiting SI. Use of Z locus nucleic acid to control the compatibility phenotype of a plant of the grass family is a further aspect of the invention.

The invention provides a method of controlling self-fertility in a plant of the grass family. For example, the invention enables the SI phenotype of allogamous Poaceae to be altered by various means, such as:

- 35 - modulating expression of a Z locus allele in the plant;
- modifying the genotype of the plant at the Z locus and/or

- altering activity of a Z locus gene product.

One or more Z locus genes may be targeted in this way. The Z locus gene may be LpGK1 and/or LpGK2.

A plant can be made self-compatible by avoiding expression of an interacting pair of Z locus alleles in the plant, for example by mutating an LpGK1 and/or LpGK2 allele in the plant so that the plant does not produce interacting gene products, or by inhibiting expression of a gene product from the Z locus, e.g. inhibiting expression of LpGK1 or LpGK2. It may be necessary to inhibit multiple Z alleles in the plant, to ensure that the plant does not express any pair of interacting Z alleles. Alternatively, a plant that expresses an interacting pair of Z alleles may be made self-compatible by inhibiting interaction between the expressed polypeptides, for example by inhibiting mutual binding of the polypeptides or by inhibiting glycerol kinase activity. A self-compatible plant may be made self-incompatible by providing an interacting pair of Z alleles in the plant and allowing expression of the alleles, allowing the gene products of the alleles to interact. Conveniently, gene expression at the Z locus can be controlled in order to modulate self-fertility, for example by providing a Z allele under control of an inducible promoter. Equally, these techniques may be applied to render a plant compatible with a second plant, by avoiding expression of a Z locus allele in the plant that would otherwise interact with a Z locus allele in the second plant.

A method of increasing self-compatibility in a self-incompatible plant of the grass family may comprise

down-regulating expression of a Z locus allele in the plant, for example using antisense or RNAi or using an inducible promoter under conditions in which the promoter is inactive; and/or

mutating nucleic acid at the Z locus to inhibit expression of the Z allele, for example by knocking-out the gene by frameshift mutation or deletion, or mutation of transcriptional control elements, to inhibit production of a functional polypeptide product. One or both of LpGK1 and/or LpGK2 may be down-regulated and/or mutated. An example frameshift mutation is shown for the LpGK1 allele haplotype 226 in Figures 10 – 12. A frameshift mutation introduced in the VR of both LpGK1 alleles produces a self compatible phenotype in the plant. A method of the invention may comprise introducing a frameshift mutation in or upstream of the VR in one or more chromosomal copies of LpGK1 in a plant, e.g. in both alleles of a diploid plant. A method may comprise deleting all or part of the VR and/or conserved regions of a Z locus gene, e.g. LpGK1. Again, the mutation may be introduced in one or all chromosomal copies in a plant. Mutation may render the gene non-functional. Mutation may render the Z locus gene unable to interact with another Z locus gene in the plant.

The method may comprise a targeted transgenic knockout of a Z locus allele, e.g. LpGK1 and/or LpGK2.

Such methods avoid interaction between Z locus alleles in the plant, thereby allowing the plant to self-fertilise.

A method of reducing self-compatibility in a self-compatible plant of the grass family may comprise

5 up-regulating expression of a Z locus allele in the plant, for example by inducing expression at the Z locus from an inducible promoter;

mutating the sequence of a Z locus allele in a plant; and/or

introducing a Z locus allele into the plant.

Such methods allow interaction between a pair of Z locus alleles in the plant, inhibiting
10 self-fertilisation in the plant.

Transgenic plants may be produced by introducing nucleic acid according to the invention into a plant so that the nucleic acid is stably integrated into the plant genome or is contained in a vector that is stably maintained in the plant cells. The introduced nucleic acid may be a Z locus allele of LpGK1 or LpGK2, for example replacing a corresponding
15 endogenous LpGK1 or LpGK2 allele at the Z locus in the plant. The transgenic plant may be engineered to contain an interacting pair of Z alleles, or to mutate or alter expression of one or both of a pair of interacting Z alleles. The introduced nucleic acid may comprise a transcriptional control element for the Z locus, and may replace a corresponding endogenous transcriptional control element at the Z locus in the plant, altering expression of the Z locus
20 allele, which may be LpGK1 and/or LpGK2. For example the nucleic acid may disrupt an endogenous Z locus promoter, or may comprise an inducible promoter or other heterologous transcriptional control element, so that expression of the Z allele is operably linked to the heterologous element. Conveniently, this may place expression of a Z allele under control of an inducible promoter, allowing the self-fertility phenotype to be controlled by suppressing or
25 allowing expression from the inducible promoter as desired.

Plants having an altered compatibility phenotype form part of the invention, and can be generated by methods according to the invention as described. The invention also extends to seeds and parts of such plants, as well as progeny of the plants and seeds and parts of those progeny. A transgenic Poaceae plant may comprise a genome having a Z locus in which
30 expression of a Z locus allele is operably linked to a heterologous genetic element, for example an inducible promoter, and/or it may comprise a Z locus containing an introduced or mutated Z locus allele, which may be LpGK1 and/or LpGK2. The transgenic plant may have an altered compatibility phenotype, for example it may be a self-compatible plant of an allogamous Poaceae species, or it may be a self-incompatible plant of a self-compatible species, or it may
35 be a plant in which self-fertility is controlled by an inducible promoter directing gene expression at the Z locus. A transgenic plant may contain nucleic acid complementary to all or part of a Z locus allele in the plant, under control of expression from a promoter. Expression of the

complementary nucleic acid in the plant may inhibit expression of a polypeptide from the Z locus Z allele, and may result in a self-compatible phenotype by preventing interaction between gene products of a pair of Z alleles.

Plants in which SI is suppressed, e.g. self-compatible plants produced in accordance with the invention, and those that are rendered compatible with other plants in accordance with the invention, may generate increased seed yields. SI has been reported to impact seed yield in perennial ryegrass, where substantial amounts of incompatible pollen in the pollen cloud significantly reduced the yield of seed per plant and per panicle (Studer et al. 2008). The present invention enables production of populations of plants, where the plants are self-compatible and/or mutually compatible, allowing seed yield to be enhanced and potentially achieving full seed yield by avoiding SI-related reductions in seed yield.

Assaying for anti-SI agents

Based on the present characterisation of the Z locus genes, and in particular their expressed polypeptides, it is possible to screen for agents that bind to and/or alter the activity of the Z locus gene products. Such agents are useful for inhibiting SI through physical application to plants, for example by spraying, providing a convenient way of inhibiting SI on a temporary or more long-lasting basis in existing species. It is desirable to identify agents that influence the compatibility phenotype of a plant, particularly those that act on physical contact with the plant or with the relevant part of a plant e.g. pollen or flower parts. Agents that inhibit SI in plants can usefully be employed for plant breeding, including in methods described herein, and also in methods of improving seed yield by treating plants in a population to inhibit SI, allowing self-pollination of the plants, and allowing the plants to set seed. Seed of the plants may then be gathered. Alternatively, agents may promote SI in plants, and such agents find use in breeding programmes including for hybrid production where plants are outcrossed.

Such agents may be identified by screening candidate compounds for the property of binding to a polypeptide product of a Z locus gene, and/or the property of altering the activity of a polypeptide product of a Z locus gene, and/or the property of influencing a compatibility phenotype of a plant. For example, a method may comprise:

bringing a candidate compound into contact with a polypeptide encoded by a Z locus gene, or a fragment of such a polypeptide; and
determining whether the compound binds the polypeptide.

Where binding is detected, this indicates that the compound is a possible agent for influencing the compatibility phenotype, and the agent may then be tested in further assays, e.g. to determine whether the candidate compound affects activity of the polypeptide or whether the candidate compound affects the compatibility phenotype of a plant on contact with the plant.

A candidate compound may be brought into contact with a polypeptide encoded by a Z locus gene, a or fragment of such a polypeptide, under conditions where the polypeptide exhibits one or more activity such as glycerol kinase activity. Detecting an increase or decrease in glycerol kinase activity in the presence of the compound compared with the absence of the compound indicates that the compound is a possible agent for influencing the compatibility phenotype of a plant.

Further screening of such compounds may involve spraying the compound on to plants and determining whether the compound affects the compatibility phenotype of the plants, e.g. determining whether a compound that binds and/or inhibits activity of a polypeptide encoded by a Z locus gene is an inhibitor of SI in the plants. Inhibition of SI in self-incompatible plants may be detected by setting of seed, or increased setting of seed by the plants, following treatment with the compound compared with plants not treated with the compound.

Kinase inhibitors, for example inhibitors of glycerol kinase, may also be designed by *in silico* methods, using computer-generated models to identify compounds having inhibitory activity. Compounds that bind to LpGK1 and/or LpGK2 may be provided. These may bind any part of the polypeptide encoded by the Z locus, for example the variable region and/or the upstream and/or downstream conserved region. The compound may bind an active portion of the polypeptide. All such inhibitors and compounds represent candidate agents for screening according to the present invention. Biological agents with the same properties may also be screened and/or used for influencing compatibility phenotype as described. For example, antibodies to a polypeptide encoded by a Z locus gene can be used for binding the polypeptide, either in vitro or in planta, including in methods of inhibiting SI as described above. Other possible candidate compounds and agents include salts, ions and protease inhibitors. Such compounds, and others from compound libraries, may be screened as described.

Use of such agents, antibodies, and compounds such as kinase inhibitors, for inhibiting SI in allogamous Poaceae is newly possible owing to the present characterisation of the polypeptides encoded by the Z locus and their role in SI, and accordingly this use represents a further aspect of the invention.

Inbreeding and hybrid production

Further aspects of the invention relate to breeding of plants by controlling fertility through the Z locus and utilising SI to control pollination in hybrid breeding schemes. By suppressing SI or rendering a plant self-compatible, Poaceae plants can be self-fertilised, allowing generation of inbred lines, which can be selected for their combining ability for desired traits, such as yield (e.g. seed yield) or biomass. Further, by inducing or introducing SI into self-compatible plants, including such inbred lines, hybrids may be generated and selected for desired traits. Heterosis associated with hybrid breeding schemes may further improve traits such as yield or biomass.

Also, as described above, SI in a plant may be inhibited in order to enhance seed yield, by allowing self-pollination of the plant, allowing the plant to set seed, and optionally gathering the seed.

One aspect of the invention is a method of plant breeding comprising self-fertilising a self-incompatible plant of the grass family, comprising increasing self-compatibility in the plant using a method of the invention as described, allowing or promoting self-pollination of the plant, and obtaining seed.

The method may further comprise generating progeny by growing the seed, and optionally performing further rounds of crossing and selection to generate inbred lines. Progeny may be selected for desirable traits such as enhanced yield or biomass. Plants may be transgenic plants, and may be transgenic at the Z locus and/or may contain transgenes or heterologous genetic elements unlinked to the Z locus, such as genes related to other traits, such as yield (e.g. seed yield) or biomass.

The method may further comprise restoring self-incompatibility in the plant, or inbred lines generated from the plant, by reducing self-compatibility in the plant using a method of the invention as described, and out-crossing the inbred line with a second plant to generate hybrid seed. Out-crossing between a first and second comprises allowing pollination of one plant by the other so that one or both plants are fertilised and produce seed. The second plant may be a genetically distinct plant of the same species, e.g. a plant of a second inbred line. The method may comprise allowing or promoting pollination of the first plant by the second plant, or allowing or promoting pollination of the second plant by the first plant, or allowing or promoting cross-pollination, and obtaining seed.

The method may further comprise generating progeny by growing the seed and selecting hybrids for desired traits such as enhanced yield or biomass, optionally including further rounds of crossing and selection.

Methods of altering incompatibility phenotypes of plants, as described herein, can be applied to maximise the productivity of polycross breeding methods. The polycross breeding involves selecting a pool of plants and allowing cross-pollination between plants in the pool, allowing the plants to set seed and gathering the seed. Progeny generated from the seed can be selected for desired traits, such as increased biomass or seed yield. Polycross methods may therefore generate new varieties with improved traits. The number of different crosses made between the plants, and the genetic diversity of the population of progeny, is dependent on the compatibility of plants in the pool – in other words it depends on the extent to which each plant in the pool can fertilise and be fertilised by the other plants in the pool. Limited compatibility between plants hampers polycross methods by reducing the number of fruitful crosses that can be made the genetic diversity of the progeny.

The present invention allows compatibility between plants to be predicted, by determining the identity of Z locus alleles in the plants. By predicting the compatibility phenotype of a number of candidate plants, the breeder can select a pool of compatible plants, avoiding or reducing combinations of plants that are incompatible. The described methods of genotyping and haplotyping can be applied to predict the compatibility phenotype of the plants. By knowing which Z locus alleles are present in the plant, pollen haplotypes can be predicted, and compatibility of pollen haplotype with stigma genotype can be determined by comparing whether particular combinations of Z alleles in the plants are compatible or incompatible, i.e. whether the Z alleles are interacting or non-interacting.

A method may comprise predicting the ability of a first plant of the grass family to fertilise a second plant of the grass family, comprising

determining the identity of Z locus genes in the plant;

determining whether a compatible combination of Z locus genes is present between the two plants; and

if a compatible combination of Z locus genes is present, including the plants in a polycross, and if a compatible combination of Z locus genes is not present, excluding the plants from a polycross.

The method may comprise identifying a pool of plants that are mutually compatible, i.e. able to fertilise one another, by determining the identity of Z locus genes in the plants and identifying compatible combinations of Z locus genes. Methods of determining the identity of Z locus genes and predicting compatibility phenotypes are described in detail herein. The method may further comprise allowing cross-pollination between plants in the pool, allowing the plants to set seed and gathering the seed. The method may further comprise generating progeny from the seed, and selecting progeny exhibiting improved traits, e.g. enhanced biomass or enhanced seed yield.

As noted above, it is recognised that incompatibility reduces seed yield in Poaceae. Using methods of the invention to predict incompatible combinations of plants, polycrosses can be performed using only the most fertile combinations, reducing time and effort spent on less fruitful polycrosses. Accordingly, when a pool of mutually compatible plants has been identified, a polycross may be performed to generate seed. Increased yield of seed can be obtained due to exclusion of crosses that are incompatible or less compatible.

Brief Description of the Drawings

The invention will now be described in more detail, with reference to the drawings, in which:

- Figure 1 Alignment of cDNA sequences of LpGK1 haplotypes P205 (SEQ ID NO: 8) and B724 (SEQ ID NO: 9) of perennial ryegrass. The alignment shows conserved regions at nucleotides 38 to 95 and 132 to 180 surrounding a variable sequence motif variable region (VR) at nucleotides 96 to 131 (referred to allele P205).
- Figure 2 Alignment of amino acid sequences of LpGK1 haplogypes P205 (SEQ ID NO: 3) and B724 (SEQ ID NO: 7) perennial ryegrass. The alignment shows conserved regions at amino acid residues 1 to 35 and 45 to 378 surrounding a variable sequence motif or hypervariable region (VR) at amino acid residues 36 to 44 (referred to allele P205).
- Figure 3 Alignment of cDNA sequences of LpGK2 haplotypes P226 (SEQ ID NO: 10) and P205 (SEQ ID NO: 12) of perennial ryegrass. The alignment shows conserved regions at nucleotides 1 to 554 and 591 to 1598 (referred to allele P205) surrounding a variable sequence motif or variable region (VR) at nucleotides 555 to 590 (referred to allele P205).
- Figure 4 Alignment of amino acid sequences of LpGK2 haplotypes P226 (SEQ ID NO: 11) and P205 (SEQ ID NO: 13) of perennial ryegrass. The alignment shows conserved regions at amino acid residues 1 to 185 and 199 to 531 (referred to allele P205) surrounding a variable sequence motif or variable region (VR) at amino acid residues 186 to 198.
- Figure 5 Alignment of amino acid sequences of LpGK1 (SEQ ID NO: 3) and LpGK2 (SEQ ID NO: 13) in perennial ryegrass haplotype P205
- Figure 6 Alignment of cDNA sequences of LpGK2 haplotypes P226 SEQ ID NO: 10 and P205 SEQ ID NO: 12 of perennial ryegrass with Z locus gene sequences from barley (gi|151421547) SEQ ID NO: 14, Brachypodium (gi|357166355) SEQ ID NO: 15, rice (gi|115460941) SEQ ID NO: 16, sorghum (gi|242077401) SEQ ID NO: 17 and maize (gi|195645871) SEQ ID NO: 18.
- Figure 7 Alignment of amino acid sequences of LpGK2 haplotypes P226 (SEQ ID NO: 11) and P205 (SEQ ID NO: 13) of perennial ryegrass with Z locus gene product sequences from Brachypodium (gi|357166356) SEQ ID NO: 19, rice (gi|38352013) SEQ ID NO: 20, maize (gi|239049939) SEQ ID NO: 21 and sorghum (gi|242077402) SEQ ID NO: 22.

Figure 8 Illustration of allele composition at LpGK1 and LpGK2 and the corresponding SI phenotype of two plants (plant 1 and 2) that are crossed. Plant 1 is producing haploid pollen consisting of haplotype 1 or haplotype 2. As both haplotypes contain the same amino acid sequence motif (VDGNGQ) (SEQ ID NO: 35) at the VR of LpGK1 and LpGK2, and the same amino acid sequence motif (VDGNGQ) is also present in the diploid pistil tissue of a second plant, pollen from plant 1 is 100% incompatible on plant 2. Similarly, haploid pollen from plant 2 can either consist of haplotype 3 or haplotype 4, both containing different amino acid sequence motifs (VDGNGQ and IGGK...LLGQ, respectively). As the amino acid sequence motif (VDGNGQ) of haplotype 3 is present in the diploid pistil tissue of plant 1, pollen containing haplotype 3 will be incompatible on plant 1. However, pollen from plant 2 that contain the amino acid sequence motif of haplotype 4 (IGGK...LLGQ) (SEQ ID NO: 36, SEQ ID NO: 37) will be able to pollinate plant 1. Thus, pollen from plant 2 is 50% incompatible on plant 1.

Figure 9 BLASTP analysis of the amino acid sequence of LpGK2 against the non-redundant protein database of GenBank. LpGK1 and LpGK2 proteins belong to the FGGY family of carbohydrate kinases and contain several active sites. The top bar represents the query protein sequence with a total length of 531 amino acids. The bars below indicated and locate active sites such as metal, MgATP and carbohydrate binding sites.

Figure 10 Multiple sequence alignment of the coding DNA sequence (CDS) of LpGK1 haplotypes P205 (SEQ ID NO: 23) and P226 (SEQ ID NO: 24). The haplotype P205 is derived from a perennial ryegrass genotype F1_30 that was used to develop the fine-mapping population segregating for the Z SI locus. The haplotype P226 is derived from a self-compatible perennial ryegrass genotype. The conserved and variable regions (as illustrated in Figure 6) are given in bold and underlined/italic, respectively. The LpGK1 allele of haplotype P226 that is derived from a self-compatible genotype shows a 2 bp (CT) insertion at position 115 and 116 of the corresponding allele.

Figure 11 Multiple sequence alignment of the amino acid sequences (AS) of LpGK1 haplotypes P205 (SEQ ID NO: 3) and P226 (SEQ ID NO: 25). The haplotype P205 is derived from a perennial ryegrass genotype F1_30 that was used to develop the fine-mapping population segregating for the Z SI locus. The

haplotype P226 is derived from a self-compatible perennial ryegrass genotype. The conserved and variable regions (as illustrated in Figure 7) are given in bold and underlined/italic, respectively. The LpGK1 allele of haplotype P226 derived from the self-compatible genotype shows a frame shift in the amino acid sequence when compared to the LpGK1 allele of haplotype P205. This frame shift is caused by a 2 bp (CT) insertion in the variable region of LpGK1 (Figure 10) and is finally leading to termination of the protein translation process. The non-functional amino acid sequence which is translated downstream of the frameshift mutation is shown in strikethrough.

Figure 12 Multiple sequence alignment of the amino acid sequences (AS) of LpGK1 (SEQ ID NO: 26) and LpGK2 (SEQ ID NO: 11) in the haplotype P226 that is derived from a self-compatible perennial ryegrass genotype. The four amino acids highlighted in bold indicate that LpGK1 and LpGK2 of the haplotype P226 contain different amino acid sequence motifs in the variable region.

Figure 13 (a) Multiple sequence alignment of cDNA sequences of LpGK2 haplotypes alleles S089 (SEQ ID NO: 27), S027 (SEQ ID NO: 31), S021 (SEQ ID NO: 33) and S065 (SEQ ID NO: 29) from *Lolium perenne*. The conserved regions are shown in bold and the variable regions are underlined. (b) Multiple sequence alignment of amino acid sequences of LpGK2 haplotypes S089 (SEQ ID NO: 28), S027 (SEQ ID NO: 32), S021 (SEQ ID NO: 34) and S065 (SEQ ID NO: 30). The conserved regions are shown in bold and the variable regions are underlined.

Detailed Description

Self-incompatibility (SI) and compatibility

In gametophytic SI (GSI), the SI phenotype of the pollen is determined by its own genotype. For example, where SI is controlled by a single multi-allelic locus, the S locus, the pollen is rejected when the S haplotype of the haploid pollen matches either of the two S haplotypes of the diploid pistil. In some GSI systems, such as in perennial ryegrass and other allogamous Poaceae, SI is controlled by two independent multi-allelic loci, S and Z. SI occurs when both pollen S and Z alleles are matched in the pistil. Otherwise pollen is compatible.

Various aspects of the invention relate to plants in which SI is determined at least in part by a Z locus, and to products and methods for reducing or controlling SI in such plants. Accordingly, a plant in the present invention may be one in which GSI is determined by at least a Z locus, for example by at least two multiple-allelic loci, S and Z.

The compatibility phenotype of a plant refers to whether or not the plant is self-compatible or self-incompatible, i.e. whether or not self-pollination results in fertilisation and seed production. The compatibility phenotype may also refer to a plant's ability to fertilise or be fertilised by a second plant. The first and second plant will usually be grass family plants of the same species. The plants may be diploid.

The present invention contributes significantly to understanding this phenomenon at the genetic and molecular level, and leads to the possibility of genetic or molecular intervention to manipulate or control SI, including generating self-compatibility in self-incompatible plants.

We describe a number of alleles and haplotypes at the Z locus, containing molecular markers that determine their compatibility or incompatibility. A pair of Z alleles may be interacting alleles or non-interacting alleles. Interaction between alleles may take place on the DNA or RNA level or at the polypeptide level. Thus, for example, a functional Z allele may express a polypeptide that is capable of interacting with a polypeptide encoded by a corresponding Z allele, where interaction between the polypeptides inhibits fertilisation of the pistil by the pollen. In allogamous Poaceae, the Z locus of a diploid plant generally comprises a homologue pair of interacting Z alleles, so that the plant is self-incompatible. Pollen produced by such a plant comprises the Z allele, and the polypeptide expressed from the pollen Z allele interacts with a Z allele expressed in the plant pistil, initiating SI.

Various aspects of the invention relate to prediction of compatibility phenotype in a plant (self-compatibility or self-incompatibility), or prediction of compatibility between two or more plants. In general, such methods involve predicting the ability of pollen to fertilise a stigma by determining whether the pollen and stigma Z locus alleles contain an incompatible combination, i.e. a pair of interacting Z locus alleles. Such methods involve determining whether the pollen contains a Z locus gene that is incompatible with a Z locus gene in the stigma. If an incompatible combination is present, then the pollen is unable to fertilise the stigma. If the pollen and stigma do not contain an incompatible combination of Z locus genes, the pollen is able to fertilise the stigma. Of course, other factors may also influence fertility, such as other genetic or external influences on plant flowering. Prediction of compatibility based on the identity of Z locus genes assumes that the Z locus genes are functional, i.e. able to interact with corresponding Z locus genes. Functionality may require expression of the full length Z locus gene product and/or expression of the Z locus gene product including both conserved regions. Z locus genes in an incompatible combination would therefore generally be capable of being expressed from the Z locus. Even where the methods of predicting phenotype are not always 100 % accurate, they are generally reliable and provide useful predictions in a majority of situations, allowing breeders to select plants for growing on and/or crossing on the basis of the predicted phenotypes, reducing inefficiencies such as inclusion of incompatible plants in polycrosses as noted elsewhere herein.

A pair of functional Z locus genes are incompatible, or interacting, if they have matching VR sequences. Matching VRs are VRs of identical amino acid sequence, or amino acid sequences with very minor variation, for example differing by only one or two point mutations. A point mutation may be a substitution (e.g. conservative substitution), insertion or deletion of an amino acid residue. Optionally, an incompatible combination of Z locus genes may share at least 90 %, 95 %, 98 % or 99 % sequence identity across their full length, or across the conserved and variable regions combined. An incompatible combination of Z locus genes may be identical in sequence in the conserved and variable regions, and optionally may be fully identical Z locus genes.

Accordingly, pollen containing a functional Z locus gene is unable to fertilise a stigma containing a functional Z locus gene with a matching VR. Where pollen and stigma are of the same plant, the plant is self-incompatible. Accordingly, interaction between the Z locus gene in the pollen of a plant and the identical Z locus gene in the stigma of the same plant leads to SI.

A method of predicting compatibility between a first plant and a second plant may comprise

- predicting Z locus haplotypes of pollen produced by the first plant;
- determining, for each predicted Z locus haplotype of the pollen, whether the pollen contains a Z locus gene that is incompatible with a Z locus gene in the genotype of the second plant.

A method of predicting the ability of a first plant to fertilise a second plant may comprise determining the identity of Z locus genes in the first plant and in the second plant; determining possible combinations of pollen Z locus genes in the first plant with stigma Z locus genes in the second plant; and

- identifying whether each possible combination of pollen and stigma Z locus genes includes an incompatible combination of Z locus genes, wherein an incompatible combination of Z locus genes occurs when the variable region (VR) of a Z locus gene in the pollen matches the VR of a Z locus gene in the stigma; wherein

- presence of an incompatible combination of Z locus genes indicates that the pollen is unable to fertilise the stigma, and absence of an incompatible combination of Z locus genes indicates that the pollen is able to fertilise the stigma.

An advantage of the invention is that compatibility phenotype can be predicted without requiring the plant or plants to flower. Phenotype may be predicted in plants that are not yet at flowering age, e.g. in seedlings. It is possible to determine whether the pollen of a first plant contains a Z locus gene that is incompatible with a Z locus gene in the stigma of a second plant, without requiring either plant to flower. Combinations of Z locus genes in pollen or stigma can be determined by determining the identity of Z locus genes present in a plant, using the genotyping or haplotyping methods described elsewhere herein.

Cells of the stigma typically contain all chromosomal copies of the Z locus, so the Z locus genes present in the stigma are determined by determining the Z locus genes in the genotype of the second plant. Where pollen contains a single chromosomal copy of the Z locus, determining the identity of Z locus genes in pollen of a plant involves determining the identity of Z locus genes in each chromosomal copy in the plant. Therefore, a Z locus gene in a predicted pollen haplotype may be a Z locus gene in one chromosomal copy of the first plant, and the Z locus genes in the stigma are the Z locus genes in all chromosomal copies of the second plant. Accordingly, there is an incompatible combination when the VR of a chromosomal copy of a Z locus gene in the first plant (representing a pollen Z locus gene) matches a VR in at least one chromosomal copy in the second plant (representing a stigma Z locus gene).

A homozygous diploid plant contains two identical chromosomal copies of the Z locus, so there is a single predicted Z locus haplotype for the pollen. A heterozygous diploid plant contains two different chromosomal copies of the Z locus, so there are two predicted Z locus haplotypes for the pollen. This may lead to different predictions of compatibility for different pollen haplotypes. The Z locus gene of one haplotype may interact with a Z locus gene of the stigma of a second plant, while the Z locus gene of another haplotype of pollen from the same plant does not interact with any Z locus gene of the stigma of the second plant. A first and second diploid plant may therefore be 0 % compatible, 50 % compatible or 100% compatible, depending on the haplotype of the pollen and the diploid genotype of the stigma.

A fully incompatible phenotype is predicted where all combinations of pollen and stigma Z locus genes include an incompatible combination, indicating that the first plant is unable to fertilise the second plant.

