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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 October 2019 (24.10.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/203942 A1

(51) International Patent Classification:
A01H 1/04 (2006.01) *C12N 15/82* (2006.01)
A01H 5/10 (2018.01) *A01H 6/46* (2018.01)
C07K 14/415 (2006.01)

(21) International Application Number:
PCT/US2019/020898

(22) International Filing Date:
06 March 2019 (06.03.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/659,164 18 April 2018 (18.04.2018) US

(71) Applicant: **PIONEER HI-BRED INTERNATIONAL, INC.** [US/US]; 7100 NW 62nd Avenue, PO Box 1014, Johnston, Iowa 50131-1014 (US).

(72) Inventors: **GOEL, Ajay Kumar**; AB-402, Aparna Cyberzon, Nallagandla, Serilingampally, Hyderabad 500019 (IN). **KANUGANTI, Satish**; 7250 NW 62nd Avenue, PO Box 552, Johnston, Iowa 50131-0552 (US). **KUMAR, Ram**; 7250 NW 62nd Avenue, PO Box 552, Johnston, Iowa 501310552 (US). **MOHANTY, Amitabh**; Greenwood Residency, Apt C-217 Hislop Road, Kowkur, Hyderabad 50010 (IN). **TAVVA, Venkata S.**; 7250 NW 62nd Avenue, Po Box 552, Johnston, Iowa 50131-0552 (US). **VIJAYAKUMAR, Chm**; House No. 1-8-529, Manisai Residency, Flat No. 302, Hyderabad 500020 (IN).

(74) Agent: **BECKER, James E.**; 7250 NW 62nd Avenue, PO Box 552, Johnston, Iowa 50131-0552 (US).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** METHODS OF IDENTIFYING, SELECTING, AND PRODUCING BACTERIAL LEAF BLIGHT RESISTANT RICE

(57) **Abstract:** Compositions and methods useful in identifying and/or selecting rice plants that have bacterial leaf blight resistance are provided herein. The resistance may be newly conferred or enhanced relative to a control plant. The methods use rice markers on chromosome 7 to identify, select and/or construct resistant plants. Rice plants generated by the methods also provided.

WO 2019/203942 A1

METHODS OF IDENTIFYING, SELECTING, AND PRODUCING BACTERIAL LEAF BLIGHT RESISTANT RICE

FIELD

The field is related to plant breeding and methods of identifying and selecting plants with resistance to bacterial leaf blight. Provided are novel genes that encode proteins providing plant resistance to bacterial leaf blight. These proteins are useful in the production of transgenic or genetically modified bacterial leaf blight resistant plants.

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 62/659,164 filed April 18, 2018 which is herein incorporated by reference in their entirety.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named 7454_SeqList.txt created on January 24, 2018, has a size of 26.2 kilobytes, and is filed concurrently with the specification. The sequence listing contained in this ASCII formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND

Bacterial leaf blight (BLB) of rice, caused by *Xanthomonas oryzae*, is a significant disease that affects rice crop and results in yield losses. However, under severe infestation it can cause complete crop failure. Chemical control measures are ineffective and the plant native defense system is an option to control this disease. BLB is one of the most explored diseases in rice and with more than 40 known resistant genes, of which several of have been cloned and characterized. Diverse host plant resistance mechanisms are identified in rice for majority of these genes. Of known BLB resistant genes, Xa21, xa5 and xa13 genes are considered as effective genes across the strain spectrum. However, recent studies have shown the existence/evolution of new BLB strains which can break or act compatibly with BLB resistance genes to cause susceptibility.

Few chemicals such as copper compounds and antibiotics are available to combat the BLB disease but they are less effective and also not feasible for large scale fields. Host plant resistance proved to be the best option to control the disease The use of rice lines that carry genetic or transgenic sources of resistance is more practical if the genes responsible for resistance can be

incorporated into elite, high yielding germplasm without reducing yield.

Selection through the use of molecular markers associated with the bacterial leaf blight resistance trait allows selections based solely on the genetic composition of the progeny. As a result, plant breeding can occur more rapidly, thereby generating commercially acceptable rice plants. Thus, it is desirable to provide compositions and methods for identifying and selecting rice plants with newly conferred or enhanced bacterial leaf blight resistance. These plants can be used in breeding programs to generate high-yielding hybrids.

SUMMARY

Compositions and methods useful in identifying and selecting rice plants with bacterial leaf blight (BLB) resistance are provided herein. The methods use markers to identify and/or select resistant plants or to identify and/or counter-select susceptible plants. Rice plants having newly conferred or enhanced resistance to BLB relative to control plants are also provided herein.