A partially incompatible phenotype is predicted where one combination of pollen and stigma Z locus genes includes an incompatible combination, and one combination of pollen and stigma Z locus genes does not include an incompatible combination, indicating that only a proportion of the pollen of the first plant is able to fertilise the second plant. For a diploid plant where only one pollen haplotype is matched in the stigma, a 50 % compatibility phenotype is predicted since half the pollen will fertilise the stigma while the other half is incompatible.

A fully compatible phenotype is predicted where no combination of pollen and stigma Z locus genes includes an incompatible combination, indicating that pollen from the first plant is able to fertilise the second plant.

The Z locus alleles considered in such methods of predicting compatibility phenotype may be LpGK1 or LpGK2, or both LpGK1 and LpGK2.

For example, the method may comprise determining the identity of LpGK1 and LpGK2 in the first plant and the second plant, where an incompatible combination occurs when a Z locus in the first plant comprises LpGK1 and LpGK2 with matching VRs, which also match the VRs of LpGK1 and LpGK2 in at least one Z locus in the second plant. The second plant may be

homozygous or heterozygous at the Z locus. Ability of the first plant to fertilise a second plant may be predicted if no incompatible combination is present, for example if the Z locus LpGK1 and LpGK2 in pollen of the first plant do not have matching VRs, and if the Z locus LpGK1 and LpGK2 in the stigma of the second plant do not have matching VRs.

5 A method of predicting the ability of a first plant to fertilise a second plant may comprise determining the identity of Z locus genes in the first plant and in the second plant; determining possible combinations of pollen Z locus genes in the first plant with stigma Z locus genes in the second plant; and

identifying whether each possible combination of pollen and stigma Z locus genes
10 includes an incompatible combination of Z locus genes, wherein an incompatible combination of Z locus genes occurs when the VRs of LpGK1 and LpGK2 Z locus genes in one chromosomal copy of the first plant are matching VRs and also match the VRs of LpGK1 and LpGK2 in a chromosomal copy of the Z locus in the second plant. A Z locus in one chromosomal copy of the first plant represents a possible pollen haplotype, which may be combined with the LpGK1
15 and LpGK2 in the Z locus of both chromosomal copies of a second diploid plant. As noted above, in a plant homozygous at the Z locus, both combinations of pollen and stigma Z locus will be identical, so in effect this can be considered a single combination of pollen and stigma Z locus genes.

An incompatible combination occurs if the Z locus pollen haplotype is found in the stigma.
20 If the Z locus in one chromosomal copy in the first plant has the same haplotype as the Z locus of at least one chromosomal copy in the second plant, an incompatible combination is present.

Presence of an incompatible combination of Z locus genes indicates that the pollen is unable to fertilise the stigma, and absence of an incompatible combination of Z locus genes indicates that the pollen is able to fertilise the stigma. Absence of an incompatible combination
25 may occur where the VRs of the Z locus genes being compared are present but do not match, or where the absence of a matching VR is because one or more Z locus genes is deleted or disrupted so that it is non-functional, for example lacking one or both conserved regions and/or are not capable of being expressed from the Z locus.

In another example, the Z locus gene considered for the prediction method may be
30 LpGK1. A stigma is unable to be fertilised by pollen that contains an LpGK1 gene that is incompatible with an LpGK1 gene in the stigma, whereas the stigma can be fertilised by pollen that does not contain an incompatible LpGK1 gene. A method of predicting ability of a first plant to fertilise a second plant may comprise determining the identity of LpGK1 Z locus genes in the first plant and in the second plant, determining possible combinations of pollen LpGK1 Z locus
35 genes in the first plant with stigma LpGK1 genes of the second plant, and identifying whether each possible combination of pollen and stigma Z locus LpGK1 genes includes an incompatible combination of Z locus LpGK1 genes. As noted already, an incompatible combination of Z

locus genes occurs when the variable region (VR) of a Z locus gene in the pollen matches the VR of a Z locus gene in the stigma. Accordingly, there is an incompatible combination when the VR of each chromosomal copy of LpGK1 in the first plant matches an LpGK1 VR in at least one chromosomal copy in the second plant. The presence of an incompatible combination of Z locus LpGK1 genes indicates that the pollen is unable to fertilise the stigma, and absence of an incompatible combination of Z locus LpGK1 genes indicates that the pollen is able to fertilise the stigma.

Full incompatibility is predicted when each LpGK1 VR in the first plant matches at least one LpGK1 VR in the second plant. In this case, all LpGK1 combinations are incompatible and pollen from the first plant is unable to fertilise the second plant.

A partially incompatible phenotype is predicted where one LpGK1 VR in the first plant matches at least one LpGK1 VR in the second plant, and one LpGK1 VR in the first plant does not match an LpGK1 VR in the second plant. In this case, one combination of pollen and stigma Z locus genes includes an incompatible combination, and one combination of pollen and stigma Z locus genes does not, indicating that only a proportion of the pollen of the first plant is able to fertilise the second plant. For a diploid plant where only one pollen haplotype is matched in the stigma, a 50 % compatibility phenotype is predicted since half the pollen will fertilise the stigma while the other half is incompatible.

A fully compatible phenotype is predicted where the first plant does not contain an LpGK1 with a VR matching an LpGK1 VR in the second plant. In this case, there is no incompatible combination, so pollen from the first plant should be able to fertilise the second plant.

We observed that LpGK1 and LpGK2 at the Z locus of a chromosome often have matching VRs. An incompatible combination may occur when the VRs of pollen LpGK1 and pollen LpGK2 in a first plant match the VRs of an LpGK1 and LpGK2 in the stigma of the second plant.

Methods of prediction may be followed by further steps of carrying out a cross between the first and second plant, and confirming whether the first plant is able to fertilise the second plant. The results of such crosses may be used to strengthen and refine further prediction methods.

Where a method predicts that a first plant is able to fertilise a second plant, the method may further comprise pollinating, or allowing pollination of, the second plant by the first plant, and allowing fertilisation and setting of seed by the second plant.

Where a method predicts that a first plant is unable to fertilise a second plant, the method may comprise excluding the first and/or second plant from a breeding program. The method may comprise discarding the first and/or second plant, and/or not using the first and/or

second plant in a polycross. Methods of crossing and selection, including polycrosses, are described in further detail elsewhere herein.

Grass family plants

Aspects of the invention have been developed and principally illustrated for perennial ryegrass. Perennial ryegrass is one of the most economically and environmentally important grass species and accounts for 70% of all agricultural land use in the United Kingdom (Klaas et al., 2011). Ryegrass is employed in pasture, in lawns and for controlling erosion. Since the GSI system including the Z locus is believed to be shared among grass family plants, the invention is also relevant to other plants of the grass family, including many cereal plants, forage crops and biomass crops. Accordingly, unless stated otherwise, references in this specification to a plant are generally to a plant of the grass family, Poaceae. The plant may be an allogamous Poaceae species, for example perennial ryegrass or rye.

The grass family, Poaceae, includes cereal crops and many cultivated forage crops. Known genera of the Poaceae include: Achnatherum, Aciachne, Acidosa, Acostia, Acrachne, Acritochaete, Acroceras, Actinocladum, Aegilops, × Aegilotriticum, Aegopogon, Aeluropus, Afrotrichloris, Ageratum, Agnesia, × Agrocalamagrostis, × Agroelymus, × Agrohordeum, × Agropogon, Agropyron, Agropyropsis, Agropyrum, × Agrositanion, Agrostis, × Agrotrigia, Aira, Airopsis, Alexfloydia, Alloechaete, Allosepis, Alloteropsis, Alopecurus, Altoparadisium, Alvimia, × Ammocalamagrostis, Ammochloa, Ammophila, Ampelocalamus, Ampelodesmos, Amphibromus, Amphicarpum, Amphigenes, Amphipogon, Anadelphia, Ancistrachne, Ancistragrostis, Andropogon, Andropterum, Aniselytron, Anisopogon, Anomochloa, Anthaenantia, Anthenantia, Anthephora, Anthochloa, Anthoxanthum, Antinoria, Apera, Aphanelytrum, Apluda, Apochiton, Apoclada, Apocopis, Arberella, Arctagrostis, × Arctodupontia, Arctophila, Arctopoa, Aristida, Arrhenatherum, Arthragrostis, Arthraxon, Arthropogon, Arthrostylidium, Arundinaria, Arundinella, Arundo, Arundoclaytonia, Asthenochloa, Astrebla, Athroostachys, Atractantha, Aulonemia, Austrochloris, Austrofestuca, Avena, Avenula, Axonopus, Bambusa, Baptochachis, Beckeropsis, Beckmannia, Bewsia, Bhidea, Blepharidachne, Blepharoneuron, Boissiera, Bonia, Bothriochloa, Bouteloua, Brachiaria, Brachyachne, Brachychloa, Brachyelytrum, Brachypodium, Briza, × Bromofestuca, Bromopsis, Bromuniola, Bromus, Brylkinia, Buchloe, Buchlomimus, Buergersiochloa, Calamagrostis, × Calammophila, Calamovilfa, Calderonella, Calyptochloa, Canastra, Capillipedium, Castellia, Catabrosa, Catalepis, Catapodium, Cathariostachys, Cathestecum, Cenchrus, Centotheca, Centrochloa, Centropodia, Cephalostachyum, Ceratochloa, Chaetium, Chaetobromus, Chaetopoa, Chaetopogon, Chamaeraphis, Chandrasekharania, Chasmanthium, Chasmopodium, Chevalierella, Chikusichloa, Chimonobambusa, Chimonocalamus, Chionachne, Chionochloa, Chloris, Chlorocalymma, Chondrosium, Chrysochloa, Chrysopogon, Chusquea,

- Cinna, Cladoraphis, Clausospicula, Cleistachne, Cleistochloa, Cleistogenes, Clementsiella, Coelachne, Coelachyrum, Coelorachis, Coix, Colantheria, Coleanthus, Colpodium, Cornucopiae, Cortaderia, Corynephorus, Cottea, Craspedorhachis, Crinipes, Crithopsis, Crypsis, Cryptochloa, Ctenium, Cutandia, Cyathopus, Cyclostachya, Cymbopogon, × Cynochloris, Cynodon,
- 5 Cynosurus, Cyperochloa, Cyphochlaena, Cyrtochloa, Cyrtococcum, Dactylis, Dactyloctenium, Dallwatsonia, Danthonia, Danthonidium, Danthoniopsis, × Danthosieglingia, Dasyochloa, Dasypyrum, Davidsea, Decaryella, Decaryochloa, Dendrocalamus, Deschampsia, Desmazeria, Desmostachya, Deyeuxia, Diandrolyra, Diarrhena, Dichaetaria, Dichanthelium, Dichanthium, Dichelachne, Dielsiochloa, Digitaria, Dignathia, Diheteropogon, Dilophotriche, Dimeria, Dinebra,
- 10 Dinochloa, Diplopogon, Dissanthelium, Dissochondrus, Distichlis, Drake-brockmania, Dregeochloa, Drepanostachyum, Dryopoa, × Dupoa, Dupontia, × Dupontopoa, Duthiea, Eccoilopus, Ectocarpus, Echinaria, Echinochloa, Echinolaena, Echinopogon, Ectrosia, Ectrosiopsis, Ehrharta, Ekmanochloa, Eleusine, Elionurus, × Elyhordeum, × Elyleymus, Elymandra, × Elymordeum, × Elymostachys, × Elymotrigia, Elymus, × Elysitanion, Elytrigia,
- 15 Elytrophorus, Elytrostachys, Enneapogon, Enteropogon, Entolasia, Entoplocamia, Eragrostiella, Eragrostis, Eremitis, Eremocaulon, Eremochloa, Eremopoa, Eremopyrum, Eriachne, Erianthecium, Erianthus, Eriochloa, Eriochrysis, Erioneuron, Euclasta, Eulalia, Eulaliopsis, Eustachys, Exotheca, Fargesia, Farrago, Ferrocalamus, Festuca, × Festulolium, × Festulpia, Filgueirasia, Fingerhuthia, Froesiochloa, Gaoligongshania, Garnotia, Gastridium, Gaudinia,
- 20 Gelidocalamus, Germainia, Gerritea, Gigantochloa, Gilgichloa, Glaziophyton, Glyceria, Glyphochloa, Gouinia, Graphephorum, Greslania, Griffithsochloa, Guadua, Guaduella, Gymnopogon, Gynerium, Habrochloa, Hackelochloa, Hainardia, × Hainardiopholis, Hakonechloa, Halopyrum, Harpachne, Harpechloa, Harpochloa, × Haynaldoticum, Helictotrichon, Hemarthria, Hemisorghum, Henrardia, Heterachne, Heteranthelium,
- 25 Heteranthoecia, Heteropholis, Heteropogon, Hickelia, Hierochloa, Hilaria, Himalayacalamus, Hitchcockella, Holcolemma, Holcus, Holttumochloa, Homolepis, Homopholis, Homozeugos, Hordelymus, Hordeum, Hubbardia, Hubbardochoa, Humbertochloa, Hydrothauma, Hygrochloa, Hygroryza, Hylebates, Hymenachne, Hyparrhenia, Hyperthelia, Hypseochoa, Hystrix, Ichnanthus, Imperata, Indocalamus, Indopoa, Indosasa, Isachne, Ischaemum, Iseilema,
- 30 Ixophorus, Jansenella, Jouvea, Kampochloa, Kaokochloa, Kengia, Kengyilia, Kerriochloa, Kinabaluchloa, Koeleria, Lagurus, Lamarckia, Lamprothyrsus, Lasiacis, Lasiurus, Lecomtella, Leersia, Leptagrostis, Leptaspis, Leptocarydion, Leptochloa, Leptocoryphium, Leptothrium, Lepturidium, Lepturopetium, Lepturus, × Leydeum, × Leymostachys, × Leymotrigia, Leymus, Libyella, Limnas, Limnodea, Limnopoa, Lindbergella, Lintonia, Lithachne, Littledalea, Lolium,
- 35 Lolium, Lophacme, Lophatherum, Lopholepis, Lophopogon, Loudetia, Loudetiopsis, Louisiella, Loxodera, Luziola, Lycochloa, Lycurus, Lygeum, Maclurochloa, Maclurolyra, Maltebrunia, Manisuris, Megalachne, Megaloprotachne, Megastachya, Melanocenchris, Melica, Melinis,

Melocalamus, Melocanna, Merostachys, Mesosetum, Metcalfia, Mibora, Micraira, Microbriza,
 Microcalamus, Microchloa, Micropyropsis, Micropyrum, Microstegium, Milium, Miscanthus,
 Mnesithea, Mniochloa, Molinia, Monachather, Monanthochloe, Monelytrum, Monocymbium,
 Monodia, Mosdenia, Muehlenbergia, Muhlenbergia, Munroa, Myriocladus, Myriostachya,
 5 Narduroides, Nardus, Nassella, Nastus, Neesiochloa, Nematopoa, Neobouteloua,
 Neohouzeaua, Neololeba, Neomicrocalamus, Neostapfia, Neostapfiella, Nephelochloa,
 Neurachne, Neurolepis, Neuropoa, Neyraudia, Nipponocalamus, Notochloe, Ocellochloa,
 Ochlandra, Ochthochloa, Odontelytrum, Odyssea, Oligostachyum, Olmeca, Olyra, Ophiochloa,
 Ophiuros, Opizia, Oplismenopsis, Oplismenus, Orcuttia, Oreobambos, Oreochloa, Orinus,
 10 Oropetium, Ortachne, Orthoclada, × Oryticum, Oryza, Oryzidium, Oryzopsis, Otachyrium,
 Otatea, Ottochloa, Oxychloris, Oxyrhachis, Oxytenanthera, Panicum, Pappophorum,
 Parabambusa, Paractaenum, Parafestuca, Parahyparrhenia, Paraneurachne, Parapholis,
 Paratheria, Pariana, Parodiolyra, Paspalidium, Paspalum, Penicillaria, Pennisetum, Pentameris,
 Pentapogon, Pentarrhaphis, Pentaschistis, Pereilema, Periballia, Perotis, Perrierbambus,
 15 Peyritschia, Phacelurus, Phaenanthoecium, Phaenosperma, Phalaris, Pharus, Pheidochloa,
 Phippsia, Phleum, Pholiurus, Phragmites, Phyllorachis, × Phyllosasa, Phyllostachys, Pinga,
 Piptatherum, Piptochaetium, Piptophyllum, Piresia, Plagiantha, Plagiosetum, Pleioblastus,
 Pleuropogon, Plinthanthesis, Poa, Poagrostis, Podophorus, Poecilostachys, Pogonachne,
 Pogonarthria, Pogonatherum, Pogonochloa, Pogononeura, Pohlidium, Poidium, Polevansia,
 20 Polypogon, × Polypogonagrostis, Polytoa, Polytrias, Pommereulla, Porteresia, Potamophila,
 Pringleochloa, Prionanthium, Prosphytochloa, Psammagrostis, Psammochloa, Psathyrostachys,
 Pseudanthistiria, Pseudechinolaena, Pseudodanthonia, Pseudodichanthium, Pseudopentameris,
 Pseudoraphis, Pseudoroegneria, Pseudosasa, Pseudosclerochloa, Pseudosorghum,
 Pseudostachyum, Pseudoxytenanthera, Pseudozoysia, Psilolemma, Psilurus, Puccinellia, ×
 25 Pucciphippsia, Puelia, Pyrrhanthera, Racemobambos, Raddia, Raddiella, Ratzeburgia,
 Redfieldia, Reederchloa, Rehia, Reimarochloa, Reitzia, Relchela, Reynaudia, Rheochloa,
 Rhipidocladum, Rhizocephalus, Rhombolytrum, Rhynchoryza, Rhytachne, Richardsiella,
 Rostraria, Rottboellia, Rytidosperma, Saccharum, Sacciolepis, Sartidia, Sasa, Schaffnerella,
 Schedonnardus, Schedonorus, Schismus, Schizachne, Schizachyrium, Schizostachyum,
 30 Schmidtia, Schoenefeldia, Sclerochloa, Sclerodactylon, Scleropogon, Scolochloa, Scribneria,
 Scutachne, Secale, Sehima, Semiarundinaria, Sesleria, Setaria, Setariopsis, Shibataea,
 Silentvalleya, Simplicia, Sinoarundinaria, Sinobambusa, Sinocalamus, Sinochasea, Sirochloa,
 Sitanion, Snowdenia, Soderstromia, Sohnsia, Sorghastrum, Sorghum, Sorgum, Spartina,
 Spartochloa, Spathia, Sphaerobambos, Sphaerocaryum, Spheneria, Sphenopholis, Sphenopus,
 35 Spinifex, Spodiopogon, Sporobolus, Steinchisma, Steirachne, Stenotaphrum, Stephanachne,
 Stereochlaena, Steyermarkochloa, Stipa, Stipagrostis, × Stiporyzopsis, Streblochaete,
 Streptochaeta, Streptogyna, Streptolophus, Streptostachys, Stypeiochloa, Sucrea, Suddia,

Swallenia, Symplectrodia, Taeniatherum, Taeniorhachis, Tarigidia, Tatianyx, Temburongia, Temochloa, Tetrachaete, Tetrachne, Tetrapogon, Thamnocalamus, Thaumastochloa, Thelepogon, Themeda, Thrasya, Thrasyopsis, Thuarea, Thyridachne, Thyridolepis, Thyrsochloa, Thysanolaena, Torreyochloa, Tovarochloa, Trachypogon, Tragus, Tribolium, Trichloris, Tricholaena, Trichoneura, Trichopteryx, Tridens, Trikeria, Trilobachne, Triniochloa, Triodia, Triplachne, Triplasis, Triplopogon, Tripogon, Tripsacum, Triraphis, Triscenia, Trisetaria, × Trisetokoeleria, Trisetum, Tristachya, × Triticale, × Triticosecale, Triticum, × Trititrigia, Tuctoria, Uniola, Uranthoecium, Urelytrum, Urochlaena, Urochloa, Urochondra, Valiha, Vaseyochloa, Ventenata, Vietnamocalamus, Vietnamochloa, Vietnamosasa, Viguierella, Vossia, Vulpia, Vulpiella, Wangenheimia, Weingaertneria, Whiteochloa, Willkommia, XAegilotriticum, XAgropogon, XAgrotrigia, XArctodupontia, XCalammophila, XDanthosieglia, XElyhordeum, XElymus, XHaynaldoticum, XLeysdeum, XLeysmotrigia, XPhyllosasa, XTrisetokoeleria, XTriticale, Xanthochloa, Xerochloa, Yakirra, Yushania, Yvesia, Zea, Zenkeria, Zeugites, Zingeria, Zizania, Zizaniopsis, Zonotriche, Zoysia and Zygochloa.

Ryegrass plants are of the genus *Lolium*. Known ryegrass plants include *Lolium canariense* (Canary Islands ryegrass), *Lolium × festucaceum*, *Lolium gracile*, *Lolium × hubbardii*, *Lolium × hybridum*, *Lolium multiflorum* (Italian ryegrass), *Lolium perenne* (perennial ryegrass), *Lolium persicum* (Persian ryegrass or Persian darnel), *Lolium remotum*, *Lolium rigidum* (Stiff darnel, Wimmera ryegrass) and *Lolium temulentum* (Darnel, poison darnel).

Reports of SI have been published for numerous grass family plants, including *Briza media*, *Cynosurus cristatus*, *Dactylis aschersoniana*, *Festuca pratensis*, Italian ryegrass, *Anthoxanthum odoratum*, *Phalaris coarulescens*, rye, bulbous barley, *Sorghastrum nutans*, *Chloris gayana*, *Molinia caerulea* and *Oryza barthii*, as well as perennial ryegrass (Yang et al., 2008) and others for which Z locus alleles are exemplified here.

Genetic constructs and sequences

Nucleic acid according to the invention may comprise a single allele isolated from the Z locus. Alternatively it may a pair of Z alleles, which may both be isolated from a Z locus, or from different Z loci. Nucleic acid may comprise a pair of Z alleles, which may be located adjacent to each other, and may be expressed from separate promoters. A pair of Z alleles may comprise interacting Z alleles, or non-interacting alleles. Nucleic acid according to the invention may comprise allele combinations that are not naturally occurring in allogamous Poaceae, including combinations in which the Z alleles are non-interacting.

Nucleic acid molecules and vectors according to the present invention may be provided isolated and/or purified from their natural environment, in substantially pure or homogeneous form, or free or substantially free of nucleic acid or genes of the species of interest or origin other than the sequence encoding a polypeptide with the required function. Nucleic acid

according to the present invention may include cDNA, RNA, genomic DNA and may be wholly or partially synthetic. The term "isolated" encompasses all these possibilities. Where a DNA sequence is specified, e.g. with reference to a figure, unless context requires otherwise the RNA equivalent, with U substituted for T where it occurs, is encompassed.

5 Nucleic acid according to the present invention may consist essentially of or consist of the relevant coding sequence. Nucleic acid according to the present invention may include a promoter or other regulatory sequence as discussed further elsewhere herein, and such regulatory sequence may be heterologous to the coding sequence, that is to say not naturally operably linked with the coding sequence. Nucleic acid according to the present invention may
10 genomic DNA or it may be cDNA or lacking one or more introns which occur naturally, or may be in any non-naturally occurring form. A coding sequence in accordance with the present invention may be included with a larger nucleic acid molecule of less than about 10,000 nucleotides, less than about 5,000 nucleotides or less than about 2,000 nucleotides.

Transcriptional control sequences will be found 5' to the open reading frame of the gene.
15 They are obtainable by probing a genomic DNA library with a nucleic acid of the invention, selecting a clone which hybridises under conditions of medium to high stringency, and sequencing the clone 5' to the open reading frame of the gene. Where only a small amount of sequence is present in the 5' region, this sequence may be used to reprobe the library to genome walk further upstream. Analysis of the upstream region will reveal control regions for
20 gene expression including control regions common to many genes (such as TATA and CAAT boxes) and other control regions, usually located from 1 to 10,000, such as 1 to 1000 or 50 to 500 nucleotides upstream of the start of transcription.

To confirm that such regions are the control regions of the gene, they may be linked to a reported gene (such as β -galactosidase) and tested in any suitable *in vitro* or *in vivo* system.
25 For example the construct of the control region (e.g. comprising 50 to 500 nucleotides upstream of the start of transcription) and the reporter gene may be used to produce a transgenic plant and the pattern of expression, both spatially and developmentally, may be compared with that of the Z locus gene. Where substantially similar patterns of expression are found, this shows that the construct comprises substantially all of the control region of the wild type gene.

30 SEQ ID NO: 1 shows the nucleotide sequence of the Z locus genomic region including promoter for perennial ryegrass, and also 3' regulatory elements including termination region.

A control region may be mutated to identify specific subregions responsible for transcriptional control. This may be achieved by a number of techniques, including DNase protection footprint assays, in which the control region is brought into contact with an extract
35 from a cell in which the Z locus gene is actively expressed, and the regions of the control region which bind factors in that extract is determined.

Isolated nucleic acid comprising such control regions obtainable by such a method are part of the present invention.

The present invention further extends to genomic DNA exon sequences found between the introns of a Z locus gene. Such exon sequences may be obtained in a manner analogous to that described above for the transcriptional control sequences, with the appropriate genome walking being conducted between the intron sequences. The locations of the exons may be determined by comparing genomic and cDNA sequences of the gene, observing where the sequences line up and diverge, and looking for consensus splice sequences which define intron/exon boundaries.

Nucleic acid according to the invention may be in the form of a recombinant and preferably replicable vector, for example a plasmid, cosmid, phage or expression vector. The nucleic acid may be under the control of an appropriate promoter or other regulatory elements for expression in a host cell such as a microbial, e.g. bacterial, or plant cell. In the case of genomic DNA, this may contain its own promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the host cell. A vector including nucleic acid according to the present invention need not include a promoter or other regulatory sequence, particularly if the vector is to be used to introduce the nucleic acid into cells for recombination into the genome.

Those skilled in the art are well able to construct vectors and design protocols for recombinant gene expression. Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. For further details see, for example, *Molecular Cloning: a Laboratory Manual*: 2nd edition, Sambrook et al, 1989, Cold Spring Harbor Laboratory Press. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Current Protocols in Molecular Biology*, Second Edition, Ausubel et al. eds., John Wiley & Sons, 1992. The disclosures of Sambrook et al. and Ausubel et al. are incorporated herein by reference. Specific procedures and vectors previously used with wide success upon plants are described by Bevan (*Nucl. Acids Res.* 12, 8711-8721 (1984)) and Guerinéau and Mullineaux (1993) (*Plant transformation and expression vectors*. In: *Plant Molecular Biology Labfax* (Croy RRD ed) Oxford, BIOS Scientific Publishers, pp 121-148).

Selectable genetic markers may be used consisting of chimaeric genes that confer selectable phenotypes such as resistance to antibiotics such as kanamycin, hygromycin, phosphinotricin, chlorsulfuron, methotrexate, gentamycin, spectinomycin, imidazolinones and glyphosate.

As described above, the invention encompasses sequences that are variants of a nucleic acid sequence or amino acid sequence shown in the figures or the accompanying sequence listing. Changes to a sequence to produce a variant may be by one or more of addition, insertion, deletion or substitution of one or more nucleotides in the nucleic acid. Such changes may lead to the addition, insertion, deletion or substitution of one or more amino acids in the encoded polypeptide. Alternatively, changes may make no difference to the encoded amino acid sequence, and the invention includes nucleic acid sequences that are degeneratively equivalent to those set out in this specification are included. Changes in sequence may or may not disrupt or alter the gene function. Introduction of a frame-shift or stop codon in a gene may abolish gene expression or polypeptide production, or may result in expression of a truncated product, which may or may not be functional e.g. it may or may not retain a capability of interacting with a polypeptide expressed by an interacting Z locus allele. A point mutation or gross mutational change to the encoded polypeptide, including insertion, deletion, substitution and/or addition of one or more amino acids or regions in the polypeptide, may also affect its function. A mutation in a promoter sequence or other regulatory region may prevent or reduce expression from the gene or affect the processing or stability of the mRNA transcript.

A variant amino acid sequence in accordance with the present invention may include within a sequence shown herein a single amino acid change with respect to the sequence shown in the relevant figure or SEQ ID NO, or 2, 3, 4, 5, 6, 7, 8, or 9 changes, about 10, 15, 20, 30, 40 or 50 changes, or greater than about 50, 60, 70, 80 or 90 changes. In addition to one or more changes within the amino acid sequence shown in the relevant figure or SEQ ID NO, a variant amino acid sequence may include additional amino acids at the C-terminus and/or N-terminus.

Percent (%) sequence identity with respect to a particular reference sequence is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the reference sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. A % sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region. The "longer" sequence is the one having the most nucleotides in the aligned region. In a similar manner, percent amino acid sequence identity with respect to a reference amino acid sequence is defined as the percentage of amino acid residues in a candidate sequence that are identical with the nucleotide residues in the reference sequence. A number of software programmes for calculating sequence identity are available, such as CLUSTAL W described by Thompson et al. (1994) *Nucleic Acids Res* 22: 4673–4680. CLUSTAL W was used for the multiple sequence alignments shown in the figures. An alternative is BLAST, originally described in Altschul et al. (1990) *J. Mol. Biol.* 215: 403-10.

Sequence identity is generally over the full-length of the relevant sequence shown, unless stated otherwise, or it may be over a contiguous sequence of about or greater than about 20, 25, 30, 33, 40, 50, 67, 133, 167, 200, 233, 267, 300, 333, 400, 450, 500, 550, 600 or
5 more amino acids or codons, compared with the relevant amino acid sequence or nucleotide sequence as the case may be.

Polynucleotides of the invention may be obtained by site directed mutagenesis of the sequences of shown in the figures or sequence listing or allelic variants thereof. This may be useful where for example silent codon changes are required to sequences to optimise codon
10 preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides. Further changes may be desirable to represent particular coding changes which are required to provide, for example, conservative substitutions.

15 Nucleic acid according to the invention may comprise or consisting essentially of a sequence of nucleotides complementary to a nucleotide sequence hybridisable with any coding sequence shown in the figures or the sequence listing. Another way of looking at this would be for nucleic acid according to this aspect to be hybridisable with a nucleotide sequence complementary to the coding sequence. Of course, DNA is generally double-stranded and
20 blotting techniques such as Southern hybridisation are often performed following separation of the strands without a distinction being drawn between which of the strands is hybridising. The hybridisable nucleic acid or its complement may encode a functional Z locus allele. Preferred conditions for hybridisation are familiar to those skilled in the art, but are generally stringent enough for there to be positive hybridisation between the sequences of interest to the exclusion
25 of other sequences.

We describe how other alleles of the Z locus genes may be obtained, for example by making or obtaining cDNA libraries made from dividing cells or tissues or genomic DNA libraries from plants of the same or other Poaceae species, and probing such libraries with probes comprising all or part of a nucleic acid of the invention under conditions of medium to high
30 stringency (for example for hybridisation on a solid support (filter) overnight incubation at 42°C in a solution containing 50 % formamide, 5 x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulphate and 20 µg/ml salmon sperm DNA, followed by washing in 0.03 M sodium chloride and 0.03 M sodium citrate (i.e. 0.2 x SSC) at from about 50°C to about 60°C).

35 Thus the present invention provides an isolated nucleic acid which hybridises to the nucleotide sequence shown in a figure or SEQ ID NO herein under the abovementioned

hybridisation and washing conditions. Such a nucleic acid is suitable for use as a probe for detecting the Z locus gene, for example in Southern blots.

In the context of cloning, it may be necessary for one or more gene fragments to be ligated to generate a full-length coding sequence. Also, where a full-length encoding nucleic acid molecule has not been obtained, a smaller molecule representing part of the full molecule, may be used to obtain full-length clones. Inserts may be prepared from partial cDNA clones and used to screen cDNA libraries. The full-length clones isolated may be subcloned into expression vectors and activity assayed by transfection into suitable host cells, e.g. with a reporter plasmid.

Z locus gene products

Polypeptides encoded by Z locus genes according to the invention are representing members of the FGGY family of carbohydrate kinases. A Z locus gene according to the invention may encode a glycerol kinase. A Z locus gene according to the invention may encode a polypeptide that binds a polypeptide expressed from another Z locus gene, i.e. the polypeptide products of interacting Z alleles may bind one another.

The present invention extends to the production and use of polypeptides encoded by the nucleic acids, and fragments of the full-length polypeptides disclosed herein, especially active portions thereof. An "active portion" of a polypeptide means a peptide which is less than said full length polypeptide, but which retains an essential biological activity, for example it may have a carbohydrate kinase activity e.g. glycerol kinase activity, and/or the ability to interact with, e.g. bind, a polypeptide expressed by an interacting Z locus allele, for example causing an SI phenotype in a plant in which both polypeptides are expressed. The active portion may be, or may comprise, the VR, upstream conserved region and/or downstream conserved region.

A "fragment" of a polypeptide means a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids. Fragments of the polypeptides may include one or more epitopes useful for raising antibodies to a portion of any of the amino acid sequences disclosed herein. Preferred epitopes are those to which antibodies are able to bind specifically, which may be taken to be binding a polypeptide or fragment thereof of the invention with an affinity which is at least about 1000x that of other polypeptides.

Purified protein according to the present invention, or a fragment or variant thereof, e.g. produced recombinantly by expression from encoding nucleic acid, may be used to raise antibodies or to isolate antibodies from a library. Antibodies and polypeptides comprising antigen-binding fragments of antibodies may be used in identifying orthologues, and also in identifying complexes containing a Z locus gene product.

A sample may be tested for the presence of a Z locus gene product using an antibody (or mixture of antibodies) specific for one or more particular variants of a polypeptide encoded by a Z locus allele. For example, an antibody may specifically bind the variable region of a polypeptide encoded by a Z locus allele, enabling specific identification of the particular allele.

5 Alternatively it may bind a conserved region of the polypeptide, enabling identification of a Z locus gene product in the sample.

The sample may be tested by being contacted with the antibody under appropriate conditions for specific binding, before binding is determined, for instance using a reporter system. Where a panel of antibodies is used, different reporting labels may be employed for
10 each antibody so that binding of each can be determined.

An antibody may be used to isolate and/or purify its binding partner polypeptide from a test sample, to allow for sequence and/or biochemical analysis of the polypeptide to determine whether it has the sequence and/or properties of the wild-type polypeptide or a particular mutant, variant or allele thereof. Amino acid sequence is routine in the art using automated sequencing
15 machines.

Candidate polypeptides for screening may for instance be the products of an expression library created using nucleic acid derived from a plant of interest, or may be the product of a purification process from a natural source. A polypeptide found to bind the antibody may be isolated and then may be subject to amino acid sequencing. Any suitable technique may be
20 used to sequence the polypeptide either wholly or partially (for instance a fragment of the polypeptide may be sequenced). Amino acid sequence information may be used in obtaining nucleic acid encoding the polypeptide, for instance by designing one or more oligonucleotides (e.g. a degenerate pool of oligonucleotides) for use as probes or primers in hybridisation to candidate nucleic acid, or by searching computer sequence databases, as discussed further
25 below.

Screening methods, probes and primers

Z locus alleles in a plant can be identified at DNA and/or polypeptide level. At the DNA level, a Z locus allele may be identified in a preparation of nucleic acid from the plant by using probes or primers to screen the sample for nucleic acid including the nucleotide sequence of a
30 Z locus gene, or by other characterising methods such as performing DNA fingerprinting to compare the restriction pattern produced when a restriction enzyme cuts nucleic acid in the sample with the restriction pattern obtained from a nucleotide sequence shown herein, or from a known mutant, allele or variant thereof. At the polypeptide level, a polypeptide including a Z locus amino acid sequence can be identified in a sample obtained from a plant, e.g. a pollen
35 sample, and it may be determined whether the polypeptide is full length, and/or is mutated, and/or is expressed at the normal level. Polypeptides may be identified by contacting the

sample with an agent capable of specifically binding to a Z locus polypeptide, such as an antibody, and determining binding.

Various aspects of the invention include methods in which nucleic acid of a plant is screened to determine the presence or identity of one or more Z locus alleles in the plant.

5 Probes and primers are provided for use in such methods. Examples of possible techniques, and factors involved in designing oligonucleotide primers and probes, are discussed below. The present disclosure provides sufficient information for a person skilled in the art to obtain genomic DNA sequence for any given new or existing allele and devise a suitable nucleic acid- and/or polypeptide-based diagnostic assay. In designing a nucleic acid assay account is taken
10 of the distinctive variation in sequence that characterises the particular variant allele.

The sequence information provided herein allows the design of diagnostic tests for determination of the presence of a specific gene or allele thereof in any given plant, cultivar, variety, population, landrace, part of a family or other selection in a breeding programme or other such genotype. A diagnostic test may be based on determination of the presence or
15 absence of a particular allele by means of nucleic acid or polypeptide determination.

At the nucleic acid level, this may involve hybridisation of a suitable oligonucleotide, such as a fragment of the gene or a homologue thereof, including any homologue disclosed herein, or any particular allele, such as an allele which gives a desirable phenotype, such as any such allele disclosed herein. The hybridisation may involve a PCR designed to amplify a
20 product from a given allelic version of the gene, with subsequent detection of an amplified product by any of a number of possible methods including but not limited to gel electrophoresis, capillary electrophoresis, direct hybridisation of nucleotide sequence probes and so on. A diagnostic test may be based on PCR designed to amplify various alleles or any allele from the relevant locus, with a test to distinguish the different possible alleles by any of a number of
25 possible methods, including DNA fragment size, restriction site variation (e.g. CAPS - cleaved amplified polymorphic sites) and so on. A diagnostic test may also be based on a great number of possible variants of nucleic acid analysis that will be apparent to those skilled in the art, such as use of a synthetic sequence as a hybridisation probe.

We have developed a marker assay based on high resolution melting of PCR-amplified
30 DNA fragments that can be used to visualise and predict haplotype combinations for the Z SI locus. High resolution melting curve analysis (HRM) is a method which measures dissociation of double stranded DNA from a PCR product amplified in the presence of a saturating fluorescence dye. The final PCR products are heated causing the dissociation of the double strand DNA, accompanied by a decrease in fluorescence. The degree of fluorescence is
35 recorded continuously over time and relates to the rates of double strand dissociation. The shape of the resulting melting curve depends on the reaction specificity and the thermal stability of a PCR amplicon determined by its length, GC content and base sequence. HRM is an easy,

fast and closed-tube procedure in one step. The sensitivity of HRM proved successful to genotype plant DNA sequence polymorphisms such as SSRs and SNPs (Studer et al. 2009) and can be used as a highly sensitive method to discriminate allelic constitutions at the Z SI locus.

HRM results in melting curves which are characteristic for each haplotype combination, i.e. it is a mixture of both alleles that is determined by the melting curve shape. Primer sequences specific to one gene (as for example primer 6922_f that is specific for LpGK2) allow discriminating allelic constitutions at either LpGK1 or LpGK2. Primer sequences used for cloning and HRM analysis of LpGK1 and LpGK2 in perennial ryegrass were:

>LpGK_VR_R_f SEQ ID NO: 38
GATCTGGAACCTCACAGGAG

>LpGK_VR_R_r SEQ ID NO: 39
CTCCTGTGAGGTTCCAGATC

>LpGK_VR_L_f SEQ ID NO: 40
TGTCACAGACTGCTCAAAATGC

>LpGK_VR_L_r SEQ ID NO: 41
GCATTTGAGCAGTCTGTGACA

>6922_f SEQ ID NO: 42
CACTTTGTGGAGACGTGTGG

Other primers specific for LpGK1 include:

>10466_f SEQ ID NO: 43
CACTTTGTGGCCTACCAATC

and

>10461_f SEQ ID NO: 44
GAACACACTTTGTGGCCTACC

Either of these forward primers may be used with a reverse primer, for example, LpGK_VR_L_r, to specifically amplify the variable region in LpGK1.

Techniques such as HRM allow the identity of Z locus alleles in a plant to be determined without a need to sequence the gene, although sequencing provides an alternative method and/or may be used to confirm the identity of alleles present.

As described above, Z locus alleles can be identified in a plant of the grass family, by a method comprising

- providing a preparation of nucleic acid from the plant;
- providing a probe or primer that hybridises to the Z locus;
- annealing the probe or primer to the Z locus in the nucleic acid from the plant; and

identifying the Z locus allele present at the Z locus by detecting hybridisation and/or by obtaining an amplification product.

The nucleic acid preparation may contain genomic DNA from the plant, either purified or unpurified, and may contain the whole genome or a part of the genome comprising the Z locus.

5 When screening for a Z locus allele nucleic acid, the nucleic acid in the sample will usually initially be amplified, e.g. using PCR, to increase the amount of the analyte as compared to other sequences present in the sample. This allows the target sequences to be detected with a high degree of sensitivity if they are present in the sample. However, this initial step may be avoided by using highly sensitive array techniques. Nucleic acid may then be sequenced
10 and/or tested in any other way to determine the presence or absence of a particular feature. Nucleic acid for testing may be prepared from nucleic acid removed from cells or in a library using a variety of other techniques such as restriction enzyme digest and electrophoresis.

A method may include hybridisation of one or more (e.g. two) probes or primers to target nucleic acid. Where the nucleic acid is double-stranded DNA, hybridisation will generally be
15 preceded by denaturation to produce single-stranded DNA. The hybridisation may be as part of a PCR procedure.

Where the probe or primer is specific for a particular Z allele sequence, the presence or absence of the hybridisation or of the amplification product respectively is a direct indicator of the presence or absence of that particular allele in the plant. Allele- or variant-specific
20 oligonucleotides may be used in PCR to specifically amplify particular sequences if present in a test sample. Assessment of whether a PCR band contains a gene variant may be carried out in a number of ways familiar to those skilled in the art. The PCR product may for instance be treated in a way that enables one to display the mutation or polymorphism on a denaturing polyacrylamide DNA sequencing gel, with specific bands that are linked to the gene variants
25 being selected.

In many situations it will be convenient to use primers that are capable of hybridising to many different Z locus alleles, in order to amplify nucleic acid for whichever Z locus allele or alleles are present in the plant. This identifies the presence of the Z locus gene in the plant, and may be followed by determining the identity of the particular allele or alleles. Such
30 identification may use sequencing and/or may use the HRM technique as described.

For example, a method of identifying a Z locus allele in a plant may comprise:
providing a preparation of nucleic acid from the plant;

amplifying all or part of a Z locus allele from the nucleic acid preparation to obtain an amplification product, e.g. by annealing a pair of primers to the nucleic acid and using PCR; and
35 characterising the amplification product and thereby determining the identity of the Z locus allele.

The primers may amplify the Z locus nucleic acid comprising LpGK1 and LpGK2. They may amplify any LpGK1 or LpGK2 haplotype present and the method may therefore produce an amplification product that contains all LpGK1 and LpGK2 alleles from the plant. This may be a mixture of different alleles where the plant is heterozygous.

Identity of the allele can be determined by sequencing, by annealing a probe specific to the allele sequence and detecting hybridisation, and/or by using a technique such as HRM. The amplification product will typically be double stranded, and may be characterised and compared with characterising features of known Z locus allele sequences in order to identify the allele or alleles present in the nucleic acid preparation.

The method may comprise identifying the Z locus haplotype of the plant, for example using a fingerprinting technique such as HRM. The Z locus haplotype is a product of the different alleles present at the Z locus.

A double stranded DNA amplification product may be characterised by measuring dissociation of the double stranded DNA in the presence of a saturating fluorescence dye, to obtain a melting curve. Since the melting curve is characteristic for the particular allele combination, comparison of the melting curve with melting curves of known alleles allows identification of the allele present in the nucleic acid preparation. If the curve does not match a melting curve of a known allele or haplotype combination, this is indicative that the allele is a previously uncharacterised Z locus allele or that the plant contains a previously uncharacterised allele or haplotype combination. Nucleic acid sequencing may be used to confirm the identification by confirming that the sequence of the allele is the same as the sequence of a known allele, and in the case of a new allele sequencing provides the new allele sequence. Sequencing may involve obtaining the sequence of the amplification product alone, and/or sequencing the full gene or full coding region.

Accordingly, a method may comprise determining the identity of Z locus genes in a plant by providing a preparation of DNA from the plant;

amplifying all or part of a Z locus allele from the DNA preparation to obtain a double stranded amplification product, e.g. by annealing a pair of primers to the nucleic acid and using PCR; and

characterising the amplification product.

Comparison of characteristics of the amplification product with characteristics of amplification products from plants with known Z locus genes, e.g. known haplotypes, may be used to identify the Z locus genes. HRM is a preferred fingerprinting technique, as described above. Accordingly, characterisation of the amplification product may comprise measuring dissociation of the two DNA strands to obtain a melting curve and comparing the melting curve with melting curves of known Z locus haplotypes. The method may comprise obtaining a match

between the characteristics of the amplification product and characteristics of known Z locus genes, thereby determining the identity of the Z locus genes or haplotype in the plant.

Such techniques are convenient for rapid identification of Z locus haplotypes and may be used for methods of predicting phenotypes and predicting compatibility in plant crosses. For populations of interest, libraries of haplotypes may be determined and characterised by HRM for this purpose. Alternatively, of course, sequencing can be used to conclusively determine Z locus genes and haplotypes in a plant.

Tests may be carried out on preparations containing genomic DNA, cDNA and/or mRNA. Nucleic acid in a test sample may be sequenced and the sequence compared with a sequence shown herein to determine whether or not a difference is present. If so, the difference can be compared with known alleles to determine whether the test nucleic acid contains one or more of the variations indicated, or the difference can be investigated for association with a desired phenotype.

Suitable oligonucleotide sequences can be generated for use as probes and primers based on the conserved or hypervariable regions of the Z locus nucleic acid. Since the VR of an allele appears to be the region responsible for determining the compatibility phenotype in the plant, this region is of particular interest for characterising and determining the identity of alleles. Preferably, therefore, the amplification product contains the VR. The amplification product may also contain all or part of one or more conserved regions, e.g. it may contain at least part of the upstream conserved region and at least part of the downstream conserved region. Primers may conveniently be annealed to conserved regions to amplify the nucleic acid. Sequences of conserved regions are illustrated in the drawings and included in the accompanying sequence listing.

PCR primers derived from the Z locus sequences described here may readily be tested for their specificity for amplifying nucleic acid according to the present invention, using both genomic DNA and RT-PCR templates. Cloning and subsequent sequencing of PCR products may be used to indicate amplification of the expected derived gene fragment. Full length cDNA clones can be obtained as described by 5' and 3' RACE technology if RT-PCR products are used as templates.

Oligonucleotides designed to amplify DNA sequences may be used in PCR reactions or other methods involving amplification of nucleic acid, using routine procedures. See for instance "PCR protocols; A Guide to Methods and Applications", Eds. Innis et al, 1990, Academic Press, New York. On the basis of amino acid sequence information, oligonucleotide probes or primers may be designed, taking into account the degeneracy of the genetic code, and, where appropriate, codon usage of the organism from which the candidate nucleic acid is derived.

A primer or probe in accordance with certain embodiments of the invention, e.g. for use in nucleic acid amplification, may be up to about 50 nucleotides, or about 40 nucleotides or about 30 or fewer nucleotides in length (e.g. up to 18, up to 21 or up to 24 nucleotides in length). Oligonucleotide probes or primers according to the present invention may be fragments of any of the sequences shown herein, or any allele associated with a desired phenotype are at least about 10 nucleotides in length, more preferably at least about 15 nucleotides in length, more preferably at least about 20 nucleotides in length, more preferably about 30 nucleotides in length. Such fragments themselves individually represent aspects of the present invention. Fragments and other oligonucleotides may be used as primers or probes as discussed but may also be generated (e.g. by PCR) in methods concerned with determining the presence in a test sample of a sequence indicative of a desired phenotype.

Z locus nucleic acid may be obtained or identified through hybridisation of an oligonucleotide or a nucleic acid molecule comprising such an oligonucleotide to target/candidate nucleic acid. Successful hybridisation may be identified and target/candidate nucleic acid isolated for further investigation and/or use.

Hybridisation may involve probing nucleic acid and identifying positive hybridisation under suitably stringent conditions (in accordance with known techniques) and/or use of oligonucleotides as primers in a method of nucleic acid amplification, such as PCR. For probing, preferred conditions are those which are stringent enough for there to be a simple pattern with a small number of hybridisations identified as positive which can be investigated further. It is well known in the art to increase stringency of hybridisation gradually until only a few positive clones remain.

For instance, screening may initially be carried out under conditions, which comprise a temperature of about 37°C or more, a formamide concentration of less than about 50%, and a moderate to low salt (e.g. Standard Saline Citrate ('SSC') = 0.15 M sodium chloride; 0.15 M sodium citrate; pH 7) concentration.

Alternatively, a temperature of about 50°C or more and a high salt (e.g. 'SSPE' = 0.180 M sodium chloride; 9 mM disodium hydrogen phosphate; 9 mM sodium dihydrogen phosphate; 1 mM sodium EDTA; pH 7.4). Preferably the screening is carried out at about 37°C, a formamide concentration of about 20%, and a salt concentration of about 5 X SSC, or a temperature of about 50°C and a salt concentration of about 2 X SSPE. These conditions will allow the identification of sequences which have a substantial degree of homology (similarity, identity) with the probe sequence, without requiring the perfect homology for the identification of a stable hybrid.

Suitable conditions include, e.g. for detection of sequences that are about 80-90% identical, hybridization overnight at 42°C in 0.25M Na₂HPO₄, pH 7.2, 6.5% SDS, 10% dextran sulphate and a final wash at 55°C in 0.1X SSC, 0.1% SDS. For detection of sequences that are

greater than about 90% identical, suitable conditions include hybridization overnight at 65°C in 0.25M Na₂HPO₄, pH 7.2, 6.5% SDS, 10% dextran sulphate and a final wash at 60°C in 0.1X SSC, 0.1% SDS.

An alternative is a solution of 5x SSPE (final 0.9 M NaCl, 0.05 M sodium phosphate, 0.005 M ethylenediaminetetraacetic acid (EDTA) pH 7.7), 5X Denhardt's solution, 0.5% SDS (sodium dodecyl sulphate), at 65°C overnight, (for high stringency, highly similar sequences) or 50°C (for low stringency, less similar sequences). Washes in 0.2 x SSC/0.1% SDS at 65°C for high stringency, alternatively at 50-60 °C in 1x SSC/0.1 % SDS for low stringency.

The present invention extends to nucleic acid selectively hybridisable under high stringency with nucleic acid shown in the figures or in the sequence listing.

As noted, nucleic acid according to the present invention is obtainable using oligonucleotides, designed on the basis of sequence information provided herein, as probes or primers. Nucleic acid isolated and/or purified from one or more cells of a plant (see above), or a nucleic acid library derived from nucleic acid isolated and/or purified from the plant (e.g. a cDNA library derived from mRNA isolated from the plant), may be probed under conditions for selective hybridisation and/or subjected to a specific nucleic acid amplification reaction such as PCR. The nucleic acid probed or used as template in the amplification reaction may be genomic DNA, cDNA or RNA. If necessary, one or more gene fragments may be ligated to generate a full-length coding sequence.

Various aspects of the present invention include the obtainable nucleic acid, methods of screening material, e.g. cell lysate, nucleic acid preparations, for the presence of nucleic acid of interest, methods of obtaining the nucleic acid, and suitable primers and primer combinations.

Nucleic acid may be screened using a variant- or allele-specific probe. Such a probe may hybridise with a region of the gene, or its complement, containing a marker of Z locus self-incompatibility, such as the VR. Under suitably stringent conditions, specific hybridisation of such a probe to test nucleic acid is indicative of the presence of the particular allele or variant comprising that VR in the test nucleic acid. For efficient screening purposes, more than one probe may be used on the same test sample.

Nucleic acid isolated and/or purified from one or more cells of a plant or a nucleic acid library derived from nucleic acid isolated and/or purified from cells (e.g. a cDNA library derived from mRNA isolated from the cells), may be probed under conditions for selective hybridisation and/or subjected to a specific nucleic acid amplification reaction such as PCR.

Approaches which rely on hybridisation between a probe and test nucleic acid and subsequent detection of a mismatch may be employed. Under appropriate conditions (temperature, pH etc.), an oligonucleotide probe will hybridise with a sequence which is not entirely complementary. The degree of base-pairing between the two molecules will be

sufficient for them to anneal despite a mis-match. Various approaches are well known in the art for detecting the presence of a mis-match between two annealing nucleic acid molecules.

For instance, RNase A cleaves at the site of a mis-match. Cleavage can be detected by electrophoresing test nucleic acid to which the relevant probe or probe has annealed and
5 looking for smaller molecules (i.e. molecules with higher electrophoretic mobility) than the full length probe/test hybrid. Other approaches rely on the use of enzymes such as resolvases or endonucleases.

Thus, an oligonucleotide probe that has the sequence of a region of the normal gene (either sense or anti-sense strand) in which mutations associated with particular phenotypes are
10 known to occur may be annealed to test nucleic acid and the presence or absence of a mis-match determined. Detection of the presence of a mis-match may indicate the presence in the test nucleic acid of a mutation. On the other hand, an oligonucleotide probe that has the sequence of a region of the gene including a mutation may be annealed to test nucleic acid and the presence or absence of a mis-match determined. The presence of a mis-match may
15 indicate that the nucleic acid in the test sample has the normal sequence, or a different mutant or allele sequence. In either case, a battery of probes to different regions of the gene may be employed.

The presence of differences in sequence of nucleic acid molecules may be detected by means of restriction enzyme digestion, such as in a method of DNA fingerprinting where the
20 restriction pattern produced when one or more restriction enzymes are used to cut a sample of nucleic acid is compared with the pattern obtained when a sample containing the normal gene or a variant or allele is digested with the same enzyme or enzymes.

The presence of absence of a lesion in a promoter or other regulatory sequence may also be assessed by determining the level of mRNA production by transcription or the level of
25 polypeptide production by translation from the mRNA.

Binding of a probe to target nucleic acid (e.g. DNA) may be measured using any of a variety of techniques at the disposal of those skilled in the art. For instance, probes may be radioactively, fluorescently or enzymatically labelled. Other methods not employing labelling of
30 probe include examination of restriction fragment length polymorphisms, amplification using PCR, RNAase cleavage and allele specific oligonucleotide probing.

Probing may employ the standard Southern blotting technique. For instance DNA may be extracted from cells and digested with different restriction enzymes. Restriction fragments may then be separated by electrophoresis on an agarose gel, before denaturation and transfer to a nitrocellulose filter. Labelled probe may be hybridised to the DNA fragments on the filter
35 and binding determined. DNA for probing may be prepared from RNA preparations from cells.

Preliminary experiments may be performed by hybridising under low stringency conditions various probes to Southern blots of DNA digested with restriction enzymes. Suitable

conditions would be achieved when a large number of hybridising fragments were obtained while the background hybridisation was low. Using these conditions nucleic acid libraries, e.g. cDNA libraries representative of expressed sequences, may be searched. As noted, those skilled in the art are well able to employ suitable conditions of the desired stringency for selective hybridisation, taking into account factors such as oligonucleotide length and base composition, temperature and so on.

The use of diagnostic tests for alleles allows the researcher or plant breeder to establish, with full confidence and independent from time consuming biochemical tests, whether or not a particular allele is present in the plant of interest (or a cell thereof), whether the plant is a representative of a collection of other genetically identical plants (e.g. an inbred variety or cultivar) or one individual in a sample of related (e.g. breeders' selection) or unrelated plants.

In a breeding scheme based on selection and selfing of desirable individuals, nucleic acid or polypeptide diagnostics for the desirable allele or alleles in high throughput, low cost assays as provided by this invention, reliable selection for outperforming plant material can be made at early generations and on more material than would otherwise be possible. This gain in reliability of selection plus the time saving by being able to test material earlier and without costly phenotype screening is of considerable value in plant breeding.

Nucleic acid-based determination of the presence or absence of one or more alleles may be combined with determination of the genotype of the flanking linked genomic DNA and other unlinked genomic DNA using established sets of markers such as RFLPs, microsatellites or SSRs, AFLPs, RAPDs etc. This enables the researcher or plant breeder to select for or against the presence of the allele and also for individual plant or families of plants which have the most desirable combinations of linked and unlinked genetic background. Such recombinations of may occur only rarely within a given segregating breeding population or backcross progeny. Direct assay of the locus allows the researcher to make a stepwise approach to fixing (making homozygous) the desired combination of flanking markers and alleles, by first identifying individuals fixed for one flanking marker and then identifying progeny fixed on the other side of the locus all the time knowing with confidence whether the allele is still present.

Control of gene expression

The invention provides methods of influencing SI in a plant, including causing or allowing expression from nucleic acid according to the invention within cells of the plant. This may be used to suppress or enhance activity of a Z locus allele. For example, SI can be suppressed as a result of under-expression of one or more Z alleles within the stigma and/or pollen.

As already noted, methods of the invention include switching off SI for inbreeding in a hybrid breeding scheme by RNAi or similar technologies, and then switching on SI again to control pollination in a subsequent step of crossing the plant with a second plant to produce

hybrid seed. A variety of known gene control techniques may be employed in such methods, and while details of some are noted here, the skilled person may equally employ any other suitable technique.

Nucleic acid according to the invention, such as an LpGK1 or LpGK2 allele, may be placed under the control of an externally inducible gene promoter to place expression under the control of the user. An advantage of introduction of a heterologous gene and/or transcriptional control element into a plant cell, particularly when the cell is comprised in a plant, is the ability to place expression of the gene under the control of a promoter of choice, in order to be able to influence gene expression and therefore SI phenotype, according to preference. Furthermore, mutants and derivatives of the wild type gene may be used in place of the endogenous gene, for example to produce different phenotype compared with wild type.

In the present invention, over expression may be achieved by introduction of the nucleotide sequence in a sense orientation. Thus, the present invention provides a method of influencing a physical characteristic of a plant, the method including causing or allowing expression of the product (polypeptide or nucleic acid transcript) encoded by heterologous nucleic acid according to the invention from that nucleic acid within cells of the plant.

Down-regulation of expression of a target gene may be achieved using anti-sense technology or "sense regulation" ("co-suppression").

In using anti-sense genes or partial gene sequences to down regulate gene expression, a nucleotide sequence is placed under the control of a promoter in a "reverse orientation" such that transcription yields RNA which is complementary to normal mRNA transcribed from the "sense" strand of the target gene. See, for example, Rothstein et al, 1987; Smith et al,(1988) Nature 334, 724-726; Zhang et al,(1992) The Plant Cell 4, 1575-1588, English et al., (1996) The Plant Cell 8, 179-188. Antisense technology is also reviewed in Bourque, (1995), Plant Science 105, 125-149, and Flavell, (1994) PNAS USA 91, 3490-3496.

As noted above, fragments of complementary sequences can be used for antisense or RNAi inhibition of Z locus gene expression. Oligonucleotide fragments of sequences complementary to Z locus nucleic acid sequences may be used. For RNAi, the nucleic acid will typically be double stranded RNA, between 18 – 25 nucleotides in length, capable of hybridising to a Z locus gene as described herein. Longer sequences may be used, as they may be processed by the cell machinery to generate double stranded sequences that inhibit Z locus gene expression by RNAi.

An alternative is to use a copy of all or part of the target gene inserted in sense, that is the same, orientation as the target gene, to achieve reduction in expression of the target gene by co-suppression. See, for example, van der Krol et al., (1990) The Plant Cell 2, 291-299; Napoli et al., (1990) The Plant Cell 2, 279-289; Zhang et al., (1992) The Plant Cell 4, 1575-1588, and US-A-5,231,020.

The complete sequence corresponding to the coding sequence (in reverse orientation for anti-sense) need not be used. For example fragments of sufficient length may be used. It is a routine matter for the person skilled in the art to screen fragments of various sizes and from various parts of the coding sequence to optimise the level of anti-sense inhibition. It may be advantageous to include the initiating methionine ATG codon, and perhaps one or more nucleotides upstream of the initiating codon. A further possibility is to target a conserved sequence of a gene, e.g. a sequence that is characteristic of one or more genes, such as a regulatory sequence.

The sequence employed may be about 500 nucleotides or less, possibly about 400 nucleotides, about 300 nucleotides, about 200 nucleotides, or about 100 nucleotides. It may be possible to use oligonucleotides of much shorter lengths, 14-23 nucleotides, although longer fragments, and generally even longer than about 500 nucleotides are preferable where possible, such as longer than about 600 nucleotides, than about 700 nucleotides, than about 800 nucleotides, than about 1000 nucleotides or more.

It may be preferable that there is complete sequence identity in the sequence used for down-regulation of expression of a target sequence, and the target sequence, though total complementarity or similarity of sequence is not essential. One or more nucleotides may differ in the sequence used from the target gene. Thus, a sequence employed in a down-regulation of gene expression in accordance with the present invention may be a wild-type sequence (e.g. gene) selected from those available, or a mutant, derivative, variant or allele, by way of insertion, addition, deletion or substitution of one or more nucleotides, of such a sequence. The sequence need not include an open reading frame or specify an RNA that would be translatable. It may be preferred for there to be sufficient homology for the respective anti-sense and sense RNA molecules to hybridise. There may be down regulation of gene expression even where there is about 5%, 10%, 15% or 20% or more mismatch between the sequence used and the target gene.

Generally, the transcribed nucleic acid may represent a fragment of a gene, or the complement thereof, or may be a mutant, derivative, variant or allele thereof, in similar terms as discussed above in relation to alterations being made to a coding sequence and the homology of the altered sequence. The homology may be sufficient for the transcribed anti-sense RNA to hybridise with nucleic acid within cells of the plant, though irrespective of whether hybridisation takes place the desired effect is down-regulation of gene expression.

Thus, the present invention also provides a method of modifying self-fertility phenotype of a plant, the method including causing or allowing anti sense transcription from heterologous nucleic acid according to the invention within cells of the plant.

The present invention further provides the use of the Z locus nucleic acid, or a fragment, or variant thereof, for down-regulation of gene expression, particularly down-regulation of

expression of a Z locus allele, preferably in order to influence compatibility of a plant, e.g. to influence SI.

When additional copies of the target gene are inserted in sense, that is the same, orientation as the target gene, a range of phenotypes is produced which includes individuals where over expression occurs and some where under expression of protein from the target gene occurs. When the inserted gene is only part of the endogenous gene the number of under expressing individuals in the transgenic population increases. The mechanism by which sense regulation occurs, particularly down regulation, is not well understood. However, this technique is also well reported in scientific and patent literature and is used routinely for gene control. See, for example, van der Krol et al., 1990; Napoli et al., 1990; Zhang et al, 1992.

Plants and methods of generating plants

A Z locus gene or other nucleic acid provided herein, including orthologs, may be used to modify SI in a plant, e.g. a transgenic plant. Nucleic acid such as a vector as described herein may be used for the production of a transgenic plant. Such a plant may possess an altered SI phenotype compared with wild-type (that is to say a plant that is wild-type for the Z locus gene).

When introducing a chosen gene construct into a cell, certain considerations must be taken into account, well known to those skilled in the art. The nucleic acid to be inserted should be assembled within a construct which contains effective regulatory elements which will drive transcription. There must be available a method of transporting the construct into the cell. Once the construct is within the cell membrane, integration into the endogenous chromosomal material either will or will not occur. Finally, as far as plants are concerned the target cell type must be such that cells can be regenerated into whole plants.

Plants transformed with the DNA segment containing the sequence may be produced by standard techniques which are already known for the genetic manipulation of plants. DNA can be transformed into plant cells using any suitable technology, such as a disarmed Ti-plasmid vector carried by *Agrobacterium* exploiting its natural gene transfer ability (EP-A-270355, EP-A-0116718, NAR 12(22) 8711 - 87215 1984), particle or microprojectile bombardment (US 5100792, EP-A-444882, EP-A-434616) microinjection (WO 92/09696, WO 94/00583, EP 331083, EP 175966, Green et al. (1987) *Plant Tissue and Cell Culture*, Academic Press), electroporation (EP 290395, WO 8706614) other forms of direct DNA uptake (DE 4005152, WO 9012096, US 4684611), liposome mediated DNA uptake (e.g. Freeman et al. *Plant Cell Physiol.* 29: 1353 (1984)), or the vortexing method (e.g. Kindle, *PNAS U.S.A.* 87: 1228 (1990d) Physical methods for the transformation of plant cells are reviewed in Oard, 1991, *Biotech. Adv.* 9: 1-11.

Agrobacterium transformation is widely used by those skilled in the art to transform dicotyledonous species. There are various approaches used for the routine production of stable,

fertile transgenic plants in almost all economically relevant monocot plants (Toriyama, et al. (1988) Bio/Technology 6, 1072-1074; Zhang, et al. (1988) Plant Cell Rep. 7, 379-384; Zhang, et al. (1988) Theor Appl Genet 76, 835-840; Shimamoto, et al. (1989) Nature 338, 274-276; Datta, et al. (1990) Bio/Technology 8, 736-740; Christou, et al. (1991) Bio/Technology 9, 957-962; 5 Peng, et al. (1991) International Rice Research Institute, Manila, Philippines 563-574; Cao, et al. (1992) Plant Cell Rep. 11, 585-591; Li, et al. (1993) Plant Cell Rep. 12, 250-255; Rathore, et al. (1993) Plant Molecular Biology 21, 871-884; Fromm, et al. (1990) Bio/Technology 8, 833-839; Gordon-Kamm, et al. (1990) Plant Cell 2, 603-618; D'Halluin, et al. (1992) Plant Cell 4, 1495-1505; Walters, et al. (1992) Plant Molecular Biology 18, 189-200; Koziel, et al. (1993) 10 Biotechnology 11, 194-200; Vasil, I. K. (1994) Plant Molecular Biology 25, 925-937; Weeks, et al. (1993) Plant Physiology 102, 1077-1084; Somers, et al. (1992) Bio/Technology 10, 1589-1594; WO92/14828). In particular, Agrobacterium mediated transformation is an efficient alternative transformation method in monocots (Hiei et al. (1994) The Plant Journal 6, 271-282).

The generation of fertile transgenic plants has been achieved in the cereals rice, maize, 15 wheat, oat (*Avena sativa*), and barley (reviewed in Shimamoto, K. (1994) Current Opinion in Biotechnology 5, 158-162.; Vasil, et al. (1992) Bio/Technology 10, 667-674; Vain et al., 1995, Biotechnology Advances 13 (4): 653-671; Vasil, 1996, Nature Biotechnology 14 page 702).

Microprojectile bombardment, electroporation and direct DNA uptake are preferred where Agrobacterium is inefficient or ineffective. Alternatively, a combination of different 20 techniques may be employed to enhance the efficiency of the transformation process, eg bombardment with Agrobacterium coated microparticles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with Agrobacterium (EP-A-486233).

Following transformation, a plant may be regenerated, e.g. from single cells, callus tissue or leaf discs, as is standard in the art. Almost any plant can be entirely regenerated from 25 cells, tissues and organs of the plant. Available techniques are reviewed in Vasil et al., Cell Culture and Somatic Cell Genetics of Plants, Vol I, II and III, Laboratory Procedures and Their Applications, Academic Press, 1984, and Weissbach and Weissbach, Methods for Plant Molecular Biology, Academic Press, 1989.

The particular choice of a transformation technology will be determined by its efficiency 30 to transform certain plant species as well as the experience and preference of the person practising the invention with a particular methodology of choice. It will be apparent to the skilled person that the particular choice of a transformation system to introduce nucleic acid into plant cells is not essential to or a limitation of the invention, nor is the choice of technique for plant regeneration.

35 The invention further encompasses a host cell transformed with nucleic acid or a vector according to the present invention, especially a plant or a microbial cell. Thus, a host cell, such as a plant cell, including heterologous nucleic acid according to the present invention is

provided. Within the cell, the nucleic acid may be incorporated within the chromosome. There may be more than one heterologous nucleotide sequence per haploid genome.

Also according to the invention there is provided a plant cell having incorporated into its genome nucleic acid, particularly heterologous nucleic acid, as provided by the present invention, under operative control of a regulatory sequence for control of expression. The coding sequence may be operably linked to one or more regulatory sequences which may be heterologous or foreign to the gene, such as not naturally associated with the gene for its expression. The nucleic acid according to the invention may be placed under the control of an externally inducible gene promoter to place expression under the control of the user.

A suitable inducible promoter is the GST-II-27 gene promoter which has been shown to be induced by certain chemical compounds which can be applied to growing plants. The promoter is functional in both monocotyledons and dicotyledons. It can therefore be used to control gene expression in a variety of genetically modified plants, including field crops, and cereals such as wheat, barley, rice, maize and sorghum. The GST-II-27 promoter is also suitable for use in a variety of tissues, including roots, leaves, stems and reproductive tissues.

A further aspect of the present invention provides a method of making such a plant cell involving introduction of nucleic acid or a suitable vector including the sequence of nucleotides into a plant cell and causing or allowing recombination between the vector and the plant cell genome to introduce the sequence of nucleotides into the genome. The invention extends to plant cells containing nucleic acid according to the invention as a result of introduction of the nucleic acid into an ancestor cell.

The term "heterologous" may be used to indicate that the gene/sequence of nucleotides in question have been introduced into said cells of the plant or an ancestor thereof, using genetic engineering, i.e. by human intervention. A transgenic plant cell, i.e. transgenic for the nucleic acid in question, may be provided. The transgene may be on an extra-genomic vector or incorporated, preferably stably, into the genome. A heterologous gene may replace an endogenous equivalent gene, i.e. one which normally performs the same or a similar function, or the inserted sequence may be additional to the endogenous gene or other sequence. An advantage of introduction of a heterologous gene is the ability to place expression of a sequence under the control of a promoter of choice, in order to be able to influence expression according to preference. Furthermore, mutants, variants and derivatives of the wild-type gene, e.g. with higher or lower activity than wild-type, may be used in place of the endogenous gene. Nucleic acid heterologous, or exogenous or foreign, to a plant cell may be non-naturally occurring in cells of that type, variety or species. Thus, nucleic acid may include a coding sequence of or derived from a particular type of plant cell or species or variety of plant, placed within the context of a plant cell of a different type or species or variety of plant. A further possibility is for a nucleic acid sequence to be placed within a cell in which it or a homologue is

found naturally, but wherein the nucleic acid sequence is linked and/or adjacent to nucleic acid which does not occur naturally within the cell, or cells of that type or species or variety of plant, such as operably linked to one or more regulatory sequences, such as a promoter sequence, for control of expression. A sequence within a plant or other host cell may be identifiably heterologous, exogenous or foreign.

Plants which include a plant cell according to the invention are also provided, along with any part or propagule thereof, seed, selfed or hybrid progeny and descendants. A plant according to the present invention may be one which does not breed true in one or more properties. Plant varieties may be excluded, particularly registrable plant varieties according to Plant Breeders' Rights. It is noted that a plant need not be considered a "plant variety" simply because it contains stably within its genome a transgene, introduced into a cell of the plant or an ancestor thereof.

In addition to a plant, the present invention provides any clone of such a plant, seed, selfed or hybrid progeny and descendants, and any part of any of these, such as cuttings, seed. The invention provides any plant propagule, that is any part which may be used in reproduction or propagation, sexual or asexual, including cuttings, seed and so on. Also encompassed by the invention is a plant which is a sexually or asexually propagated off-spring, clone or descendant of such a plant, or any part or propagule of said plant, off-spring, clone or descendant.

Examples

As described in the Examples below, we have cloned the Z SI locus in perennial ryegrass by a classical fine mapping approach using two independent mapping populations (VrnA_Z and DTZ, consisting of more than 6,000 individuals each) that are segregating for Z. According to the expression pattern of positional candidate genes that was determined by RNA-seq of mature pollen, unpollinated stigma tissue, as well as pollinated stigmas representing incompatible and compatible pollen-stigma interactions, we found two glycerol kinase-like genes next to each other and being in complete linkage in all the 12,000 individuals tested. The two genes, referred to as LpGK1 and LpGK2, contain two and three introns, respectively. The DNA sequences of the LpGK1 and LpGK2 genes, as well as the VRs that are predictive for the Z SI locus, have been established for self-incompatible, self-compatible and double haploid plants. The corresponding proteins belong to the FGGY family of carbohydrate kinases and contain several active sites such as metal, MgATP and carbohydrate binding sites. In both genes, we have identified a highly conserved exon region surrounding a variable region (VR) that is similar for each haplotype and fully predictive for the Z SI locus. The amino acid sequence of these VRs are in complete linkage with functional Z alleles tested using an in vitro pollination test. We further developed a marker assay based on high resolution melting of PCR-amplified DNA

fragments (HRM markers, Studer et al. 2009) that can be used to visualise and predict haplotype combinations for the Z SI locus.

Results and Discussion

In well characterized single-locus SI systems of other plant families, it has been established that pollen and pistil SI determinants are physically and genetically linked in order to maintain functionality. To identify the Z SI component in perennial ryegrass, a fine mapping approach in two independent mapping populations (consisting of more than 6,000 individuals each) that were segregating for Z, has been used. As a result, the Z locus was pinpointed to a single BAC clone. BAC clone sequencing, assembly and annotation identified ten positional candidate genes. However, none of these genes showed DNA sequence homology to known Brassicaceae, Solanaceae or Papaveraceae SI genes, supporting that GSI in Poaceae is determined by a previously undescribed recognition system.

In addition, RNA-seq was used to determine tissue-specific expression patterns of positional candidate genes and to quantify gene expression in self-incompatible (after self-pollination) and compatible (after cross-pollination) pollen-pistil interactions. As a result, no significant difference in quantitative gene expression was detected when comparing self-incompatible and compatible pollen-pistil interactions. This further supported the hypothesis that gene products determining the initial SI reaction are already expressed and functionally active at the time of the pollen-pistil contact.

Based on sequence annotation and gene expression analysis, we further investigated a glycerol kinase-like gene as a prime candidate gene for Z. We identified a duplication of this gene in self-incompatible ryegrass genotypes. The two genes, thereafter referred to as LpGK1 and LpGK2, were located side by side and contain two and three introns, respectively. The corresponding proteins belong to the FGGY family of carbohydrate kinases and contain several active sites such as metal, MgATP and carbohydrate binding sites (Figure 9). In both genes, we identified a highly conserved exon region surrounding a variable sequence motif that is fully predictive for Z (Figure 1, Figure 2, Figure 3 and Figure 4). Despite a high degree of sequence conservation, several randomly distributed sequence polymorphisms between LpGK1 and LpGK2 were identified (Figure 5). For the conserved region surrounding the functional Z site, however, both the DNA and the amino acid sequence of LpGK1 and LpGK2 were highly similar, even between the grass species *Brachypodium*, barley and rice (Figure 7). Notably, the functional amino acid sequence motif of LpGK1 and LpGK2 was similar on each haplotype, as tested by cloning and sequencing of haplotypes containing both genes in a set of perennial ryegrass plants.

SI is not always fully effective and self-compatibility (SC) has been described in various SI systems (de Nettancourt. 1977). For example, it has been shown for the single-locus GSI

system in *Arabidopsis thaliana* that SC is caused by a mutation in a gene (PUB8) encoding a U-box-containing protein that is linked to the S-gene and which regulates SRK transcript levels (Liu et al., 2007). In contrast, other SC systems are based on modifier genes located outside SI loci and suppressing specific gene functions (Nasrallah, 2004; Nasrallah et al., 2002). In
5 grasses, the involvement of additional, possibly equivalent loci independent of S and Z has been revealed through studies on the breakdown of SI in at least three grass species; for example a SC locus called T was identified in perennial ryegrass (Thorogood and Hayward, 1991; Thorogood et al., 2005) and in Blue canary grass (Hayman and Richter, 1992), and the S5 locus was found in rye (Voylokov et al., 1993). We have sequenced LpGK1 and LpGK2 in a
10 self-compatible perennial ryegrass genotypes that is homozygous for the haplotype P226 at the Z locus genes. We found a 2 bp insertion (CT) in the VR of LpGK1 (Figure 10). This insertion is causing a frame shift in the amino acid sequence, finally leading to a stop codon and termination of the protein translation process (Figure 11). Notably, this haplotype P226 that is derived from a self-compatible perennial ryegrass genotype contain different amino acid
15 sequence motifs in the VR of LpGK1 and LpGK2 (Figure 12). This in contrast to sequence information of haplotype P205 (Figure 5) and has not been observed in any other haplotypes coding for functional Z SI locus genes.

In order to confirm our findings in other Poaceae crops, we have sequenced LpGK1 and LpGK2 in self compatible Tibetan rye (*Secale sylvestre* Host), and a self-incompatible genotype
20 of each of the hybrid rye cultivars Askari and Picasso. Notably, we found a retrotransposon element that was inserted in LpGK1 of Tibetan rye, affecting the gene function and leading to its SC phenotype.

A more detailed structural analysis of the FGGY family of carbohydrate kinases revealed that the monomers contain two large domains, which are separated by a deep cleft forming the
25 active site. In addition, LpGK1 and LpGK2 also contain a putative homodimer and homotetramer interface (polypeptide binding site), further supporting that they may interact either with each other or with the S locus protein. This will advance the identification of the S component as both S and Z possibly interact to induce a successful SI reaction. For example, a yeast two-hybrid system can be used to discover protein-protein (or protein-DNA) interactions
30 between the Z and the S SI component by testing for physical interaction (e.g. binding) between the LpGK1 and LpGK2 (or any heteropolymers of these two genes) with a protein or a DNA molecule that is representing S. Similarly, an antibody that is specific for the Z locus gene can be used to isolate the S component out of a pistil and/or pollen protein solution by complex immunoprecipitation (Co-IP). Immunoprecipitation of intact protein complexes (i.e. an antigen
35 along with any proteins or ligands that are bound to it) works by designing an antibody that targets the Z locus gene which is believed to be a member of a larger complex of proteins involving the S component. By targeting Z with an antibody it may become possible to pull the

entire protein complex out of solution and thereby identify the S component as well as other unknown members of this complex. Isolated polypeptides may be subject to amino acid sequencing to obtain amino acid sequence information of S and other components involved in SI.

5 Interestingly, we found Arabidopsis NHO1 (also known as NONHOST1, or non-host resistant 1) that share strong homologies with the glycerol kinases LpGK1 and LpGK2 that are representing the Z locus genes. Nonhost resistance of plants refers to the phenomenon observed when all members of a plant species are typically resistant to a specific parasite. NHO1, a single copy gene in Arabidopsis, is required for nonspecific resistance to nonhost
10 Pseudomonas bacteria and to the fungal pathogen *Botrytis cinerea*, indicating that NHO1 is not limited to bacterial or fungal resistance mechanisms, and that glycerol kinase-based cell–cell recognition is widespread in nature.

Parallels between self-nonsel and host-pathogen recognition systems have intrigued biologists for decades. NHO1 knockout mutant in Arabidopsis shows a normal phenotype, and
15 we speculate that LpGK1 represents a suitable target for a transgenic approach as a molecular switch to control SI by inducible promoters.

Origin of Plant Material

The perennial ryegrass mapping population used for mapping and sequencing work was derived from a cross between a plant of the Italian cultivar Veyo“ and a Danish ecotype
20 collected on Falster, Denmark. Two selected F1 genotypes (referred to as F1_30 and F1_39) were then used to develop a F2 population consisting of more than 6,000 F2 genotypes.

References

All documents cited below, and anywhere in this specification, are incorporated by reference.

Baumann U., Juttner J., Bian X., Langridge P. (2000) Self-incompatibility in the grasses. *Annals of Botany* 85:203-209.

- 5 Bosch M., Franklin-Tong V.E. (2007) Temporal and spatial activation of caspase-like enzymes induced by self-incompatibility in *Papaver* pollen. *Proc. Natl Acad. Sci. USA* 104:18327-18332.

- Cheng J.H., Han Z.H., Xu X.F., Li T.Z. (2006) Isolation and identification of the pollen-expressed polymorphic F-box genes linked to the S-locus in apple (*Malus Xdomestica*). *Sex. Plant*
10 *Reprod.* 19:175-183.

de Graaf B.H.J. (2006) Self-incompatibility in *Papaver* targets soluble inorganic pyrophosphatases in pollen. *Nature* 444:490-493.

de Nettancourt D. (1977) Incompatibility in angiosperms, Springer, New York, NY.

- Duvick D.N. (2005) The contribution of breeding to yield advances in maize (*Zea mays* L.)
15 Elsevier Academic Press, San Diego, CA.

East E.M. (1940) The distribution of self-sterility in the flowering plants. *Proc. Am. Philos. Soc.* 82:449-518.

- Entani T., Iwano M., Shiba H., Che F.S., Isogai A., Takayama S. (2003) Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-
20 expressed F-box gene with allelic diversity. *Genes Cells* 8:203-213.

Franklin-Tong V.E., Franklin F.C.H. (1992) Gametophytic self-incompatibility in *Papaver rhoeas* L. *Sex. Plant Reprod.* 5:1-7.

Geiger H.H., Miedaner T (2009) Rye breeding. In *Cereals*. Edited by Carena MJ. Springer, New York

- 25 Hackauf B., Wehling P. (2005) Approaching the self-incompatibility locus Z in rye (*Secale cereale* L.) via comparative genetics.

Hayman D.L., Richter J. (1992) Mutations affecting self-incompatibility in *Phalaris coerulescens* Desf. (*Poaceae*). *Heredity* 68:495–503.

- Kakeda K. (2009) S locus-linked F-box genes expressed in anthers of *Hordeum bulbosum*. *Plant Cell Rep.*
30 28:1453-1460.

Klaas M., Yang B., Bosch M., Thorogood D., Manzanares C., Armstead I.P., Franklin F.C.H., Barth S. (2011) Progress towards elucidating the mechanisms of self-incompatibility in the grasses: further insights from studies in *Lolium*. *Annals of Botany*:in press.

Krol et al., (1990) *The Plant Cell* 2, 291 229

- Lamkey K.R., Edwards J.W. (1999) Quantitative genetics of heterosis, in: J. G. Coors and S. Pandey (Eds.), The genetics and exploitation of heterosis in crops, ASA, CSSA, and SSSA, Madison, WI. pp. 31-58.
- Lee H.S., Huang S.S., Kao T.H. (1994) S-proteins control rejection of incompatible pollen in
 5 *Petunia inflata*. *Nature* 367:560-563.
- Li X., Nield J., Hayman D., Langridge P. (1994) Cloning a putative self-incompatibility gene from the pollen of the grass *Phalaris coerulescens*. *Plant Cell* 6:1923-1932.
- Li X., Paech N., Nield J., Hayman D., Langridge P. (1997) Self-incompatibility in the grasses: evolutionary relationship of the S gene from *Phalaris coerulescens* to homologous
 10 sequences in other grasses. *Plant Mol. Biol.* 34:223-232.
- Li J.M., Yuan L.P. (2000) Hybrid rice, genetics, breeding, and seed production, in: J. Janick (Ed.), *Plant Breed. Rev.*, John Wiley & Sons, New York. pp. 15-158.
- Liu P., Sherman-Broyles S., Nasrallah M.E., Nasrallah J.B. (2007) A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. *Curr. Biol.* 17:734-740.
- 15 Lundqvist A. (1954) Studies on self-sterility in rye, *Secale cereale* L. *Hereditas* 40:278–294.
- Manzanares C., Studer B., Hayes R., Thorogood D. (2011) Population genetics of the grass self-incompatibility system – Practical implications for grass breeding programmes, in: S. Barth (Ed.), *EUCARPIA - 29th Fodder Crops and Amenity Grass Section Meeting*, Dublin, Ireland.
- 20 Murfett J., Atherton T.L., Mou B., Gasser C.S., McClure B.A. (1994) S-RNase expressed in transgenic *Nicotiana* causes S-allele-specific pollen rejection. *Nature* 367:563-566.
- Napoli et al., (1990) *The Plant Cell* 2, 279-289
- Nasrallah J.B. (2004) Recognition and rejection of self in plant reproduction. *Mol. Biol. Cell* 15:236A-236A.
- 25 Nasrallah M.E., Liu P., Nasrallah J.B. (2002) Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science* 297:247-249.
- Pandey K.K. (1977) Origin of complementary incompatibility systems in flowering plants. *Theor. Appl. Genet.* 49:101-109.
- Posselt U.K. (2010) Identification of heterotic patterns in perennial ryegrass, in: C. Huyghe (Ed.),
 30 *Sustainable Use of Genetic Diversity in Forage and Turf Breeding*, Springer Netherlands. pp. 569-572.
- Schopfer C.R., Nasrallah M.E., Nasrallah J.B. (1999) The male determinant of self-incompatibility in *Brassica*. *Science* 286:1697-1700.
- Shinozuka H., Cogan N., Smith K., Spangenberg G., Forster J. (2009) Fine-scale comparative
 35 genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* L.). *Plant Mol. Biol.* 72:343-355.

- Sijacic P., Wang X., Skirpan A.L., Wang Y., Dowd P.E., McCubbin A.G., Huang S., Kao T.H. (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* 429:302-305.
- Snowman B.N., Kovar D.R., Shevchenko G., Franklin-Tong V.E., Staiger C.J. (2002) Signal-mediated depolymerization of actin in pollen during the self-incompatibility response. *Plant Cell* 14:2613-2626.
- Studer B., et al. (2008) Genetic characterisation of seed yield and fertility traits in perennial ryegrass (*Lolium perenne* L.), *Theor. Appl. Genet*, 117:781-791
- Studer B., Jensen L.B., Fiil A., Asp T. (2009) "Blind" mapping of genic DNA sequence polymorphisms in *Lolium perenne* L. by high resolution melting curve analysis. *Mol. Breed.* 24:191-199.
- Studer B., Asp T., Thorogood D., Barth S., Armstead I., Yang B., Lübberstedt T. (2010a) Map-based cloning of the Z self-incompatibility locus in perennial ryegrass (*Lolium perenne* L.), The 6th International Symposium on the Molecular Breeding of Forage and Turf, Ediciones INTA, Buenos Aires, Argentina. pp. 35.
- Takasaki T., Hatakeyama K., Suzuki G., Watanabe M., Isogai A., Hinata K. (2000) The S receptor kinase determines self-incompatibility in *Brassica stigma*. *Nature* 403:913-916.
- Thomas S.G., Franklin-Tong V.E. (2004) Self-incompatibility triggers programmed cell death in *Papaver* pollen. *Nature* 429:305-309.
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
- Thorogood D., Hayward M.D. (1991) The genetic control of self-compatibility in an inbred line of *Lolium perenne* L. *Heredity* 67:175-181.
- Thorogood D., Kaiser W.J., Jones J.G., Armstead I. (2002) Self-incompatibility in ryegrass 12. Genotyping and mapping the S and Z loci of *Lolium perenne* L. *Heredity* 88:385-390.
- Thorogood D., Armstead I.P., Turner L.B., Humphreys M.O., Hayward M.D. (2005) Identification and mode of action of self-compatibility loci in *Lolium perenne* L. *Heredity* 94:356-363.
- Ushijima K., Sassa H., Dandekar A.M., Gradziel T.M., Tao R., Hirano H. (2003) Structural and transcriptional analysis of the self-incompatibility locus of almond: Identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell* 15:771-781
- Van Daele I., Van Bockstaele E., Martens C., Roldán-Ruiz I. (2008b) Identification of transcribed derived fragments involved in self-incompatibility in perennial ryegrass (*Lolium perenne* L.) using cDNA-AFLP. *Euphytica* 163:67-80.

- Van Daele I., Muylle H., Van Bockstaele E., Roldán-Ruiz I. (2008a) Mapping of markers related to self-incompatibility, disease resistance, and quality traits in *Lolium perenne* L. Genome 51:644-656.
- 5 Voylokov A.V., Fuong F.T., Smirnov V.G. (1993) Genetic studies of self-fertility in rye (*Secale cereale* L.). 1. The identification of genotypes of self-fertile lines for the Sf alleles of self-incompatibility genes. Theor. Appl. Genet. 87:616-618.
- Wang Y., Xue Y., Li J. (2005) Towards molecular breeding and improvement of rice in China. Trends Plant Sci. 10:610-614.
- 10 Wheeler M.J., de Graaf B.H.J., Hadjosif N., Perry R.M., Poulter N.S., Osman K., Vatovec S., Harper A., Franklin F.C.H., Franklin-Tong V.E. (2009) Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. Nature 459:992-995.
- Yang B., Thorogood D., Armstead I., Barth S. (2008) How far are we from unravelling self-incompatibility in grasses? New Phytol. 178:740-753.
- 15 Yang B., Thorogood D., Armstead I., Franklin F., Barth S. (2009) Identification of genes expressed during the self-incompatibility response in perennial ryegrass (*Lolium perenne* L.). Plant Mol. Biol. 70:709-723.
- Zhang et al, 1992 The Plant Cell 4, 1575 1588.

What is claimed is:

1. Isolated nucleic acid comprising a nucleotide sequence of a Z locus Poaceae gene.
2. Isolated nucleic acid according to claim 1, wherein the nucleotide sequence is:
 - (i) the LpGK1 allele of perennial ryegrass Z locus haplotype P205 shown in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 8 or SEQ ID NO: 23;
 - (ii) the LpGK1 allele of perennial ryegrass Z locus haplotype B724 shown in SEQ ID NO: 6 and SEQ ID NO: 9;
 - (iii) the LpGK2 allele of perennial ryegrass Z locus haplotype P226 shown in SEQ ID NO: 10;
 - (iv) the LpGK2 allele of perennial ryegrass Z locus haplotype P205 shown in SEQ ID NO: 1 or SEQ ID NO: 12;
 - (v) the LpGK2 allele of barley shown in SEQ ID NO: 14;
 - (vi) the LpGK2 allele of Brachypodium shown in SEQ ID NO: 15;
 - (vii) the LpGK2 allele of rice shown in SEQ ID NO: 16;
 - (viii) the LpGK2 allele of sorghum shown in SEQ ID NO: 17;
 - (ix) the LpGK2 allele of maize shown in SEQ ID NO: 18;
 - (x) the LpGK1 allele of perennial ryegrass Z locus haplotype P226 shown in SEQ ID NO: 24;
 - (xi) the LpGK2 allele of perennial ryegrass Z locus haplotype S089 shown in SEQ ID NO: 27;
 - (xii) the LpGK2 allele of perennial ryegrass Z locus haplotype S065 shown in SEQ ID NO: 29;
 - (xiii) the LpGK2 allele of perennial ryegrass Z locus haplotype S027 shown in SEQ ID NO: 31;
 - (xiv) the LpGK2 allele of perennial ryegrass Z locus haplotype S021 shown in SEQ ID NO: 33;
 - (xv) a nucleotide sequence that is at least 90 % identical with any of (i) to (xiv);
 - (xvi) a nucleotide sequence that encodes a polypeptide that has at least 90 % amino acid sequence identity with the polypeptide encoded by any of (i) to (xiv); or
 - (xvii) a nucleotide sequence that encodes a polypeptide sequence encoded by any of (i) to (xiv).
2. Isolated nucleic acid comprising a Z locus gene, comprising a VR flanked by an upstream conserved region and a downstream conserved region, wherein the VR and

conserved regions share at least 90 % identity with the corresponding VR and conserved regions of a reference sequence selected from any of (i) to (xiv).

3. Isolated nucleic acid according to claim 2, wherein the nucleic acid comprises a functional Z locus gene that interacts with another functional Z locus gene in a plant to confer a self-incompatibility phenotype on the plant.
4. Nucleic acid according to claim 2 or claim 3, wherein the gene comprises the VR and conserved regions of the reference sequence.
5. Nucleic acid according to any of claims 2 to 4, wherein the nucleotide sequence is at least 70 % identical with the reference sequence.
6. Nucleic acid according to any of claims 1 to 5, comprising a coding sequence operably linked to an inducible promoter.
7. A vector comprising nucleic acid according to any of the preceding claims.
8. A polypeptide encoded by nucleic acid according to any of claims 1 to 5.
9. A polypeptide comprising the amino acid sequence of any of SEQ ID NOS: 3, 5, 7, 11, 13, 19-22, 25-26, 28, 30, 32 or 34.
10. A primer or probe that specifically hybridises to a Z locus gene according to any of claims 1 to 5.
11. A primer or probe according to claim 10, comprising an oligonucleotide sequence that specifically hybridises to a conserved region of a Z locus gene according to any of claims 1 to 5.
12. A primer or probe according to claim 11, comprising an oligonucleotide sequence that specifically hybridises to a variable region of a Z locus gene according to any of claims 1 to 5.
13. Use of a primer or probe according to any of claims 10 to 12 for identifying a Z locus gene in a plant of the grass family.
14. A method of identifying a Z locus gene in a plant of the grass family, comprising providing a preparation of nucleic acid from the plant;

providing a probe that hybridises to a conserved region of the Poaceae Z locus;
contacting the nucleic acid in the preparation with the probe under conditions for hybridisation; and

identifying a Z locus gene if present by its hybridisation with the probe.

15. A method of identifying a Z locus gene in a plant of the grass family, comprising
providing a preparation of nucleic acid from the plant;
providing a pair of primers for hybridisation to a conserved region of the Poaceae Z locus, suitable for amplifying the conserved region;

contacting the nucleic acid in the preparation with the primers under conditions for hybridisation and amplification; and

determining the presence or absence of an amplification product, where the presence of an amplification product indicates the presence of a Z locus gene in the plant.

16. A method according to claim 14 or claim 15, further comprising determining the sequence of the Z locus gene in the plant.

17. A method of determining the identity of a Z locus gene alleles in a plant of the grass family, comprising

providing a preparation of nucleic acid from the plant;

providing a probe or primer that hybridises to the Z locus;

annealing the probe or primer to the Z locus in the nucleic acid from the plant; and

identifying the Z locus gene present at the Z locus by detecting hybridisation and/or by obtaining an amplification product.

18. A method according to claim 17, comprising

providing a probe that hybridises to a variable region of a Z locus gene at the Poaceae Z locus;

contacting the nucleic acid in the preparation with the probe under conditions for hybridisation; and

identifying the Z locus gene if present by its hybridisation with the probe.

19. A method according to claim 17, comprising

providing a pair of primers for hybridisation to a variable region of a Z locus gene at the Poaceae Z locus;

contacting the nucleic acid in the preparation with the primers under conditions for hybridisation and amplification; and

determining the presence or absence of an amplification product, where the presence of an amplification product indicates the presence of the Z locus gene.

20. A method according to any of claims 14 to 19, further comprising determining the haplotype of the plant at the Z locus.
21. A method of predicting the ability of a plant of the grass family to self-fertilise, comprising determining the identity of Z locus genes in the plant; and determining whether a self-incompatible combination of a Z locus genes is present.
22. A method according to claim 21, comprising determining possible combinations of pollen Z locus genes and stigma Z locus genes; and identifying whether each possible combination of pollen and stigma Z locus genes includes an incompatible combination of Z locus genes, wherein an incompatible combination of Z locus genes occurs when the VR of a Z locus gene in the pollen matches the VR of a Z locus gene in the stigma; wherein presence of an incompatible combination of Z locus genes indicates that the pollen is unable to fertilise the stigma, and absence of an incompatible combination of Z locus genes indicates that the pollen is able to fertilise the stigma.
23. A method of predicting the ability of a first plant to fertilise a second plant, comprising determining the identity of Z locus genes in the first plant and in the second plant; determining possible combinations of pollen Z locus genes in the first plant with stigma Z locus genes in the second plant; and identifying whether each possible combination of pollen and stigma Z locus genes includes an incompatible combination of Z locus genes, wherein an incompatible combination of Z locus genes occurs when the VR of a Z locus gene in the pollen matches the VR of a Z locus gene in the stigma; wherein presence of an incompatible combination of Z locus genes indicates that the pollen is unable to fertilise the stigma, and absence of an incompatible combination of Z locus genes indicates that the pollen is able to fertilise the stigma.
24. A method according to claim 23, comprising determining that all combinations of pollen and stigma Z locus genes include an incompatible combination, and thereby predicting a full incompatibility phenotype wherein the first plant is unable to fertilise the second plant.

25. A method according to claim 23, comprising determining that one combination of pollen and stigma Z locus genes includes an incompatible combination, and one combination of pollen and stigma Z locus genes does not include an incompatible combination, and thereby predicting a partial compatibility phenotype wherein only a proportion of the pollen of the first plant is able to fertilise the second plant.

26. A method according to claim 23, comprising determining that no combination of pollen and stigma Z locus genes includes an incompatible combination, and thereby predicting a full compatibility phenotype wherein the first plant is able to fertilise the second plant.

27. A method according to any of claims 23 to 26, wherein an incompatible combination of Z locus genes occurs when the VRs of LpGK1 and LpGK2 in pollen are matching VRs and also match the VRs of an LpGK1 and an LpGK2 in the stigma.

28. A method according to claim 27, wherein the first and second plants are diploid, and the method comprises

- determining the identity of Z locus genes in the first plant and in the second plant;
- determining possible combinations of pollen Z locus genes in the first plant with stigma Z locus genes in the second plant; and

- identifying whether each possible combination of pollen and stigma Z locus genes includes an incompatible combination of Z locus genes, where an incompatible combination occurs when a Z locus in the first plant comprises LpGK1 and LpGK2 with matching VRs, which also match the VRs of LpGK1 and LpGK2 in at least one Z locus in the second plant; wherein

- presence of an incompatible combination of Z locus genes indicates that the pollen is unable to fertilise the stigma, and absence of an incompatible combination of Z locus genes indicates that the pollen is able to fertilise the stigma, and wherein

- if both combinations of pollen and stigma Z locus genes include an incompatible combination, the first plant is predicted to be unable to fertilise the second plant; and/or

- if one combination of pollen and stigma Z locus genes includes an incompatible combination, and the other combination of pollen and stigma Z locus genes does not include an incompatible combination, the first plant is predicted to be partially compatible with the second plant, wherein half the pollen of the first plant is able to fertilise the second plant; and/or

- if neither combination of pollen and stigma Z locus genes includes an incompatible combination, the first plant is predicted to be able to fertilise the second plant.

29. A method according to claim 28, where the incompatible combination occurs when a pollen Z locus haplotype is the same as a stigma Z locus haplotype.

30. A method of altering self-fertility in a plant of the grass family, comprising modulating expression of one or more Z locus genes in the plant, modifying the genotype of the plant at the Z locus and/or altering activity of one or more Z locus gene products.
31. A method according to claim 30, comprising providing the one or more Z locus genes under control of an inducible promoter.
32. A method according to claim 30, comprising increasing self-compatibility in a self-incompatible plant of the grass family, comprising
down-regulating expression of one or more Z locus genes in the plant; and/or
mutating nucleic acid at the Z locus to inhibit expression of one or more Z locus genes.
33. A method according to claim 30, comprising reducing self-compatibility in a self-compatible plant of the grass family, comprising
up-regulating expression of one or more Z locus genes in the plant;
mutating the sequence of one or more Z locus genes in a plant; and/or
introducing one or more Z locus genes into the plant.
34. A method according to claim 30, comprising providing or generating a transgenic plant in which expression of one or more Z locus genes is under control of an inducible promoter.
35. A method according to any of claims 30 to 34, wherein the one or more Z locus genes is or are one or more Z locus genes according to any of claims 1 to 5.
36. A method according to any of claims 30 to 35, wherein the one or more Z locus genes comprise an LpGK1 gene.
37. A method according to claim 36, wherein the one or more Z locus genes comprise LpGK1 and LpGK2.
38. A method of self-fertilising a plant of an allogamous Poaceae species, comprising increasing self-compatibility of the plant using the method of claim 32, allowing or promoting self-pollination of the plant, and obtaining seed.
39. A method according to claim 38, further comprising growing the seed to generate progeny, and performing further steps of crossing and selection to obtain an inbred line.

40. A method according to claim 38 or claim 39, further comprising reducing self-compatibility in a plant of the inbred line using a method according to claim 33, and outcrossing the inbred plant with a second plant, and generating hybrid seed.
41. A method according to claim 40, further comprising growing the hybrid seed to generate hybrid progeny, and performing further steps of crossing and selection to generate hybrids.
42. A method according to claim 41, further comprising obtaining seed from the hybrids.
43. Seeds or plants, or parts of plants, generated by a method according to any of claims 38 to 42.
44. A transgenic plant comprising in its genome a Z locus gene operably linked to an inducible promoter.
45. A method of screening for an agent that influences a compatibility phenotype of a plant of the grass family, comprising
bringing a candidate compound into contact with a polypeptide encoded by a Z locus Poaceae gene; and
determining whether the compound binds the polypeptide and/or affects activity of the polypeptide;
where detecting binding and/or an effect on activity indicates that the compound may influence a compatibility phenotype of a plant.
46. Use of a kinase inhibitor for inhibiting self-incompatibility in a grass family plant.
47. A nucleic acid, polypeptide, method, use or plant substantially as described herein, with reference to and/or as illustrated in any of the drawings.

LpGK1_CDS_P205	ATGGAGAATGTTGATGCTGTCAAGGATGCAGTCCGGGCTGGTGATGCATTATTTGGTACA	60
LpGK1_CDS_B724	ATGGAGAATGTTAACTCTGTCAAGGATGCAGTTCGGGCTGGTGATGCATTATTTGGCACA	60
	***** * *****	
LpGK1_CDS_P205	ATTGACACCTGGTTGATCTGGAACCTCACAGGAGGCATTGGTGCCAAAGACACTGATGGA	120
LpGK1_CDS_B724	ATCGACACCTGGTTGATCTGGAACCTCACAGGAGGTGTTGACGGGAA-----	107
	** ***** *** ** *	
LpGK1_CDS_P205	AAAGAATTACTTGGCCAACATGTCACAGACTGCTCAAATGCATCTCGCACGATGCTTATG	180
LpGK1_CDS_B724	-----TGGGCAGCATGTCACAGACTGCTCAAATGCATCTCGCACGATGCTTATG	156
	*** ** *****	
LpGK1_CDS_P205	AATCTAAAGGCACTTGATTGGAACAAGCCTACACTTGAGGCATTAGGCATTCCTGCTGGC	240
LpGK1_CDS_B724	AATCTAAAGGCACTTGATTGGGACAAGCCTACACTTGAGGCATTAGGCATTCCTGCTGGC	216
	***** *****	
LpGK1_CDS_P205	ATCTTGCCGAAGATTATCAGCAACTCAGAGAAAATTGGCGTGGTTGCCAGTGGGTTTCCT	300
LpGK1_CDS_B724	ATCTTGCCGAAGATTATCAGCAACTCGGAAAAAATTGGCGTGGTTGCCAGTGGGTTCCCT	276
	***** ** *****	
LpGK1_CDS_P205	TTGGCTGGCGTCTCCATCTCAGGGTGTCTAGGAGACCAGCATGCTGCAATGCTCGGGCAG	360
LpGK1_CDS_B724	TTGGCTGGCGTCTCCATCTCAGGGTGTCTAGGAGACCAGCATGCTGCAATGCTCGGGCAG	336

LpGK1_CDS_P205	CTATGCCAGAAAGGTGAAGCCAAAAGCACCTATGGAAC TGGTGCCTTCATCCTTCTGAAC	420
LpGK1_CDS_B724	CTATGCCAGAAAGGTGAAGCCAAAAGCACCTACGGAAC TGGTGCCTTCATCCTTCTCAAC	396
	***** *****	
LpGK1_CDS_P205	ACAGGGGAGGAGGTTACCCAGTCCTCCCACGGTCTTCTTAGCACTATCGCTTACAAGCTT	480
LpGK1_CDS_B724	ACAGGGGAGGAGGTTACCCAGTCCTCCCATGGTCTTCTTAGCACCATTGCTTACAAGCTT	456
	***** ** *****	
LpGK1_CDS_P205	GGCCCAGATGTACCCACCAATTATGCTCTAGAAGGCTCAATTGCAATTGCTGGTGCAGCA	540
LpGK1_CDS_B724	GGCCCAGATGTACCAACCAATTACGCTCTAGAAGGCTCAATTGCTATTGCTGGTGCAGCA	516
	***** *****	
LpGK1_CDS_P205	GTGCAGTGGTTGAGGGACAGCCTTGGGATCATTAGCTCAGCAGCAGAAATTGAAGGGTTG	600
LpGK1_CDS_B724	GTTCAGTGGTTGAGGGACAGCCTTGGGATCATTAGCTCAGCAGCAGAAATTGAAGGGTTG	576
	** *****	
LpGK1_CDS_P205	GCTGAAAGTGTGCAGGATT CAGGTGGSATATACTTTGTGCCAGCATTCAATGGATTGTTT	660
LpGK1_CDS_B724	GCTGAAAGTGTGCAGGATT CAGGTGGGATATACTTTGTCCCGGCATTCAATGGATTGTTT	636
	***** ** *****	
LpGK1_CDS_P205	GCACCATGGTGGAGAGATGATGCAAGAGGGATCTGCATAGGAATCACAAGGTT CACAAAT	720
LpGK1_CDS_B724	GCACCATGGTGGAGGGATGATGCAAGAGGTATCTGCATAGGAATCACAAGGTT CACAAAT	696
	***** *****	
LpGK1_CDS_P205	AAGGGGCACATTGCTCGGGCAGTGCTTGAGAGTATGTGCTTT CAGGTGAATGATGTCCTC	780
LpGK1_CDS_B724	AAGGGGCACATTGCTCGAGCAGTGCTAGAGAGCATGTGCTTT CAGGTGAATGATGTCCTT	756
	***** *****	
LpGK1_CDS_P205	AGCTCCATGCATAAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCAGAAGGGGAGTTC	840
LpGK1_CDS_B724	AGCTCCATGCATAAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCAGAAGGGGAGTTC	816

Figure 1

LpGK1_AS_P205	MENVDAVKDAVRAGDALFGTIDTWLIWNLTGGIGGKDTDGKELLGQHVTDCSNASRTMLM	60
LpGK1_AS_B724	MENVNSVKDAVRAGDALFGTIDTWLIWNLTGGVDGN-----GQHVTDCSNASRTMLM	52
	****:;*****:.*:*****	
LpGK1_AS_P205	NLKALDWNKPTLEALGIPAGILPKIISNSEKIGVVASGFPLAGVSI SGCLGDQHAAMLGQ	120
LpGK1_AS_B724	NLKALDWDKPTLEALGIPAGILPKIISNSEKIGVVASGFPLAGVSI SGCLGDQHAAMLGQ	112
	*****.******	
LpGK1_AS_P205	LCQKGAEAKSTYGTGAFILLNTGEEVTQSSHGLLSTIAYKLGPDPVTNYALEGSIAIAGAA	180
LpGK1_AS_B724	LCQKGAEAKSTYGTGAFILLNTGEEVTQSSHGLLSTIAYKLGPDPVTNYALEGSIAIAGAA	172

LpGK1_AS_P205	VQWLRDSLGIISAAEIEGLAESVQDSGXIIYFVPAFNGLFAPWWRDDARGICIGITRFTN	240
LpGK1_AS_B724	VQWLRDSLGIISAAEIEGLAESVQDSGGIIYFVPAFNGLFAPWWRDDARGICIGITRFTN	232
	***** *****	
LpGK1_AS_P205	KGHIARAVLESMCFQVNDVLSSMHKDAGEAGEVKSAEGEFLLRVDGGATVNNLLMQIQAD	300
LpGK1_AS_B724	KGHIARAVLESMCFQ-----	247

LpGK1_AS_P205	LLGSPVVRPADIETALGAAYAAGLAAGVWTKDEIFAGLHKENTTVFRPKLDDAHRKKRG	360
LpGK1_AS_B724	-----	
LpGK1_AS_P205	DSWYKAVSRSFDLADLSL	378
LpGK1_AS_B724	-----	

Figure 2

CLUSTAL 2.1 multiple sequence alignment

```

LpGK2_P205_cds      ATGGCGGGGACGG-----AGGAGGTGTACGTGGCGTCGATCGACCAGGGCACCACC 51
LpGK2_P226_cds      ATGGCGGGGACGGCGGCGGAGGAGGAGGTGTACGTGGCGTCGATCGACCAGGGCACCACC 60
                      *****
                      *****

LpGK2_P205_cds      AGCACCAGGTTTCATCGTCTACGACCGCCACGCCAGGCCCGTCGCCTCGCACCAGCTCGAG 111
LpGK2_P226_cds      AGCACCAGGTTTCATCGTCTACGACCGCCACGCCAGGCCCGTCGCCTCGCACCAGCTCGAG 120
                      *****

LpGK2_P205_cds      TTCAAGCAGCACTACCCGGAGGCAGGATGGGTCGAGCATGATCCTATGGAGATTATAGAG 171
LpGK2_P226_cds      TTCAAGCAGCACTACCCGGAGGCAGGATGGGTCGAGCATGATCCTATGGAGATTATAGAG 180
                      *****

LpGK2_P205_cds      AGCGTGAAGATTTGCATGTCAGAGACGCTTGGCAAAGCTGCAGCTAGTGGACTTAATGTG 231
LpGK2_P226_cds      AGCGTGAAGATTTGCATGTCAGAGACACTTGGCAAAGCTGCAGCTAGTGGACTTAATGTG 240
                      *****

LpGK2_P205_cds      GATGCTGGTTTGAAGGCCATTGGGATCACAAATCAGAGGGAGACAACAGTTATGTGGAGC 291
LpGK2_P226_cds      GATGCTGGTTTGAAGGCCATTGGGATCACAAATCAGAGGGAGACAACAGTTATGTGGAGC 300
                      *****

LpGK2_P205_cds      AAATCCACAGGTCTTCCACTATAACAATGCCATAGTGTGGATGGATGTTTCGCACAAGCTCT 351
LpGK2_P226_cds      AAATCCACAGGTCTTCCACTATAACAATGCCATAGTGTGGATGGATGTTTCGCACAAGCTCC 360
                      *****

LpGK2_P205_cds      ATTTGCAGGAGACTGGAAAGTGAGCTATCAGGTGGTAGAACACACTTTGTGGAGACGTGT 411
LpGK2_P226_cds      ATTTGCAGGAGATTGGAAAGTGAGCTATCAGGTGGCAGAACACACTTTGTGGAGACATGT 420
                      *****

LpGK2_P205_cds      GGCCTACCAATCAGTACCTATTTTCAGTGCTCTTAAGTTATTATGGTTGATGGAGAATGTT 471
LpGK2_P226_cds      GGCTTACCAATCAGCACATATTTTCAGTGCTCTCAAGTTATTGTGGTTGATGGAGAATGTT 480
                      *** *****

LpGK2_P205_cds      GACGCTGTCAAGGATGCAGTCAAGGCTGGTGATGCATTATTTGGCACAATCGACACCTGG 531
LpGK2_P226_cds      AACTCTGTCAAGGATGCAGTTCGGGCTGGTGATGCATTATTTGGCACAATCGACACCTGG 540
                      ** *****

LpGK2_P205_cds      TTGATCTGGAACCTCACAGGAGGCATTGGTGGGAAAGATACTGATGGAAAAGAATTACTT 591
LpGK2_P226_cds      TTGATCTGGAACCTCACAGGAGGTGTTGACGGGAA-----T 576
                      *****

LpGK2_P205_cds      GGGCAGCATGTCACAGACTGCTCAAATGCATCTCGCACGATGCTTATGAATCTAAAGGCA 651
LpGK2_P226_cds      GGGCAGCATGTCACAGACTGCTCAAATGCATCTCGCACGATGCTTATGAATCTAAAGGCA 636
                      *****

LpGK2_P205_cds      CTTGATTGGGACAAGCCTACACTTGAGGCATTAGGCATTCTGTTGGCATCTTGCCGAAG 711
LpGK2_P226_cds      CTTGATTGGGACAAGCCTACACTTGAGGCATTAGGCATTCTGCTGGCATCTTGCCGAAG 696
                      *****

LpGK2_P205_cds      ATTATCAGCAACTCGGAGAAAATTGGCGTGGTTGCCAGTGGGTTCCCTTTGGCTGGCGTC 771
LpGK2_P226_cds      ATTATCAGCAACTCGGAAAAAATTGGCGTGGTTGCCAGTGGGTTCCCTTTGGCTGGCGTC 756
                      *****

LpGK2_P205_cds      TCCATCTCAGGGTGTTTtaggagaccagcatgctgcaatgctcggacagctatgccagaaa 831
LpGK2_P226_cds      TCCATCTCAGGGTGTTtaggagaccagcatgctgcaatgctcgggcagctatgccagaaa 816
                      *****

```

Figure 3

LpGK2_P205_cds	GGTGAAGCCAAAAGCACCTACGGAACCTGGTGCCTTCATCCTTCTCAACACAGGGGAGGAG	891
LpGK2_P226_cds	GGTGAAGCCAAAAGCACCTACGGAACCTGGTGCCTTCATCCTTCTCAACACAGGGGAGGAG	876

LpGK2_P205_cds	GTTACCCAGTCCTCCCATGGTCTTCTTAGCACCATTGCTTACAAGCTTGGCCCAGATGTA	951
LpGK2_P226_cds	GTTACCCAGTCCTCCCATGGTCTTCTTAGCACCATTGCTTACAAGCTTGGCCCAGATGTA	936

LpGK2_P205_cds	CCCACCAATTATGCTCTAGAAGGCTCAATTGCTATTGCTGGTGCAGCAGTTCAGTGGTTG	1011
LpGK2_P226_cds	CCAACCAATTACGCTCTAGAAGGCTCAATTGCTATTGCTGGTGCAGCAGTTCAGTGGTTG	996
	** *****	
LpGK2_P205_cds	AGGGACAGCCTTGGGATCATTAGCTCAGCAGCAGAAATTGAAGGGTTGGCTGAAAGTGTG	1071
LpGK2_P226_cds	AGGGACAGCCTTGGGATCATTAGCTCAGCAGCAGAAATTGAAGGGTTGGCTGAAAGTGTG	1056

LpGK2_P205_cds	CAAGATTTCAGGCGGGATATACTTTGTGCCAGCATTCAATGGATTGTTGCGACCATGGTGG	1131
LpGK2_P226_cds	CAGGATTTCAGGTGGGATATACTTTGTCCCGGCATTCAATGGATTGTTGCGACCATGGTGG	1116
	** ***** ** *****	
LpGK2_P205_cds	AGGGATGACGCGAGAGGTATCTGCATAGGAATCACGAGGTTTACAAACAAGGGGCACATT	1191
LpGK2_P226_cds	AGGGATGATGCAAGAGGTATCTGCATAGGAATCACAGGTTTACAAATAAGGGGCACATT	1176
	***** ** *****	
LpGK2_P205_cds	GCTCGAGCAGTGCTTGAGAGTATGTGCTTTCAAGTGAATGATGTCCTCAGCTCCATGCA	1251
LpGK2_P226_cds	GCTCGAGCAGTGCTAGAGAGCATGTGCTTTCAAGTGAATGATGTCCTCAGCTCCATGCA	1236
	***** ***** *****	
LpGK2_P205_cds	AAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCTGAAGGGGAGTTCTTGTGCGCGTT	1311
LpGK2_P226_cds	AAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCTGAAGGGGAGTTCTTGTGCGAGTT	1296
	***** ***** **	
LpGK2_P205_cds	GATGGTGGTGCTACTGTTAACAACCTTCTCATGCAGATCCAGGCTGATTTGCTGGGTAGC	1371
LpGK2_P226_cds	GATGGTGGTGCTACTGTTAACAACCTTCTCATGCAGATCCAGGCTGATTTGCTGGGTAGC	1356

LpGK2_P205_cds	CCTGTTGTCAGACCAGCTGACATCGAGACAACAGCCCTTGGAGCTGCATACGCAGCAGGG	1431
LpGK2_P226_cds	CCTGTTGTCAGACCAGCTGACATTGAGACAACAGCCCTTGGAGCTGCATACGCAGCAGGG	1416
	***** *****	
LpGK2_P205_cds	TTAGCTGCTGGAGTTTGGACGAAGGAGGAAATATTTGCTGGCTTACACAAAGATAACACT	1491
LpGK2_P226_cds	TTAGCTGCTGGAGTCTGGACAAAGGAGGAAATATTTGCTGGCTTACACAAAGATAACACG	1476
	***** *****	
LpGK2_P205_cds	ACAGTGTTCCGCCCCAAACTGGATGATGCCCATAGGAAGAAGAGAGGAGACTCCTGGTAC	1551
LpGK2_P226_cds	ACGGTGTTCCGCCCCGAAACTGGATGATGCCCATAGGAAGAAGAGAGGCGACTCCTGGTAC	1536
	** *****	
LpGK2_P205_cds	AAGGCGGTTTCGAGATCATTGATCTGGCCGATCTTTCTCTCTAGAT	1598
LpGK2_P226_cds	AAGGCGGTTTCGAGATCATTGATCTGGCTGATCTTTCTCTCTAGAA	1583

Figure 3 continued

CLUSTAL 2.1 multiple sequence alignment

```

LpGK2_P205_as      MAGTEE---VYVASIDQGTTSTRFIVYDRHARPVASHQLEFKQHYPEAGWVEHDPMEIIE  57
LpGK2_P226_as      MAGTAAEEEVYVASIDQGTTSTRFIVYDRHARPVASHQLEFKQHYPEAGWVEHDPMEIIE  60
                    *****
LpGK2_P205_as      SVKICMSETLGKAAASGLNVDAGLKAIGITNQRETTVMWSKSTGLPLYNAIVWMDVRTSS  117
LpGK2_P226_as      SVKICMSETLGKAAASGLNVDAGLKAIGITNQRETTVMWSKSTGLPLYNAIVWMDVRTSS  120
                    *****
LpGK2_P205_as      ICRRLESELSSGGRTHFVETCGLPISTYFSALKLLWLMENVDAVKDAVKAGDALFGTIDTW  177
LpGK2_P226_as      ICRRLESELSSGGRTHFVETCGLPISTYFSALKLLWLMENVNSVKDAVRAGDALFGTIDTW  180
                    *****
LpGK2_P205_as      LIWNLTGGIGGKDTDGKELLGQHVTDCSNASRTMLMNLKALDWDKPTLEALGIPVGILPK  237
LpGK2_P226_as      LIWNLTGGVDGNG-----QHVTDCSNASRTMLMNLKALDWDKPTLEALGIPAGILPK  232
                    *****
LpGK2_P205_as      IISNSEKIGVVASGFPLAGVSI SGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEE  297
LpGK2_P226_as      IISNSEKIGVVASGFPLAGVSI SGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEE  292
                    *****
LpGK2_P205_as      VTQSSHGLLSTIAYKLGPDVPTNYALEGSIAIAGAAVQWLRDSLGISSAAEIEGLAESV  357
LpGK2_P226_as      VTQSSHGLLSTIAYKLGPDVPTNYALEGSIAIAGAAVQWLRDSLGISSAAEIEGLAESV  352
                    *****
LpGK2_P205_as      QDSGGIYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHIARAVLESMCFQVNDVLSSMH  417
LpGK2_P226_as      QDSGGIYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHIARAVLESMCFQVNDVLSSMH  412
                    *****
LpGK2_P205_as      KDAGEAGEVKSAGEFEFLLRVDGGATVNNLLMQIQADLLGSPVVRPADIETALGAAYAAG  477
LpGK2_P226_as      KDAGEAGEVKSAGEFEFLLRVDGGATVNNLLMQIQADLLGSPVVRPADIETALGAAYAAG  472
                    *****
LpGK2_P205_as      LAAGVWTKEEIFAGLHKDNTTVFRPKLDDAHRKKRGDSWYKAVSRSFDLADLSL  531
LpGK2_P226_as      LAAGVWTKEEIFAGLHKDNTTVFRPKLDDAHRKKRGDSWYKAVSRSFDLADLSL  526
                    *****

```

Figure 4

CLUSTAL 2.1 multiple sequence alignment

```

LpGK2_P205_as      MAGTEEVYVASIDQGTTSTRFIVYDRHARPVASHQLEFKQHYPEAGWVEHDPMEIIESVK 60
LpGK1_P205_as      -----

LpGK2_P205_as      ICMSETLGKAAASGLNVDAGLKAIGITNQRETTVMWSKSTGLPLYNAIVWMDVRTSSICR 120
LpGK1_P205_as      -----

LpGK2_P205_as      RLESELSSGGRTHFVETCGLPISTYFSALKLLWLMENVDVAVKDAVKAGDALFGTIDTWLIW 180
LpGK1_P205_as      -----MENVDVAVKDAVRAGDALFGTIDTWLIW 27
                        *****

LpGK2_P205_as      NLTGGIGGKDTDGKELLGQHVTDCSNASRTMLMNLKALDWDKPTLEALGIPVGILPKIIS 240
LpGK1_P205_as      NLTGGIGGKDTDGKELLGQHVTDCSNASRTMLMNLKALDWNKPTLEALGIPAGILPKIIS 87
                        *****

LpGK2_P205_as      NSEKIGVVASGFPLAGVSIISGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEEVTQ 300
LpGK1_P205_as      NSEKIGVVASGFPLAGVSIISGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEEVTQ 147
                        *****

LpGK2_P205_as      SSHGLLSTIAYKLGPDPVPTNYALEGSIAIAGAAVQWLRDLSLGISSAAEIEGLAESVQDS 360
LpGK1_P205_as      SSHGLLSTIAYKLGPDPVPTNYALEGSIAIAGAAVQWLRDLSLGISSAAEIEGLAESVQDS 207
                        *****

LpGK2_P205_as      GGIYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHIARAVLESMCFQVNDVLSSMHKDA 420
LpGK1_P205_as      GXIYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHIARAVLESMCFQVNDVLSSMHKDA 267
      * *****

LpGK2_P205_as      GEAGEVKSAGEFLLRVDGGATVNNLLMQIQADLLGSPVVRPADIETTALGAAYAAGLAA 480
LpGK1_P205_as      GEAGEVKSAGEFLLRVDGGATVNNLLMQIQADLLGSPVVRPADIETTALGAAYAAGLAA 327
                        *****

LpGK2_P205_as      GVWTKEEIFAGLHKDNTTVFRPKLDDAHRKKRGDSWYKAVSRSFDLADLSL 531
LpGK1_P205_as      GVWTKDEIFAGLHKENTTVFRPKLDDAHRKKRGDSWYKAVSRSFDLADLSL 378
      *****

```

Figure 5

CLUSTAL 2.1 multiple sequence alignment

```

LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|
TACACGAACTGATAGCTAATCTGAAGCTTTAGCCGCCATTCTTTTTTCGA 50

LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|
GGATAATAATCCCCGCGTCATGTGATCGCCCGCTGCACGATTGTGCGCCG 100

LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|
TGGGCGAACCAATCAGCGCGGCGAGGGCTCAAAAAAATCCAGCTTTTT 150

LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|
TTGGCGCGGACGCCAGGAATCCCGTCAAGTCCCAGAGGACAAGGTCCTGG 200

LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|
AGCGAGCGCGAGCGCGAGGAGATAGAGACGAACCACTAGCAGTCTCAGAT 62
-----GGCAGGGCCGAG 12
-----CCAAT 5
-----ATT 3
CGCCGGGCTGAGCGCGAGAAGAGCGGAGCGGATAGGCAAGAAGAGCAGA- 249

LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|
CAGAGTCAGACAACCGATAGTCATTGACTTCCACGC-----CCAG 102
CGGATCCAA-----CCAT 18
TGACGCCA-----TCTG 15
-GGTGTAGACAGCGCGACGCCATTGACTTCCCCGCTACCTGCTCCGT 298

LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|
CTCATCCACCTCCCTCCCAATCGGATCCACCCACCTTCTCGCCATAAC 152
CTAATCCACCCCGCGCCG-----CCGCGCCTATAAA 49
CTCGTCTGCCTCCTCCCGAT-----TTCCTCTCCT----- 46
CCCGTCCGTCTCCGTCCCAATCGGATCCACCCG---TCGTGCGCTATAAA 345
-----ACCCG---T----- 6

```

Figure 6

LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	CTCCCTC--CCTCCGCCCCGCTGCCGCGAGTCCGCTCTCAGCTCCCAACCG	200
LpGK2_P226_comp1791_c0_seq1_cd	-----	
gi 357166355 ref XM_003580635.	CTCTCCC--CTGTTAACCCA-----AGTCCAC-----CTCCGAACGTG	84
gi 115460941 ref NM_001060606.	CTCTCTC--TCTCCCTCTC-----TCCA-----TCCCTCCCG	76
gi 242077401 ref XM_002448592.	CTCTCCCTTCTCTCCGCTCT--CCGGGGCCCGGCTCC--TTCTCAACTG	390
gi 195645871 gb EU970286.1	-----	
LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	CTTCCCA---GTTTGACGCCGCCC---CCCTCCTCGGTTGTT-CCACCAG	243
LpGK2_P226_comp1791_c0_seq1_cd	-----	
gi 357166355 ref XM_003580635.	CT-CCCA---ATTCGACGCCGCGGCTTCTCTCCTCCCTCGTTTCCACCGG	130
gi 115460941 ref NM_001060606.	GCGTCCA---ATCTG-----TTCT-----G	93
gi 242077401 ref XM_002448592.	CCACCGATTGTTTGACTCCATCTCCAGTCCCTCCCTTCCCTTCCCACTTA	440
gi 195645871 gb EU970286.1	-----TCCCACTTC	15
LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	CTCACTC-----CCGCCCGCTCCCTCTGTGCCCGCCGCGG	281
LpGK2_P226_comp1791_c0_seq1_cd	-----ATGG	4
gi 357166355 ref XM_003580635.	CTCACTCAGCCCCCTGCTCTTGCTGCTGCTTCCCTGTGCCCGCCGCGGT	180
gi 115460941 ref NM_001060606.	CTT-CTC-----GCTGCGCCCCGGAATTGT-----GG	120
gi 242077401 ref XM_002448592.	ATTGGTC-----GCTGCTCCCTT-----CCTCCACGGC	468
gi 195645871 gb EU970286.1	ATTGGTC-----GCTGCTCCCTC-----CCTCCACGGC	43
LpGK2_P205_cds	-----ATGGCGG	7
gi 151421547 dbj AK252700.1	CGGAGCTTGCTCTAGGCGTGTGGGGACTGGGACTCTGAGAG-CATGGCAG	330
LpGK2_P226_comp1791_c0_seq1_cd	CGGGG-----ACGGCGG	16
gi 357166355 ref XM_003580635.	TG---TTTTCCGGCGAGCTTCGGGGGCCG---TCTGGGAGGTATGGCGG	223
gi 115460941 ref NM_001060606.	TGAGGTTT--TGTGAGGGTGTGGGGGCCG-----AGGAGGATGGCGG	160
gi 242077401 ref XM_002448592.	CGA--TCTCGTGCCGGCGAGAGGGATC-----ATGGCGG	500
gi 195645871 gb EU970286.1	CGA--TCCCGTGCCGGCGAGAGGGACC-----ATGGCGG	75
	* * * * *	
LpGK2_P205_cds	GGACGG-----AGGAGGTGTACGTGGCGTCGATCGACCAGGGCACC	48
gi 151421547 dbj AK252700.1	GGAAGGGGG-----AGGAGGTGTACGTGGCGTCGATCGATCAGGGGACG	374
LpGK2_P226_comp1791_c0_seq1_cd	CGGAGG-----AGGAGGTGTACGTGGCGTCGATCGACCAGGGCACC	57
gi 357166355 ref XM_003580635.	GGAAGGAGGGGGAGGAGGAGGTGTACGTGGCGTCGATCGACCAGGGCACC	273
gi 115460941 ref NM_001060606.	-----AGGAGGTGTATGTCGCGTCGATCGACCAGGGCACC	195
gi 242077401 ref XM_002448592.	GGGAAGGGG-----AGGAGGTGTACGTGGCGGCCATCGACCAGGGCACC	544
gi 195645871 gb EU970286.1	GGAAAGGGA-----AGGAGGTGTACGTGGCGGCCATCGACCAGGGCACC	119
	***** * * * * * * * * * *	
LpGK2_P205_cds	ACCAGCACCAGGTTTCATCGTCTACGACCGCCACGCCAGGCCCGTCGCCTC	98
gi 151421547 dbj AK252700.1	ACCAGCACCAGGTTTCATCGTCTACGACCGCCACGCCAGGCCCGTCGCCTC	424
LpGK2_P226_comp1791_c0_seq1_cd	ACCAGCACCAGGTTTCATCGTCTACGACCGCCACGCCAGGCCCGTCGCCTC	107
gi 357166355 ref XM_003580635.	ACCAGCACCAGGTTTCATCGTCTACGACCGCCACGCCAGGCCCGTCGCCTC	323
gi 115460941 ref NM_001060606.	ACGAGCACGCGGTTTCATCGTCTACGACCGCCACGCCAAGCCCGTCGCGTC	245
gi 242077401 ref XM_002448592.	ACCAGCACCAGGTTTCATCGTCTACGACCGCCACGCCAAGCCCGTCGCGTC	594
gi 195645871 gb EU970286.1	ACAAGCACCCGGTTTCATCGTCTACGACCGCCACGCCAACCCTCGCATC	169
	** * * * * * * * * * * * * * * * *	
LpGK2_P205_cds	GCACCAGCTCGAGTTCAAGCAGCACTACCCGGAGGCAGGATGGGTCGAGC	148
gi 151421547 dbj AK252700.1	CCACCAGCTCGAGTTCAAGCAGCACTACCCGGAGGCAGGATGGGTCGAGC	474
LpGK2_P226_comp1791_c0_seq1_cd	GCACCAGCTCGAGTTCAAGCAGCACTACCCGGAGGCAGGATGGGTCGAGC	157
gi 357166355 ref XM_003580635.	GCACCAGCTCGAGTTCAAGCAGCACTACCCGGAGGCAGGATGGGTCGAGC	373
gi 115460941 ref NM_001060606.	GCATCAGCTCGAGTTCAAGCAGCACTACCCGGAGGCTGGATGGGTTGAGC	295
gi 242077401 ref XM_002448592.	GCACCAGCTCGAGTTACGCAGCACTACCCGGAGGCAGGTTGGGTTGAGC	644
gi 195645871 gb EU970286.1	GCACCAGCTCGAGTTCAAGCAACACTACCCGGAGGCAGGTTGGGTTGAGC	219
	** * * * * * * * * * * * * * * * *	

Figure 6 continued

LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	ATGATCCTATGGAGATTATAGAGAGCGTGAAGATTTGCATGTCAGAGACG 198 ATGATCCTATGGAGATCATGGAGAGTGTGAAGGTCTGCATGGCAAAGACA 524 ATGATCCTATGGAGATTATAGAGAGCGTGAAGATTTGCATGTCAGAGACA 207 ATGATCCTATGGAGATCATGGAGAGCGTGAGGGCGTGCATGGCAAATGCA 423 ATGATCCTATGGAGATTATGGAGAGTGTGAAGATATGCATGGCGAAGGCA 345 ATGATCCTATGGAGATTATAGAGACTGTTATGGTGTGTATGAATGAGGCA 694 ATGATCCTATGGAGATTATAGAGACTGTTAAGGTGTGTATGAAAGAGGCA 269 ***** ** * * * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	CTTGCCAAAGCTGCAGCTAGTGGACTTAATGTGGATGCTGGTTTGAAGGC 248 CTCGAAAAGCTGCAGCCAGTGGACTTAATGTGGATGCTAGTTTGAAGGC 574 CTTGCCAAAGCTGCAGCTAGTGGACTTAATGTGGATGCTGGTTTGAAGGC 257 CTTGAGAAAGCTGCTGCTGATGGATTAATGTGTCTGCTGGCTTGAAGGC 473 CTTGACAAAGCTGCAGCTGATGGTCATAATATGGATGTTGGACTGAAGGC 395 GTTGGCAAAGCCAAAGATGGTAAATACAATGTGGTTGCTGGTTTGAAGGC 744 GTTGGCAAAGCCAAAGATGGTAAACACAATGTGGTTGCTGGTTTGAAGGC 319 * * * * * * * * * * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	CATTGGGATCACAATCAGAGGGAGACAACAGTTATGTGGAGCAAATCCA 298 CATTGGAATCACAATCAGAGGGAGACTACCGTTGTGTGGAGCAAATCCA 624 CATTGGGATCACAATCAGAGGGAGACAACAGTTATGTGGAGCAAATCCA 307 TATAGGGATCACAATCAGAGGGAGACCACTGTTATGTGGAGCAAATCCA 523 CATTGGTATTACAATCAGAGGGAAACCACTGTTATGTGGAGCAAATCCA 445 CATTGGGATCACAATCAGAGGGAAACCACTGTTATGTGGAGTAAATCCA 794 CATTGGGATCACAATCAGAGGGAAACCACTGTTATGTGGAGTAAATCCA 369 * * * * * * * * * * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	CAGGTCTTCCACTATACAATGCCATAGTGTGGATGGATGTTTCGCACAAGC 348 CAGGCCTTCCGCTCTACAATGCCATTGTGTGGATGGATGTTTCGCAC TAGC 674 CAGGTCTTCCACTATACAATGCCATAGTGTGGATGGATGTTTCGCACAAGC 357 CAGGCCTTCCGCTCTATAATAGCCATTGTGTGGATGGACGTTTCGCACAAGC 573 CAGGCCTTCCCTCTATAACAATGCCATTGTATGGATGGATGCTCGCACAAGC 495 CTGGCCATCCACTTTATAATGCCATTGTGTGGATGGATGCTCGCACAAGC 844 CTGGCCCTCCACTGTATAATGCTATTGTGTGGATGGATGCTCGCACAAGC 419 * * * * * * * * * * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	TCTATTTGCAGGAGACTGGAAGTGAGCTATCAGGTGGTAGAACACACTT 398 TCTATTTGCAGGAGACTGGAAGTAAGCTATCAGGTGGTAGAACGCACTT 724 TCCATTTGCAGGAGATTGGAAGTGAGCTATCAGGTGGCAGAACACACTT 407 TCCATTTGCAGGAGATTGGAAGGTGAGCTGGCGGGTGGTAGAACACACTT 623 CCCATTTGCAGGAGATTGGAAGTGATCTTTCTGGTGGTAGAACCCACTT 545 CCCGTTTGCAGGAGATTAGAAAGTGAGCTATCAGGCGGTAGAACCCACTT 894 CCTGTTTGCAGGAGATTGAAAAATGAGCTGTGAGGCGGTAGGACCCACTT 469 * * * * * * * * * * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	TGTGGAGACGTGTGGCTACCAATCAGTACCTATTTTCAGTGCTCTTAAGT 448 TGTGGAGACATGTGGTTTACCAATCAGTACCTACTTCAGTGCTCTGAAGT 774 TGTGGAGACATGTGGCTTACCAATCAGCACATATTTTCAGTGCTCTCAAGT 457 TGTGGAGACATGTGGATTACCAATCAGTACCTATTTTAGTGCTCTGAAGC 673 TGTCGAGACATGTGGATTGGCAATCAGTACCTATTTTCAGTGCTTTGAAGG 595 TGTGGAGACATGTGGGCTGCCAATCAGTACCTATTTTCAGTGCTCTGAAGT 944 CGTGGAGACATGCGGGTTGCCAATCAGTACCTATTTTCAGTGCTCTGAAGT 519 * * * * * * * * * * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	TATTATGGTTGATGGAGAATGTTGACGCTGTCAAGGATGCAGTCAAGGCT 498 TATTATGGTTGATGGAGAATGTTGACGCTGTCAAGGATGCAGTCAAGGCT 824 TATTGTGGTTGATGGAGAATGTTAACTCTGTCAAGGATGCAGTTCGGGCT 507 TACTTTGGTTGATGGAAAATGTTGATGCTGTCAAGGATGCAGTCCGGGCC 723 TGTTATGGTTGATCGAAAATGTGGATGCCGTCAAGAATGCAGTCCGAGCT 645 TATTATGGTTGATGGAAAATGTGGATGCTGTCAAGGATGCAATCAAGACT 994 TATTATGGTTGATGGAAAATGTGGATGCTGTCAAGGATGCAGTCCGGACT 569 * * * * * * * * * * * * * *

Figure 6 continued

LpGK2_P205_cds	GGTGATGCATTATTTGGCACAATCGACACCTGGTTGATCTGGAACCTCAC	548
gi 151421547 dbj AK252700.1	GGTGATGCATTATTTGGCACAATCGATACCTGGTTGATCTGGAACCTAAC	874
LpGK2_P226_comp1791_c0_seq1_cd	GGTGATGCATTATTTGGCACAATCGACACCTGGTTGATCTGGAACCTCAC	557
gi 357166355 ref XM_003580635.	GGGGATGCATTATTTGGCACAATCGACACCTGGCTGATCTGGAACCTGAC	773
gi 115460941 ref NM_001060606.	GGTGATGCCTTATTTGGCACAATTGACACCTGGTTAATCTGGAACCTCAC	695
gi 242077401 ref XM_002448592.	GGTGACGCCCTTATTTGGCACAATCGACACCTGGTTGATTTGGAACCTTAC	1044
gi 195645871 gb EU970286.1	GGTGACGCTTTATTCGGCAGCATCGACACCTGGTTGATTTGGAACCTTAC	619
	*** **	
LpGK2_P205_cds	AGGAGGCATTGGTGGGAAAGATACTGATGGAAAAGAATTACTTGGGCAGC	598
gi 151421547 dbj AK252700.1	TGGAGGCATTGGTGGGAAAGACTGTGATGGAAAGGAGTTAGTTGGGCGGC	924
LpGK2_P226_comp1791_c0_seq1_cd	AGGAGGTGTTGACGGGAA-----TGGGCAGC	583
gi 357166355 ref XM_003580635.	AGGAGGCATTGGCGGTAAAGA-----TGGAAAACAG---ATCGGACAGC	814
gi 115460941 ref NM_001060606.	AGGAGGCATTGGCGGGACAGATCGTGATGGAAATAAAGTATTTGGGCACC	745
gi 242077401 ref XM_002448592.	TGGAGGTGTTGCTGG-----TGGGCAGC	1067
gi 195645871 gb EU970286.1	AGGAGGTGTTGCTGG-----TGGGCAGC	642
	***** **	
LpGK2_P205_cds	ATGTCACAGACTGCTCAAATGCATCTCGCACGATGCTTATGAATCTAAAG	648
gi 151421547 dbj AK252700.1	ATGTCACAGACTGCTCAAACGCATCACGGACAATGCTTATGAATCTGAAG	974
LpGK2_P226_comp1791_c0_seq1_cd	ATGTCACAGACTGCTCAAATGCATCTCGCACGATGCTTATGAATCTAAAG	633
gi 357166355 ref XM_003580635.	ATGTCACAGATTGCTCAAATGCATCTCGCACGATGCTTATGAATCTAAAG	864
gi 115460941 ref NM_001060606.	ATGTCACAGACTGCTCCAATGCATCTCGCACAATGCTTATGAACCTGAAG	795
gi 242077401 ref XM_002448592.	ATGTCACAGATTGCTCAAATGCATCTCGTACAATGCTTATGAATCTAAAG	1117
gi 195645871 gb EU970286.1	ATGTCACGGATTGCTCAAATGCATCTCGTACAATGCTTATGAATCTAAAG	692
	***** **	
LpGK2_P205_cds	GCACTTGATTGGGACAAGCCTACACTTGAGGCATTAGGCATTCTGTTGG	698
gi 151421547 dbj AK252700.1	GCACTTGACTGGGACAAGCCTACACTTGAGGCATTAGGCATTCTGCTGG	1024
LpGK2_P226_comp1791_c0_seq1_cd	GCACTTGATTGGGACAAGCCTACACTTGAGGCATTAGGCATTCTGCTGG	683
gi 357166355 ref XM_003580635.	GCACTTGATTGGGACAAGCCTACACTCAAGGCATTAGGAATTCTGCTGA	914
gi 115460941 ref NM_001060606.	GCACTTGACTGGGACAAACCAACACTTGAGACGTTAGGAATTCTGCTGA	845
gi 242077401 ref XM_002448592.	ACACTTGACTGGGATAAGCCAACACTTGATGTGTTAGGAGTTCTGTTGA	1167
gi 195645871 gb EU970286.1	ACACTTGACTGGGATAAGCCAACACTTGCTGTGTTAGGAGTTCTGTTGA	742
	***** **	
LpGK2_P205_cds	CATCTTGCCGAAGATTATCAGCAACTCGGAGAAAATTGGCGTGGTTGCCA	748
gi 151421547 dbj AK252700.1	CATCTTGCCAAAGATTGTGAGCAACTCCGAGAAAATTGGTGTGTTGCCA	1074
LpGK2_P226_comp1791_c0_seq1_cd	CATCTTGCCGAAGATTATCAGCAACTCGGAAAAAATTGGCGTGGTTGCCA	733
gi 357166355 ref XM_003580635.	CATCTTACCAAAGATTATCAGTAACCTCAGAGAAAATTGGTTGGTTGCCG	964
gi 115460941 ref NM_001060606.	GATATTGCCCAAAATTATCAGTAACCTCGGAGAGAATTGGTGTGGTTGCCA	895
gi 242077401 ref XM_002448592.	GATTTTGCCAAAGATTATCAGTAATTCAGAGAAAATTGGTGTGGTCGCCA	1217
gi 195645871 gb EU970286.1	GATTTTGCCAAAGATTATCAGTAATTCAGAGAAAATCGGTGTGGTCGCCA	792
	** **	
LpGK2_P205_cds	GTGGGTTCCCTTTGGCTGGCGTCTCCATCTCAGGGTGTITTAGGAGACCAG	798
gi 151421547 dbj AK252700.1	GTGGGTTCCCTTTGGCTGGTGTGTCTATCTCAGGGTGTCTAGGAGATCAG	1124
LpGK2_P226_comp1791_c0_seq1_cd	GTGGGTTCCCTTTGGCTGGCGTCTCCATCTCAGGGTGTCTAGGAGACCAG	783
gi 357166355 ref XM_003580635.	ATGGGTTTCCTTTGGCTGGTGTCCCCATCTCAGGGTGTCTAGGAGATCAG	1014
gi 115460941 ref NM_001060606.	ATGGGTTTCCCTTTGGCGGGCGTCCCTATTGCTGGGTGCCCTAGGAGACCAG	945
gi 242077401 ref XM_002448592.	AAGAGTTCCCTTTGCGAGGTGTCCCATCTCAGGGTGCCTTGGAGATCAG	1267
gi 195645871 gb EU970286.1	AAGAGTTCCCGTTTGCGAGGTGTCCCATCTCGGGGTGTCTTGGAGATCAG	842
	* **	
LpGK2_P205_cds	CATGCTGCAATGCTCGGACAGCTATGCCAGAAAGGTGAAGCCAAAAGCAC	848
gi 151421547 dbj AK252700.1	CATGCTGCGATGCTTGGGCAGCTATGCCAGAAAGGTGAAGCAAAAAGCAC	1174
LpGK2_P226_comp1791_c0_seq1_cd	CATGCTGCAATGCTCGGGCAGCTATGCCAGAAAGGTGAAGCCAAAAGCAC	833
gi 357166355 ref XM_003580635.	CATGCAGCAATGCTCGGGCAATTATGCCAAAAAGGTGAAGCCAAAAGCAC	1064
gi 115460941 ref NM_001060606.	CATGCTGCAATGCTTGGGCAGCTGTGCCAGAAAGGTGAAGCCAAAAGCAC	995
gi 242077401 ref XM_002448592.	CATGCTGCTATGCTTGGGCAGTTGTGCAAGAAAGGTGAAGCGAAAAGCAC	1317
gi 195645871 gb EU970286.1	CATGCTGCTATGCTTGGGCAGCTGTGCCAGAAAGGTGAAGCGAAAAGCAC	892
	***** **	

Figure 6 continued

LpGK2_P205_cds	CTACGGAAC TGGTGCCTTCATCCTTCTCAACACAGGGGAGGAGGTTACCC	898
gi 151421547 dbj AK252700.1	CTATGGAAC TGGTGCCTTCATCCTTCTCAACACAGGGGAGGAGGTTACTC	1224
LpGK2_P226_comp1791_c0_seq1_cd	CTACGGAAC TGGTGCCTTCATCCTTCTCAACACAGGGGAGGAGGTTACCC	883
gi 357166355 ref XM_003580635.	CTATGGAAC TGGTGCCTTCATCCTTCTCAACACAGGGGAGGAGGTTACCC	1114
gi 115460941 ref NM_001060606.	CTACGGAAC TGGTGCCTTCATTCTTCTCAACACAGGGGAGGAGCCTACCC	1045
gi 242077401 ref XM_002448592.	CTATGGAAC TGGTGCCTTCATTCTTCTTAACACAGGGGAAGAACCTACCC	1367
gi 195645871 gb EU970286.1	CTATGGAAC TGGTGCCTTCATCCTTCTTAACACAGGGGAAGAGCCTACCC	942
	*** *****	
LpGK2_P205_cds	AGTCCTCCCATGGTCTTCTTAGCACCATTTGCTTACAAGCTTGCCCCAGAT	948
gi 151421547 dbj AK252700.1	AGTCCACCCACGGTCTTCTTAGCACCATTTGCTTACAACTTGCCCCAGAT	1274
LpGK2_P226_comp1791_c0_seq1_cd	AGTCCTCCCATGGTCTTCTTAGCACCATTTGCTTACAAGCTTGCCCCAGAT	933
gi 357166355 ref XM_003580635.	AATCTTCCCATGGGCTTCTTAGCACAAATTGCTTATAAGCTTGCCCCAAGC	1164
gi 115460941 ref NM_001060606.	AATCCTCCCATGGTCTTCTTAGCACGATTGCTTATAAGCTTGCCCCAAGC	1095
gi 242077401 ref XM_002448592.	AGTCCTCCCACGGTCTTCTTAGCACTATCGCATATAAGCTTGCCCCAAGC	1417
gi 195645871 gb EU970286.1	AATCCTCCCATGGCCTTCTTAGTACCATTGCTTACAAGCTTGGTCCAGCT	992
	* ** ***** **	
LpGK2_P205_cds	GTACCCACCAATTATGCTCTAGAAGGCTCAATTGCTATTGCTGGTGCAGC	998
gi 151421547 dbj AK252700.1	GTGCCAACTAATTATGCTCTAGAAGGGTCAATTGCAATTGCTGGTGCAGC	1324
LpGK2_P226_comp1791_c0_seq1_cd	GTACCAACCAATTACGCTCTAGAAGGCTCAATTGCTATTGCTGGTGCAGC	983
gi 357166355 ref XM_003580635.	GCACCCACTAATTATGCTCTAGAAGGCTCGATTGCAATTGCTGGTGCAGC	1214
gi 115460941 ref NM_001060606.	GCACCCACTAATTATGCTCTAGAAGGCTCCATCGCAATTGCTGGTGCAGC	1145
gi 242077401 ref XM_002448592.	GCACCCACTAATTATGCCCCTTGAAGGATCCATTGCAATTGCTGGAGCAGC	1467
gi 195645871 gb EU970286.1	GCACCCACTAATTATGCTCTTGAAGGCTCCATTGCAATTGCAGGCGCAGC	1042
	* ** *	
LpGK2_P205_cds	AGTTCAGTGGTTGAGGGACAGCCTTGGGATCATTAGCTCAGCAGCAGAAA	1048
gi 151421547 dbj AK252700.1	AGTTCAGTGGTTGAGGGACAGCCTTGGAATCATTTCCTCGGCTGCAGAAA	1374
LpGK2_P226_comp1791_c0_seq1_cd	AGTTCAGTGGTTGAGGGACAGCCTTGGGATCATTAGCTCAGCAGCAGAAA	1033
gi 357166355 ref XM_003580635.	AGTTCAGTGGTTGAGGGACAGCCTCGGAATCATTGCTCAGCAGCAGAAA	1264
gi 115460941 ref NM_001060606.	AGTTCAGTGGTTGAGAGACAGCCTTGGAATCATTAGCACAGCAGCTGATA	1195
gi 242077401 ref XM_002448592.	AGTTCAGTGGTTGAGGGATAGCCTTGGAATCATTCAAACAGCAGCCGAAA	1517
gi 195645871 gb EU970286.1	AGTTCAGTGGCTGAGGGACAGCCTTGGAATCATTAGTCCAGCAGCTGAGA	1092
	***** *	
LpGK2_P205_cds	TTGAAGGGTTGGCTGAAAGTGTGCAAGATTACAGCGGGATATACTTTGTG	1098
gi 151421547 dbj AK252700.1	TTGAAGGGTTAGCTGAAAGTGTGCAAGATTACAGGTGGCATATACTTTGTG	1424
LpGK2_P226_comp1791_c0_seq1_cd	TTGAAGGGTTGGCTGAAAGTGTGCAAGATTACAGGTGGGATATACTTTGTG	1083
gi 357166355 ref XM_003580635.	TTGAAGAGTTGGCTGAAAGCGTGCAGGATTACAGGTGGGATATACTTTGTG	1314
gi 115460941 ref NM_001060606.	TTGAAAAGTTGGCTGATACGTGTCAGGATTACAGGTGGAATATACTTTGTT	1245
gi 242077401 ref XM_002448592.	TTGAAAAGTTGGCTGAAACTGTACCGGATTACAGGTGGTGTATACTTTGTG	1567
gi 195645871 gb EU970286.1	TCGAAAAGTTGGCTGAAACAGTGCCAGATTACAGGTGGAGTGTACTTTGTG	1142
	* ** *	
LpGK2_P205_cds	CCAGCATTCATGGATTGTTTCGCACCATGGTGGAGGGATGACGCGAGAGG	1148
gi 151421547 dbj AK252700.1	CCAGCATTCATGGATTGTTTCGCGCCATGGTGGAGGGATGATGCTAGGGG	1474
LpGK2_P226_comp1791_c0_seq1_cd	CCGGCATTCATGGATTGTTTCGCACCATGGTGGAGGGATGATGCAAGAGG	1133
gi 357166355 ref XM_003580635.	CCAGCATTCATGGATTGTTTCGCACCATGGTGGAGGGATGATGCCAGGGG	1364
gi 115460941 ref NM_001060606.	CCAGCATTCATGGGTTGTTTGCACCATGGTGGAGGGATGATGCAAGAGG	1295
gi 242077401 ref XM_002448592.	CCAGCATTCATGGGTTGTTTGCACCATGGTGGAGGGATGATGCAAGGGG	1617
gi 195645871 gb EU970286.1	CCAGCATTTAATGGGTTGTTTGCACCATGGTGGAGAGATGATGCGAGGGG	1192
	** *****	
LpGK2_P205_cds	TATCTGCATAGGAATCACAGGTTTACAAACAAGGGGCACATTGCTCGAG	1198
gi 151421547 dbj AK252700.1	GATCTGCATAGGAATCACAGGTTTACCAACAAGGGGCACATTGCTCGAG	1524
LpGK2_P226_comp1791_c0_seq1_cd	TATCTGCATAGGAATCACAGGTTTACAAATAAGGGGCACATTGCTCGAG	1183
gi 357166355 ref XM_003580635.	GATCTGCATAGGAATCACAGGTTTACCAATAAGGGGCACATTGCTCGAG	1414
gi 115460941 ref NM_001060606.	CATATGTATAGGAATCACACGGTTTACTAACAAGGGGCACATTGCTCGAG	1345
gi 242077401 ref XM_002448592.	GATCTGCATTGGAATCACAGGTTTACGAATAAGGGTCACATTGCTCGAG	1667
gi 195645871 gb EU970286.1	AATTTGCATCGGAATCACAGGTTTACAAATAAGGGGCACATTGCTCGAG	1242
	** *	

Figure 6 continued

LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	CAGTGCCTTGAGAGTATGTGCTTTCAAGTGAATGATGTCCTCAGCTCCATG 1248 CAGTGCCTGAGAGTATGTGCTTTCAGGTGAATGATGTCCTCAGCTCCATG 1574 CAGTGCCTGAGAGTATGTGCTTTCAGGTGAATGATGTCCTCAGCTCCATG 1233 CAGTGCCTTGAGAGTATGTGCTTTCAGGTGAATGATGTCCTCAGCTCCATG 1464 CAGTGCCTGAGAGTATGTGCTTTCAGGTGAATGATGTCCTCAGCTCCATG 1395 CAGTGCCTGAGAGTATGTGCTTTCAGGTGAATGATGTCCTCAGCTCCATG 1717 CAGTGCCTGAGAGTATGTGCTTTCAGGTGAATGATGTCCTCAGCTCCATG 1292 *****
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	CATAAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCTGAAGGGGAGTT 1298 CATAAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCTGAAGGGGAGTT 1624 CATAAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCTGAAGGGGAGTT 1283 CACAAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCTGAAGGAGAGTT 1514 CACAAGGATGCTGGGGAGGCAGGAGAAGTGAAGAGTGCTGAAGGGGAGTT 1445 CACAAGGATGCTGGAGAGGCGGGAGAAGTAAAGAGTGCTGAAGGGGAATT 1767 CACAAGGATGCTGGAGAGGCAGGAGAAGTAAAGAGCGCAGAAGGAGAGTT 1342 ** *****
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	CTTGTTGCGCGTTGATGGTGGTGCTACTGTTAACAACCTTCTCATGCAGA 1348 CTTGCTACGTGTTGATGGTGGTGCTACTGTTAACAACCTTCTCATGCAGA 1674 CTTGTTGCGAGTTGATGGTGGTGCTACTGTTAACAACCTTCTCATGCAGA 1333 CTTGTTGCGTGTGATGGTGGTGCTACTGTCAATAACCTTCTCATGCAGA 1564 CCTGCTGCGCGTTGATGGTGGTGCTACTGTCAACAATCTTCTCATGCAGA 1495 CTTGTTGCGTGTGATGGTGGTGCTACTGTTAATAATCTTCTCATGCAGA 1817 CTTATTGCGTGTGATGGTGGTGCTACTGTTAATAATCTTCTCATGCAGA 1392 * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	TCCAGGCTGATTTGCTGGGTAGCCCTGTTGTCAGACCAGCTGACATCGAG 1398 TCCAGGCTGATTTGCTGGGTAGCCCTGTTGTCAGGCCAGCCGACATTGAG 1724 TCCAGCTGATTTGCTGGGTAGCCCTGTTGTCAGACCAGCTGACATTGAG 1383 TCCAGGCTGATTTGCTGGGTAGCTCCTGTTGTCAGACCAGCTGACATCGAG 1614 TCCAGGCTGATTTATTGGGTAGCCCTGTTGTCAGACCAGCTGACATCGAG 1545 TTCAGGCTGATCTATTGGGTAGCCCTGTTGTCAGACCAGCTGACATAGAG 1867 TCCAGGCTGATTTATTAGGCAGCCCTGTTGTCAGACCAGCTGACATAGAG 1442 * *****
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	ACAACAGCCCTTGAGCTGCATACGCAGCAGGGTTAGCTGCTGGAGTTTG 1448 ACAACGCGCCCTCGGGGCGCATACGCAGCAGGGTTAGCTGCCGGAGTCTG 1774 ACAACAGCCCTTGAGCTGCATACGCAGCAGGGTTAGCTGCTGGAGTCTG 1433 ACAACAGCCATTGAGCCGCATATGCAGCAGGGTTAGCTGCTGGTGTCTG 1664 ACAACGCGCCCTTGAGCTGCGTATGCAGCTGGATTAGCTGTTGGAGTCTG 1595 ACGACAGCCCTTGAGCTGCATATGCAGCTGGTTTGGCCGCTGGAGTCTG 1917 ACCACAGCCCTCGGAGCTGCATATGCTGCTGGGTGGCTGCAGGAGTTTG 1492 ** * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	GACGAAGGAGGAAATATTTGCTGGCTTACACAAAGATAACACTACAGTGT 1498 GACGAAGGAGGAAATATTCGAGGCCTGCACAAGGAGAACAACGGTGT 1824 GACAAAGGAGGAAATATTTGAGGCTTACACAAAGATAACACGACGGTGT 1483 GACAAAGGAGGAAATATTTGAGGTTTGCACAAGGAGAACAACGACAGTCT 1714 GTCGAAGGAACAAATTTTCGCTGGTTTGCACAAGGAGAACAACGAGTCT 1645 GACGAAGGAGCAAGTATTCGAGGTTTGCACAAGGAGAACAACGACGGTCT 1967 GACCAAGGAGAAGGTTTTTTCAGGTTTGCACAAGGAGAACAACGACAGTCT 1542 * * * * *
LpGK2_P205_cds gi 151421547 dbj AK252700.1 LpGK2_P226_comp1791_c0_seq1_cd gi 357166355 ref XM_003580635. gi 115460941 ref NM_001060606. gi 242077401 ref XM_002448592. gi 195645871 gb EU970286.1	TCCGCCCCAACTGGATGATGCCCATAGGAAGAAGAGAGGAGACTCCTGG 1548 TCCGCCCCAAGTTAGACGATGCACACAGGAAGAAAAGAGGAGACTCCTGG 1874 TCCGCCCCGAACTGGATGATGCCCATAGGAAGAAGAGAGGCGACTCCTGG 1533 TCCGCCCCGAAATTGGATGAGGCACACAGGAAGAAGAGAGGGGACTCCTGG 1764 TCCGCCCCGAAATTGGATGAGGCCACACAGGAGGAAGAGAGCAGACTCCTGG 1695 TCCGCCCCAACTGGATGAGGCCACACAGGAAGAAGCAGCAGATTTCATGG 2017 TCCGCCCCGAACTGGATGAGGCCACACAGGAAGAAGAGGGCGGATTTCATGG 1592 *****

Figure 6 continued

LpGK2_P205_cds	TACAAGGCGGTTTCGAGATCATTTGATCTGGCCGATCTTTCTCTCTAGAT	1598
gi 151421547 dbj AK252700.1	TACAAGGCGGTTTCGAGATCATTTGACCTGGCTGACCTTTCTCTCTAGTT	1924
LpGK2_P226_comp1791_c0_seq1_cd	TACAAGGCGGTTTCGAGATCATTTGATCTGGCTGATCTTTCTCTCTAG--	1581
gi 357166355 ref XM_003580635.	TACAAGGCGGCTCAAGATCATTTGATCTGGCTGATCTTTCTCTCTAGAT	1814
gi 115460941 ref NM_001060606.	TACAAGGCCGTTTCAAGATCATTTGATTGGCTGACCTTTCTCTTTAGAA	1745
gi 242077401 ref XM_002448592.	TACAAGGCGGTGTCAAGATCATTTGACTTGGCTGATCTTTCCCTTTAAAT	2067
gi 195645871 gb EU970286.1	TACAAGGCGGTTTCAAGATCATTCGATCTGGCTGATCTTTCTCTGTAAAT	1642
	***** ** ** ***** ** **** ** ***** ** **	
LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	TG-----TTACACATTGCTGCGCCGTAAATTGGACCAG-----TTT	1960
LpGK2_P226_comp1791_c0_seq1_cd	-----AA-----	1583
gi 357166355 ref XM_003580635.	TG-----TTACACCTTGTTCACGGTAAATTGGATCAGGTGTGTTT	1855
gi 115460941 ref NM_001060606.	TGACGAGGGATTTGTGAATTACTTC----TGGATTGTATAAGGTGC--TTA	1790
gi 242077401 ref XM_002448592.	TGG-----CACACCTCGTCGC--CAGTCAGTTGTG-----	2095
gi 195645871 gb EU970286.1	T-----TATCCTGCCAG--CGCTCCTCTAT-----	1665
LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	G-----TGCC-----AGCCCTTGTTTG-----TCTTGTCACAG	1988
LpGK2_P226_comp1791_c0_seq1_cd	-----	
gi 357166355 ref XM_003580635.	GAAGGTGGTATATC-----ATTTAACCCCTG-----T--TTGTCAATG	1891
gi 115460941 ref NM_001060606.	GCACTTGTATGTCCATGAATAAGTTGGTCAATGGAACATCCTTCAAAAG	1840
gi 242077401 ref XM_002448592.	-----ACCTGCGTATGTAGCACATTATGAATA	2122
gi 195645871 gb EU970286.1	-----CTG---GTTTTGGAAGCTGCAATG	1687
LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	AATAAACGG-----TG-TGTTGTA--CATT-----TTTCG	2016
LpGK2_P226_comp1791_c0_seq1_cd	-----	
gi 357166355 ref XM_003580635.	AATAAATGG-----AAACGTTGTG--CATT-----CTTCG	1920
gi 115460941 ref NM_001060606.	AATTAGTTGATTCCCCACGCTTTGCTGCTGGAATTATTAAGTAATTTTCA	1890
gi 242077401 ref XM_002448592.	AGCAGGCAG-----CCCTGTAGTGTGCATTG-----CTTCG	2153
gi 195645871 gb EU970286.1	AATAAATTG-----GTCTTCGCTGGGACTAA-----TCTTA	1718
LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	GTGCGATTCTTCTTTGGCT-----TTATAGGAAGAT-----	2047
LpGK2_P226_comp1791_c0_seq1_cd	-----	
gi 357166355 ref XM_003580635.	GTG--GGTCTCTCTTCAGTT-----TCATAGGAAGAA-----	1950
gi 115460941 ref NM_001060606.	ATATACTTTTAAACAATTGAGATAGGAGTAAAAAGAAAAATAATGAAA	1940
gi 242077401 ref XM_002448592.	GTG-----	2156
gi 195645871 gb EU970286.1	ACT-----	1721
LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	-----CTGGAGCA-----TAAATGTC	2063
LpGK2_P226_comp1791_c0_seq1_cd	-----	
gi 357166355 ref XM_003580635.	-----TCAGAGCA-----TACTTGT--	1965
gi 115460941 ref NM_001060606.	TGTCAGTGTTTATTTCAGTACTTATCCTGAAACAAACAAATGATGCCTGCG	1990
gi 242077401 ref XM_002448592.	-----GGGTGTT	2163
gi 195645871 gb EU970286.1	-----GCTTGTT	1728
LpGK2_P205_cds	-----	
gi 151421547 dbj AK252700.1	CTT-----GTATATCTATAAGTACACG---AGCATC-----	2091
LpGK2_P226_comp1791_c0_seq1_cd	-----	
gi 357166355 ref XM_003580635.	-TC-----GTATATCTATAGGTACATG---GATATC-----	1992
gi 115460941 ref NM_001060606.	CTTTAATGTGGAACATGTTGATGAGTATATGTTAAATATTATAAAAAATGA	2040
gi 242077401 ref XM_002448592.	-CC-----TGTATTATTTCCTCA-----	2179
gi 195645871 gb EU970286.1	ATT-----TATATATAACATCA-----	1745

Figure 6 continued

```
LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|

LpGK2_P205_cds
gi|151421547|dbj|AK252700.1|
LpGK2_P226_comp1791_c0_seq1_cd
gi|357166355|ref|XM_003580635.
gi|115460941|ref|NM_001060606.
gi|242077401|ref|XM_002448592.
gi|195645871|gb|EU970286.1|

-----
----TGGATTCTTCAGAA-----AGAAGTGTGTGTCG 2119
-----
----TGGATTCTTCAGAA-----AGAAATATTTGTGT 2020
TAGATAGATTTATTAAAATATTGTAAAAATGATGTAAGAGATGATTGGAT 2090
----TGCAGTAAGCAGGA-----GAGGATG-ACAGTT 2206
----CGATGTTTATAAAA-----GTGAGTACATGAAT 1773

-----
AGTATCTGCCT-----CCAAAAAAAAAAAAAAAA----- 2146
-----
AGTATTTGTAT-----TTTTCTGAAGTA----- 2043
A-TGTTTGTATGTTGATTTGTAAAATTAAATAGATTATAGATGATATTGC 2139
G-TATTTA--T-----TTACGGACATAAGTACAGAAACAT----- 2238
A-TCCTCAGAT-----TTCCAAAAAAAAAAAAAAAA----- 1803
```

Figure 6 continued

CLUSTAL 2.1 multiple sequence alignment

```

LpGK2_P226_comp1791_c0_seq1      MAGTAAEEVYVASIDQGTSTRFIVYDRHARPVASHQLEFKQHYPEAGW 50
LpGK2_F1_30_P205_AS_full          MAGT---EEVYVASIDQGTSTRFIVYDRHARPVASHQLEFKQHYPEAGW 47
gi|357166356|ref|XP_003580683.    MAGKEGEEVYVASIDQGTSTRFIVYDRHARPVASHQLEFKQHYPEAGW 50
gi|38352013|gb|AAR18687.1|        -----MAEEVYVASIDQGTSTRFIVYDRHAKPVASHQLEFKQHYPEAGW 45
gi|239049939|ref|NP_001132106.    MAGK--GKEVYVAAIDQGTSTRFIVYDRHAKPVASHQLEFKQHYPEAGW 48
gi|242077402|ref|XP_002448637.    MAGE--GEEVYVAAIDQGTSTRFIVYDRHAKPVASHQLEFKQHYPEAGW 48
                                     :*****:*****:*****:*****:*****:
LpGK2_P226_comp1791_c0_seq1      VEHDPMEIIESVKICMSETLGKAAASGLNVDAGLKAIGITNQRETTVMWS 100
LpGK2_F1_30_P205_AS_full          VEHDPMEIIESVKICMSETLGKAAASGLNVDAGLKAIGITNQRETTVMWS 97
gi|357166356|ref|XP_003580683.    VEHDPMEIMESVRACMANALEKAAADGLNVSAGLKAIGITNQRETTVMWS 100
gi|38352013|gb|AAR18687.1|        VEHDPMEIMESVKICMAKALDKAAADGHNMDVGLKAIGITNQRETTVMWS 95
gi|239049939|ref|NP_001132106.    VEHDPMEI IETVKVCMKEAVGKAKDGKHNVVAGLKAIGITNQRETTVMWS 98
gi|242077402|ref|XP_002448637.    VEHDPMEI IETVMVCMNEAVGKAKDGKYNVVAGLKAIGITNQRETTVMWS 98
                                     *****:*:*  **  :::  **  .  *:  .*****:*****
LpGK2_P226_comp1791_c0_seq1      KSTGLPLYNIAIVWMDVRTSSICRRLESELSSGGRTHFVETCGLPITYFSA 150
LpGK2_F1_30_P205_AS_full          KSTGLPLYNIAIVWMDVRTSSICRRLESELSSGGRTHFVETCGLPITYFSA 147
gi|357166356|ref|XP_003580683.    KSTGLPLYNIAIVWMDVRTSSICRRLESELSSGGRTHFVETCGLPITYFSA 150
gi|38352013|gb|AAR18687.1|        KSTGLPLYNIAIVWMDARTSPICRRLESLSGGRTHFVETCGLPITYFSA 145
gi|239049939|ref|NP_001132106.    KSTGRPLYNIAIVWMDARTSPVCRRELENELSGGRTHFVETCGLPITYFSA 148
gi|242077402|ref|XP_002448637.    KSTGHPLYNIAIVWMDARTSPVCRRESELSSGGRTHFVERCGLPITYFSA 148
                                     ****  *****.*.*.:*****.:*:*****  *****
LpGK2_P226_comp1791_c0_seq1      LKLLWLMENVNSVKDAVRAGDALFGTIDTWLIWNLTGGVDGN----- 192
LpGK2_F1_30_P205_AS_full          LKLLWLMENVDAVKDAVKA GDALFGTIDTWLIWNLTGGIGGKDTDGKELL 197
gi|357166356|ref|XP_003580683.    LKLLWLMENVDAVKDAVRAGDALFGTIDTWLIWNLTGGIGGKD---GKQI 197
gi|38352013|gb|AAR18687.1|        LKVLWLIENVDVKNNAVRAGDALFGTIDTWLIWNLTGGIGGTDTRDGNKVF 195
gi|239049939|ref|NP_001132106.    LKLLWLMENVDAVKDAVRTGDALFGTIDTWLIWNLTGGVAGG----- 190
gi|242077402|ref|XP_002448637.    LKLLWLMENVDAVKDAIKTGDALFGTIDTWLIWNLTGGVAGG----- 190
                                     **:***:***:***:***:***:***:*****:*****:
LpGK2_P226_comp1791_c0_seq1      GQHVTDCSNASRTMLMNLKALDWDKPTLEALGIPAGILPKIISNSEKIGV 242
LpGK2_F1_30_P205_AS_full          GQHVTDCSNASRTMLMNLKALDWDKPTLEALGIPVGILPKIISNSEKIGV 247
gi|357166356|ref|XP_003580683.    GQHVTDCSNASRTMLMNLKALDWDKPTLKGALGIPADILPKIISNSEKIGL 247
gi|38352013|gb|AAR18687.1|        GHVTDCSNASRTMLMNLKALDWDKPTLETGIPAEILPKIISNSERIGV 245
gi|239049939|ref|NP_001132106.    -QHVTDCSNASRTMLMNLKTLDWDKPTLAVLGVPVEILPKIISNSEKIGV 239
gi|242077402|ref|XP_002448637.    -QHVTDCSNASRTMLMNLKTLDWDKPTLDVLGVPVEILPKIISNSEKIGV 239
                                     :*****:*****.*:*.*.*****:***:
LpGK2_P226_comp1791_c0_seq1      VASGFPLAGVVISGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEE 292
LpGK2_F1_30_P205_AS_full          VASGFPLAGVVISGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEE 297
gi|357166356|ref|XP_003580683.    VADGFPLAGVPIISGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEE 297
gi|38352013|gb|AAR18687.1|        VANGFPLAGVPIAGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEE 295
gi|239049939|ref|NP_001132106.    VAKEFPFAGVPIISGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEE 289
gi|242077402|ref|XP_002448637.    VAKEFPFAGVPIISGCLGDQHAAMLGQLCKKGEAKSTYGTGAFILLNTGEE 289
                                     **.  **:*:*.*:*:*****:*****:*****:*****
LpGK2_P226_comp1791_c0_seq1      VTQSSHGLLSTIAYKLGPDVPTNYALEGSIAIAGAAVQWLRDSLGI ISSA 342
LpGK2_F1_30_P205_AS_full          VTQSSHGLLSTIAYKLGPDVPTNYALEGSIAIAGAAVQWLRDSLGI ISSA 347
gi|357166356|ref|XP_003580683.    VTQSSHGLLSTIAYKLGPSAPTNYALEGSIAIAGAAVQWLRDSLGI IRSA 347
gi|38352013|gb|AAR18687.1|        PTQSSHGLLSTIAYKLGPSAPTNYALEGSIAIAGAAVQWLRDSLGI ISTA 345
gi|239049939|ref|NP_001132106.    PTQSSHGLLSTIAYKLGPAAPTNYALEGSIAIAGAAVQWLRDSLGI IQSA 339
gi|242077402|ref|XP_002448637.    PTQSSHGLLSTIAYKLGPTAPTNYALEGSIAIAGAAVQWLRDSLGI IQTA 339
                                     *****.*.*****:*****:*****:
LpGK2_P226_comp1791_c0_seq1      AEIEGLAESVQDSGGIYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHI 392
LpGK2_F1_30_P205_AS_full          AEIEGLAESVQDSGGIYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHI 397
gi|357166356|ref|XP_003580683.    AEIEELAESVQDSGGIYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHI 397
gi|38352013|gb|AAR18687.1|        ADIEKLADTVQDSGGIYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHI 395
gi|239049939|ref|NP_001132106.    AEIEKLAETVPDSGGVYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHI 389
gi|242077402|ref|XP_002448637.    AEIEKLAETVPDSGGVYFVPAFNGLFAPWWRDDARGICIGITRFTNKGHI 389
                                     *:**  **::*  ****:*****:*****:*****:*****

```

Figure 7

```
LpGK2_P226_comp1791_c0_seq1      ARAVLESMCFQVNDVLSSMHKDAGEAGEVKS AEGEFLLRVDGGATVNNLL 442
LpGK2_F1_30_P205_AS_full          ARAVLESMCFQVNDVLSSMHKDAGEAGEVKS AEGEFLLRVDGGATVNNLL 447
gi|357166356|ref|XP_003580683.    ARAVLESMCFQVNDVLSSMHKDAGEAGEVKS AEGEFLLRVDGGATVNNLL 447
gi|38352013|gb|AAR18687.1|       ARAVLESMCFQVNDVLSSMHKDAGEAGEVKS AEGEFLLRVDGGATVNNLL 445
gi|239049939|ref|NP_001132106.    ARAVLESMCFQVNDVLSSMHKDAGEAGEVKS AEGEFLLRVDGGATVNNLL 439
gi|242077402|ref|XP_002448637.    ARAVLESMSFQVNDVLSSMHKDAGEAGEVKS AEGEFLLRVDGGATVNNLL 439
*****.*****

LpGK2_P226_comp1791_c0_seq1      MQIQADLLGSPVVRPADIET TALGAAYAAGLAAGVWTKEEIFAGLHKDNT 492
LpGK2_F1_30_P205_AS_full          MQIQADLLGSPVVRPADIET TALGAAYAAGLAAGVWTKEEIFAGLHKDNT 497
gi|357166356|ref|XP_003580683.    MQIQADLLGSPVVRPADIET TALGAAYAAGLAAGVWTKEEIFAGLHKDNT 497
gi|38352013|gb|AAR18687.1|       MQIQADLLGSPVVRPADIET TALGAAYAAGLAAGVWTKEEIFAGLHKDNT 495
gi|239049939|ref|NP_001132106.    MQIQADLLGSPVVRPADIET TALGAAYAAGLAAGVWTKEEIFAGLHKDNT 489
gi|242077402|ref|XP_002448637.    MQIQADLLGSPVVRPADIET TALGAAYAAGLAAGVWTKEEIFAGLHKDNT 489
*****:*****.***:**:*****:**

LpGK2_P226_comp1791_c0_seq1      TVFRPKLDDAHRKKRGDSWYKAVSR SFDLADLSL 526
LpGK2_F1_30_P205_AS_full          TVFRPKLDDAHRKKRGDSWYKAVSR SFDLADLSL 531
gi|357166356|ref|XP_003580683.    TVFRPKLDDAHRKKRGDSWYKAVSR SFDLADLSL 531
gi|38352013|gb|AAR18687.1|       RVFRPKLDEAHRKKRADSWYKAVSR SFDLADLSL 529
gi|239049939|ref|NP_001132106.    TVFRPKLDEAHRKKRADSWYKAVSR SFDLADLSL 523
gi|242077402|ref|XP_002448637.    TVFRPKLDEAHRKKRADSWYKAVSR SFDLADLSL 523
*****:***:**.*****
```

Figure 7 continued

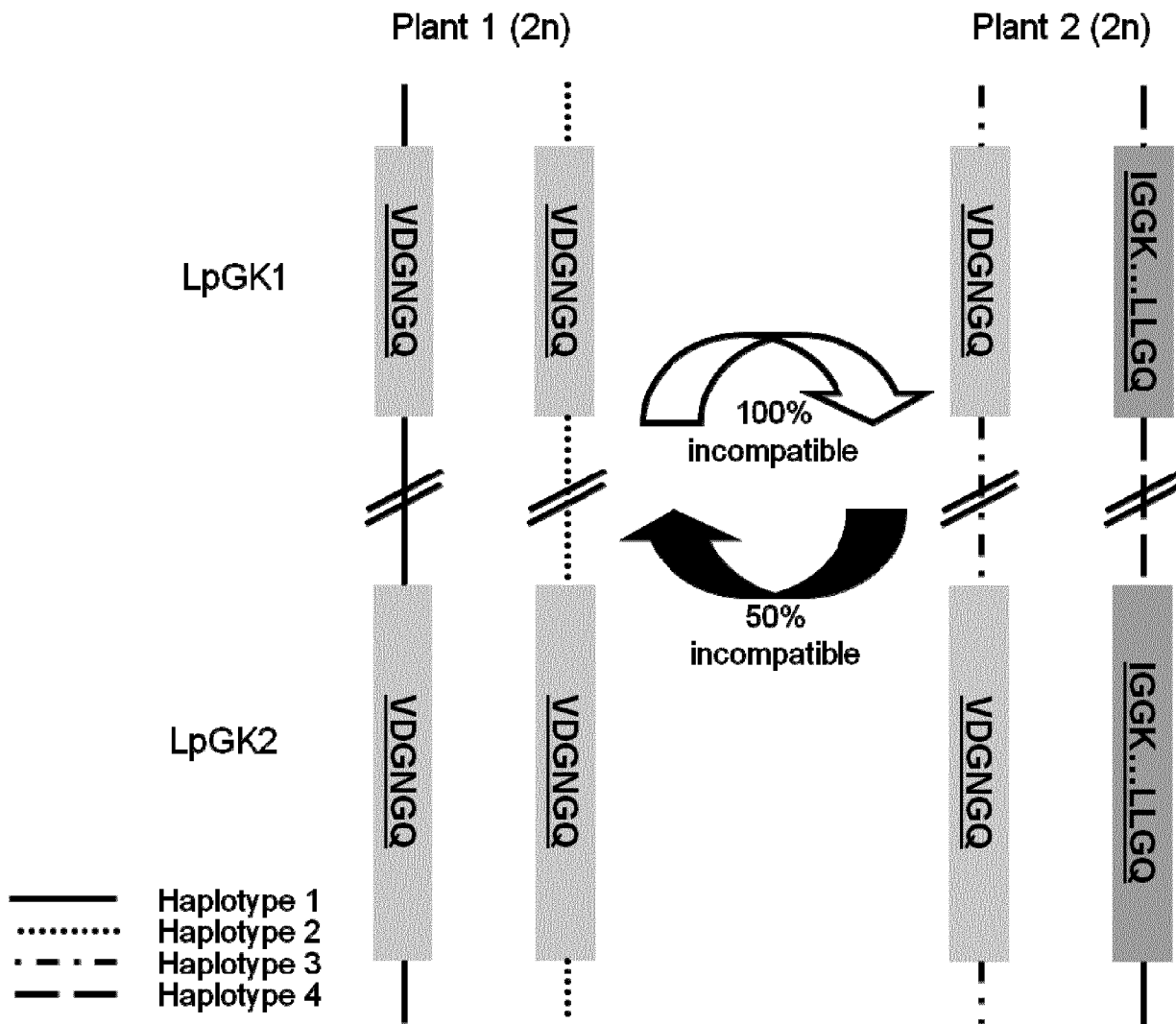


Figure 8

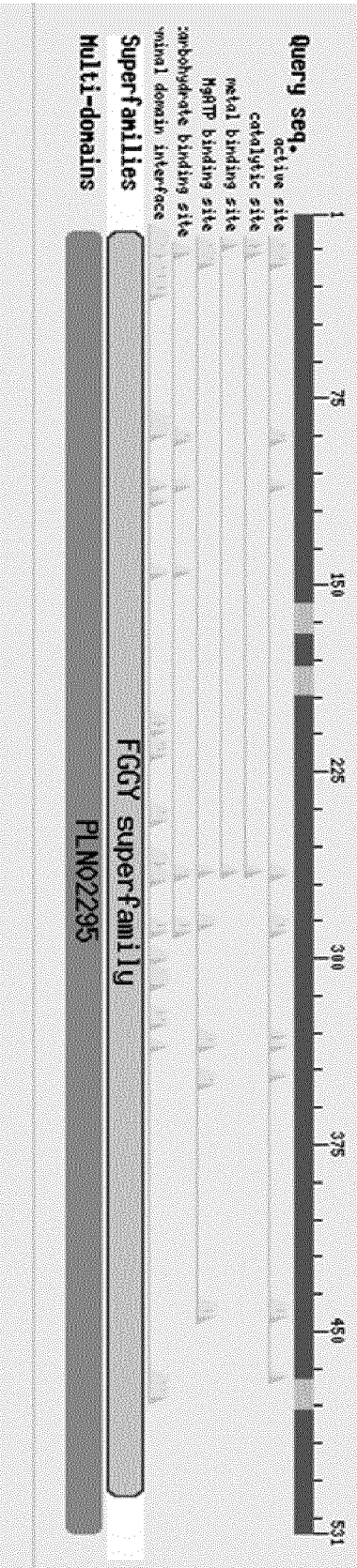


Figure 9

CLUSTAL 2.1 multiple sequence alignment

```

LpGK1_P205_cds      ATGGAGAATGTTGATGCTGTCAAGGATGCAGTCCGGGCTGGTGATGCATTATTTGGTACA 60
LpGK1_P226_cds      ATGGAGAATGTTGATGCTGTCAAGGATGCAGTCAAGTCTGGTGATGCATTATTTGGCACA 60
                      ***** * *****

LpGK1_P205_cds      ATTGACACCTGGTTGATCTGGAACCTCACAGGAGGCATTGGTGGCAAAGACACT--GATG 118
LpGK1_P226_cds      ATCGACACCTGGTTGATCTGGAACCTCACAAGAGGCATTGGTGGAAAAGACACTCTGATG 120
                      ** *****

LpGK1_P205_cds      GAAAAGAATTACTTGGCCAACATGTCACAGACTGCTCAAATGCATCTCGCACGATGCTTA 178
LpGK1_P226_cds      GAAAAGAATTACTTGGCCAACATGTCACAGACTGCTCAAATGCATCTCGCACGATGCTTA 180
                      *****

LpGK1_P205_cds      TGAATCTAAAGGCACTTGATTGGAACAAGCCTACACTTGAGGCATTAGGCATTCCTGCTG 238
LpGK1_P226_cds      TGAATCTAAAGGCACTTGATTGGGACAAGCCTACACTTGAGGCATTAGGCATTCCTGCTG 240
                      *****

LpGK1_P205_cds      GCATCTTGCCGAAGATTATCAGCAACTCAGAGAAAATTGGCGTGTTGCCAGTGGGTTTC 298
LpGK1_P226_cds      GCATCTTGCCGAAGATTATCAGCAACTCGGAGAAAATTGGTGTGGTTGCCACTGGGTTCC 300
                      *****

LpGK1_P205_cds      CTTTGGCTGGCGTCTCCATCTCAGGGTGTCTAGGAGACCAGCATGCTGCAATGCTCGGGC 358
LpGK1_P226_cds      CTTTGGCTGGCGTCTCCATCTCAGGGTGTCTAGGAGACCAGCATGCTGCAATGCTCGGGC 360
                      *****

LpGK1_P205_cds      AGCTATGCCAGAAAGGTGAAGCCAAAAGCACCTATGGAAGTGGTGCCTTCATCCTTCTGA 418
LpGK1_P226_cds      AGCTATGCCAGAAAGGTGAAGCCAAAAGCACCTACGGAAGTGGTGCCTTCATCCTTCTCA 420
                      *****

LpGK1_P205_cds      ACACAGGGGAGGAGGTTACCCAGTCCTCCCACGGTCTTCTTAGCACTATCGCTTACAAGC 478
LpGK1_P226_cds      ACACAGGGGAGGAGGTTACCCAGTCCTCCCAGTGGTCTTCTTAGCACCATTGCTTACAAGC 480
                      *****

LpGK1_P205_cds      TTGGCCCAGATGTACCCACCAATTATGCTCTAGAAGGCTCAATTGCAATTGCTGGTGCAG 538
LpGK1_P226_cds      TTGGCCCAGATGTACCCACCAATTATGCTCTAGAAGGCTCAACTGCAATTGCTGGTGCAG 540
                      *****

LpGK1_P205_cds      CAGTGCAGTGGTTGAGGGACAGCCTTGGGATCATTAGCTCAGCAGCAGAAATTGAAGGGT 598
LpGK1_P226_cds      CAGTTCAGTGGTTGAGGGACAGCCTTGGGATCATTAGGTCAGCAGCAGAA--TTGAAGGGT 599
                      **** *****

LpGK1_P205_cds      TGGCTGAAAGTGTGCAGGATTCAGGTGGSATATACTTTGTGCCAGCATTCAATGGATTGT 658
LpGK1_P226_cds      TGGCTGAAAGTGTGCAGGATTCAGGTGGGATATACTTCCTGCCAGCATTCAATGGATTGT 659
                      *****

LpGK1_P205_cds      TTGCACCATGGTGGAGAGATGATGCAAGAGGGATCTGCATAGGAATCACAAAGGTTACAA 718
LpGK1_P226_cds      TTGCACCATGGTGGAGGATGATGCTAGGGGTATCTGCATAGGAATCACAAAGGTTTACAA 719
                      *****

LpGK1_P205_cds      ATAAGGGGCACATTGCTCGGGCAGTGCTTGAGAGTATGTGCTTTCAGGTGAATGATGTCC 778
LpGK1_P226_cds      ACAAGGGGCACATTGCTCGAGCAGTGCTTGAGAGCATGTGTTTTTCAGGTGAACGATGTCC 779
                      * *****

LpGK1_P205_cds      TCAGCTCCATGCATAAGGATGCTGGGGAGGCAGGAGAAGTCAAGAGTGCAGAAGGGGAGT 838
LpGK1_P226_cds      TCAGCTCCATGCATAAGGATGCTGGGGAGGCAGGAGAAGTTAAGAGTGCAGAAGCGGAAT 839
                      *****

LpGK1_P205_cds      TCTTGTTGCGTGTTGATGGTGGTGCTACTGTTAATAACCTTCTCATGCAGATCCAGGCTG 898
LpGK1_P226_cds      TCTTGTTGCGTATTGATGGTGGTGCTACTGTAAACAACCATCTCATGCAGATCCAGGCTG 899
                      *****

```

Figure 10

CLUSTAL 2.1 multiple sequence alignment

```

LpGK1_P205_as  MENVDAVKDAVRAGDALFGTIDTWLIWNLTGGIGGKDTDGKELLGQHVTDCSNASRTMLM 60
LpGK1_P226_as  MENVDAVKDAVKSGDALFGTIDTWLIWNLTRGIGGKDT----LMEKNYLG--NMSQTAQM 54
                *****:*****          *****      *: :: .  * *: * *
                :

LpGK1_P205_as  NLKALDWNKPTLEALGIPAGILPKIISNSEKIGVVASGFPLAGVSISGCLGDQHAAMLGQ 120
LpGK1_P226_as  HL-----ARCL----- 60
                : *                               : **

LpGK1_P205_as  LCQKGEAKSTYGTGAFILLNTGEEVTQSSHGLLSTIAYKLGPDVPTNYALEGSIAIAGAA 180
LpGK1_P226_as  -----

LpGK1_P205_as  VQWLRDSLGIISSAAEIEGLAESVQDSGXIFYVPAFNGLFAPWWRDDARGICIGITRFTN 240
LpGK1_P226_as  -----

LpGK1_P205_as  KGHIARAVLESMCFQVNDVLSSMHKDAGEAGEVKSAEGEFLLRVDGGATVNNLLMQIQAD 300
LpGK1_P226_as  -----

LpGK1_P205_as  LLGSPVVRPADIETTALGAAYAAGLAAGVWTKDEIFAGLHKENTTVFRPKLDDAHRKKRG 360
LpGK1_P226_as  -----

LpGK1_P205_as  DSWYKAVSRSFDLADLSL 378
LpGK1_P226_as  -----

```

Figure 11

CLUSTAL 2.1 multiple sequence alignment

```

LpGK2_P226_as      MAGTAAEEEVYVASIDQGTSTRFIVYDRHARPVASHQLEFKQHYPEAGWVEHDPMEIIE 60
LpGK1_P226_as      -----

LpGK2_P226_as      SVKICMSETLGKAAASGLNVDAGLKAIGITNQRETTVMWSKSTGLPLYNAIVWMDVRTSS 120
LpGK1_P226_as      -----

LpGK2_P226_as      ICRRLESELSSGGRTHFVETCGLPISTYFSALKLLWLMENVNSVKDAVRAGDALFGTIDTW 180
LpGK1_P226_as      -----MENVDAVKDAVKSGDALFGTIDTW 24
                      *****:*****:*****

LpGK2_P226_as      LIWNLITGGVDGNGQHVTDCSNASRTMLMNLKALDWDKPTLEALGIPAGILPKIISNSEKI 240
LpGK1_P226_as      LIWNLTRGIGG-----KDTLMEKN----- 43
                      ***** *:.*          :  **:  :

LpGK2_P226_as      GVVASGFPLAGVSI SGCLGDQHAAMLGQLCQKGEAKSTYGTGAFILLNTGEEVTQSSHGL 300
LpGK1_P226_as      -----YLGNMSQTAQMHL----- 56
                      **:.*.*.:  :

LpGK2_P226_as      LSTIAYKLGPDPVTNYALEGSIAIAGAAVQWLRDSLGIISAAEIEGLAESVQDSGGIYF 360
LpGK1_P226_as      -----

LpGK2_P226_as      VPAFNGLFAPWWRDDARGICIGITRFTNKGHIARAVLESMCFQVNDVLSSMHKDAGEAGE 420
LpGK1_P226_as      -----

LpGK2_P226_as      VKSAEGEFLLRVDGGATVNNLLMQIQADLLGSPVVRPADIETTALGAAYAAGLAAGVWTK 480
LpGK1_P226_as      -----

LpGK2_P226_as      EEIFAGLHKDNTTVFRPKLDDAHRKKRGDSWYKAVSRSFDLADLSL 526
LpGK1_P226_as      -----

```

Figure 12

CLUSTAL 2.1 multiple sequence alignment

```

LpGK2_S089_cds      -----GTGTGGCCTACCAATCAGTACCTATTTTCAGTGCTCTTAAGTTATTATGGT  50
LpGK2_S027_cds      -----GTGTGGCCTACCAATCAGTACATATTTTCAGTGCTCTCAAGTTATTGTGGT  50
LpGK2_S021_cds      -----GTGTGGCCTACCAATCAGTACCTATTTTCAGTGCTCTCAAGTTATTATGGT  50
LpGK2_S065_cds      TTGTGGAGACGTGTGGCTTACCAATCAGTACATATTTTCAGTGCTCTGAAGTTATTATGGT  60
                      *****
LpGK2_S089_cds      TGATGGAGAATGTTGACGCTGTCAAGGATGCAGTCAAGGCTGGTGATGCATTATTTGGCA 110
LpGK2_S027_cds      TGATGGAGAATGTTAACGCTGTCAAGGATGCAGTCAAGGCCGGTGATGCATTATTTGGCA 110
LpGK2_S021_cds      TGATGGAGAATGTTGATGCTGTCAAGGATGCAGTCAAGGCTGATGATGCATTATTTGGCA 110
LpGK2_S065_cds      TGATGGAGAATGTTGACGCTGTAAAGGATGCAGTCAAGGCTGGTGATGCATTATTTGGCA 120
                      *****
LpGK2_S089_cds      CAATCGACACCTGGTTGATCTGGAATCTCACAGGAGGCATTGGTGGGAAAGATACTGATG 170
LpGK2_S027_cds      CAATCGACACCTGGTTGATCTGGAACCTCACAGGAGGCATTGGTGGGAAAGACACTGATG 170
LpGK2_S021_cds      CAATTGACACCTGGTTGATCTGGAACCTCACAGGAGGTGTTGACGGGAAT----- 160
LpGK2_S065_cds      CAATCGACACCTGGTTGATCTGGAACCTCACAGGAGGCATTGGTGGGAAAGACAATGATG 180
                      *****
LpGK2_S089_cds      GAAAAGAATTACTTGGGCAGCATGTTCACAGACTGCTCAAATGCATCCCGTACGATGCTTA 230
LpGK2_S027_cds      GAAAAGAATTACTTGGGCAACATGTTCACAGACTGCTCAAATGCATCTCGCACGATGCTTA 230
LpGK2_S021_cds      -----GGGCAGCATGTTCACAGACTGCTCAAATGCATCTCGTACCATGCTTA 206
LpGK2_S065_cds      GAAAAGAATTACTTGGGCAGCATGTTCACAGACTGCTCAAATGCATCTCGCACGATGCTTA 240
                      *****
LpGK2_S089_cds      TGAATCTAAAGGCACTTGATTGGGACAAGCCTACACTTGAGGCA 274
LpGK2_S027_cds      TGAATCTAAAGGCACTTGATTGGGACAAGCCTA-ACCTTGAGGCA 273
LpGK2_S021_cds      TGAATCTAAAGGCACTTGATTGGGACAAGCCT----- 238
LpGK2_S065_cds      TGAATCTAAAGGCACTTGATTGGGACAAGCC----- 271
                      *****

```

Figure 13a

CLUSTAL 2.1 multiple sequence alignment

```

LpGK2_S089_as      ---CGLPISTYFSALKLLWLMENVDVAVKDAVKAGDALFGTIDTWLIWNLTGGIGGKDTDG  57
LpGK2_S027_as      ---CGLPISTYFSALKLLWLMENVNAVKDAVKAGDALFGTIDTWLIWNLTGGIGGKDTDG  57
LpGK2_S065_as      VETCGLPISTYFSALKLLWLMENVDVAVKDAVKAGDALFGTIDTWLIWNLTGGIGGKDNDG  60
LpGK2_S021_as      ---CGLPISTYFSALKLLWLMENVDVAVKNAVKADDALFGTIDTWLIWNLTGGVDGN----  53
                      *****
LpGK2_S089_as      KELLGQHVTDCSNASRTMLMNLKALDWDKPTLEA 91
LpGK2_S027_as      KELLGQHVTDCSNASRTMLMNLKALDWDKPNLR- 90
LpGK2_S065_as      KELLGQHVTDCSNASRTMLMNLKALDWDK----- 89
LpGK2_S021_as      ----GQHVTDCSNASRTMLMNLKALDWDKP---- 79
                      *****

```

Figure 13b

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/067519

A. CLASSIFICATION OF SUBJECT MATTER
INV. 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)
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, WPI Data, Sequence Search, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HIROSHI SHINOZUKA ET AL: "Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (<i>Lolium perenne</i> L.)", PLANT MOLECULAR BIOLOGY, KLUWER ACADEMIC PUBLISHERS, DORDRECHT, NL, vol. 72, no. 3, 27 November 2009 (2009-11-27), pages 343-355, XP019773808, ISSN: 1573-5028	1
Y	page 354, column 1, paragraph 2; figures 2,3; table 1 ----- -/--	2-37, 43-47



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

19 November 2013

Date of mailing of the international search report

06/12/2013

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Krüger, Julia

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2013/067519

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	M. KLAAS ET AL: "Progress towards elucidating the mechanisms of self-incompatibility in the grasses: further insights from studies in Lolium", ANNALS OF BOTANY, vol. 108, no. 4, 1 September 2011 (2011-09-01), pages 677-685, XP055087527, ISSN: 0305-7364, DOI: 10.1093/aob/mcr186	1
Y	page 4, column 2, paragraph 3	2-37, 43-47
X	----- HACKAUF B ET AL: "Approaching the self-incompatibility locus Z in rye (Secale cereale L.) via comparative genetics", THEORETICAL AND APPLIED GENETICS ; INTERNATIONAL JOURNAL OF PLANT BREEDING RESEARCH, SPRINGER, BERLIN, DE, vol. 110, no. 5, 1 March 2005 (2005-03-01), pages 832-845, XP019321857, ISSN: 1432-2242, DOI: 10.1007/S00122-004-1869-4 abstract; figure 2	1
X	----- US 2006/107345 A1 (ALEXANDROV NICKOLAI [US] ET AL) 18 May 2006 (2006-05-18) sequence 36007	1-12,47
X	----- DATABASE UniProt [Online] 13 June 2012 (2012-06-13), "SubName: Full=Uncharacterized protein;", XP002716557, retrieved from EBI accession no. UNIPROT:I1J2K5 Database accession no. I1J2K5 sequence & VOGEL JOHN P ET AL: "Genome sequencing and analysis of the model grass Brachypodium distachyon", NATURE (LONDON), vol. 463, no. 7282, February 2010 (2010-02), pages 763-768, ISSN: 0028-0836	1-12,47
X	----- JP 2005 185101 A (NAT INST OF AGROBIO SCIENCES; SASAKI CO; RIKAGAKU KENKYUSHO; FOUND ADV) 14 July 2005 (2005-07-14) sequence 29805	1-12,47
X	----- US 2010/037352 A1 (ALEXANDROV NICKOLAI [US] ET AL) 11 February 2010 (2010-02-11) sequence 36009 ----- -/--	1-12,47

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2013/067519

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2009/094717 A1 (TROUKHAN MAXIM [US] ET AL) 9 April 2009 (2009-04-09) sequence 1300 -----	1-12,47
X	WO 2006/069610 A2 (METANOMICS GMBH [DE]; PLESCH GUNNAR [DE]; PUZIO PIOTR [DE]; BLAU ASTRI) 6 July 2006 (2006-07-06) sequence 68478 -----	1-12,47

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2013/067519

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.: 38-42
because they relate to subject matter not required to be searched by this Authority, namely:
Rule 39.1(ii) PCT - Essentially biological process for the production of plants
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2013/067519

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006107345 A1	18-05-2006	US 2006107345 A1	18-05-2006
		US 2010037352 A1	11-02-2010

JP 2005185101 A	14-07-2005	JP 2005185101 A	14-07-2005
		US 2006123505 A1	08-06-2006

US 2010037352 A1	11-02-2010	US 2006107345 A1	18-05-2006
		US 2010037352 A1	11-02-2010

US 2009094717 A1	09-04-2009	US 2009094717 A1	09-04-2009
		US 2013117886 A1	09-05-2013

WO 2006069610 A2	06-07-2006	AU 2005321630 A1	06-07-2006
		CA 2559760 A1	06-07-2006
		US 8541208 B1	24-09-2013
		WO 2006069610 A2	06-07-2006