In one embodiment, methods for identifying and/or selecting rice plants having resistance to BLB are presented. In these methods DNA of a rice plant is analyzed for the presence of a resistant gene allele on chromosome 7 that is associated with BLB resistance, wherein said resistant gene allele comprises: a "G" at R1091-7-06 (reference sequence SEQ ID NO: 10), a "G" at R1091-7-021 (reference sequence SEQ ID NO: 11), a "C" at R1091-7-022 (reference sequence SEQ ID NO: 12), a "G" at R1091-7-027 (reference sequence SEQ ID NO: 13), a "G" at R1091-7-028 (reference sequence SEQ ID NO: 14), a "G" at R1091-7-026 (reference sequence SEQ ID NO: 15), an "T" at R1091-7-042 (reference sequence SEQ ID NO: 16), a "A" at R1091-7-037 (reference sequence SEQ ID NO: 17), a "C" at R1091-7-030 (reference sequence SEQ ID NO: 18), a "G" at R1091-7-018 (reference sequence SEQ ID NO: 19), a "G" at R1091-7-053 (reference sequence SEQ ID NO: 20), a "C" at R1091-7-052 (reference sequence SEQ ID NO: 21), a "G" at R1091-7-040 (reference sequence SEQ ID NO: 22), a "T" at R1091-7-044 (reference sequence SEQ ID NO: 23), a "A" at R1091-7-039 (reference sequence SEQ ID NO: 24), a "C" at R1091-7-024 (reference sequence SEQ ID NO: 25), and a "G" at R1091-7-007 (reference sequence SEQ ID NO: 26); and a rice plant is identified and/or selected as having BLB resistance if said resistant gene allele is detected. In another embodiment, identifying and/or selecting rice plants having resistance to BLB comprises detecting the presence of a chromosomal interval between the flanking markers R1091-7-06 and R1091-7-007. In another embodiment, identifying and/or selecting rice plants having resistance to BLB comprises detecting the presence of an "T" at R1091-7-042 (reference sequence SEQ ID NO: 16) in the coding sequence of the BLB4 resistance gene. The selected rice plant may be crossed to a second rice plant in order to obtain a progeny plant that has the resistant gene allele. In some embodiments, the methods for identifying and/or selecting rice plants having resistance to BLB

comprise detecting or selecting a genomic region comprising SEQ ID NO: 1. The BLB resistance may be newly conferred or enhanced relative to a control plant that does not have the favorable allele. The resistant gene allele may be further refined to a chromosomal interval defined by and including markers R1091-7-06 and R1091-7-007. The analyzing step may be performed by isolating nucleic acids and detecting one or more marker alleles linked to and associated with the resistant gene allele. In a further embodiment, the BLB resistant region comprises a gene encoding a BLB4 polypeptide that confers or enhances resistance to *Xanthomonas oryzae* or BLB. In some embodiments, the BLB4 polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 3.

In another embodiment, methods of identifying and/or selecting plants with BLB resistance are provided in which one or more marker alleles linked to and associated with any of: a "G" at R1091-7-06, a "G" at R1091-7-021, a "C" at R1091-7-022, a "G" at R1091-7-027, a "G" at R1091-7-028, a "G" at R1091-7-026, an "T" at R1091-7-42, a "A" at R1091-7-037, a "C" at R1091-7-030, a "G" at R1091-7-018, a "G" at R1091-7-053, a "C" at R1091-7-052, a "G" at R1091-7-040, a "T" at R1091-7-044, a "A" at R1091-7-039, a "C" at R1091-7-024, and a "G" at R1091-7-007, are detected in a plant, and a plant having the one or more marker alleles is selected. The one or more marker alleles may be linked by 10 cM, 9 cM, 8 cM, 7 cM, 6 cM, 5 cM, 4 cM, 3 cM, 2 cM, 1 cM, 0.9 cM, 0.8 cM, 0.7 cM, 0.6 cM, 0.5 cM, 0.4 cM, 0.3 cM, 0.2 cM, or 0.1 cM or less on a single meiosis based genetic map. The selected plant may be crossed to a second plant to obtain a progeny plant that has one or more marker alleles linked to and associated with any of: a "G" at R1091-7-06, a "G" at R1091-7-021, a "C" at R1091-7-022, a "G" at R1091-7-027, a "G" at R1091-7-028, a "G" at R1091-7-026, an "T" at R1091-7-42, a "A" at R1091-7-037, a "C" at R1091-7-030, a "G" at R1091-7-018, a "G" at R1091-7-053, a "C" at R1091-7-052, a "G" at R1091-7-040, a "T" at R1091-7-044, a "A" at R1091-7-039, a "C" at R1091-7-024, and a "G" at R1091-7-007.

In another embodiment, methods of introgressing a gene allele associated with BLB resistance are presented herein. In these methods, a population of rice plants is screened with one or more markers to determine if any of the rice plants has a gene allele associated with BLB resistance, and at least one rice plant that has the gene allele associated with BLB resistance is selected from the population. The gene allele comprises a "G" at R1091-7-06, a "G" at R1091-7-021, a "C" at R1091-7-022, a "G" at R1091-7-027, a "G" at R1091-7-028, a "G" at R1091-7-026, an "T" at R1091-7-42, a "A" at R1091-7-037, a "C" at R1091-7-030, a "G" at R1091-7-018, a "G" at R1091-7-053, a "C" at R1091-7-052, a "G" at R1091-7-040, a "T" at R1091-7-044, a "A" at R1091-7-039, a "C" at R1091-7-024, and a "G" at R1091-7-007.

In some embodiments, introgression of BLB resistant genes from resistant to susceptible lines may be achieved either by marker-assisted trait introgression, transgenic, or genome editing approaches.

Embodiments include an isolated polynucleotide comprising a nucleotide sequence encoding a BLB4 polypeptide capable of conferring resistance to BLB, wherein the BLB4 polypeptide has an amino acid sequence of at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, and at least 95% identity, when compared to SEQ ID NO:3. In another embodiment, an isolated polynucleotide comprises a promoter region and a nucleotide sequence encoding a BLB4 polypeptide capable of conferring resistance to BLB, wherein the BLB4 polypeptide has an amino acid sequence of at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, and at least 95% identity, when compared to SEQ ID NO:3, and wherein the promoter region has a nucleic acid sequence of at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, and at least 95% identity, when compared to SEQ ID NO:4.

Additional embodiments of the present disclosure include a vector comprising a polynucleotide of the disclosure, such as SEQ ID NO: 2, or a recombinant DNA construct comprising a polynucleotide disclosed herein operably linked to at least one regulatory sequence. A plant cell, as well as a plant, each comprising the recombinant DNA construct of an embodiment disclosed herein, and a seed comprising the recombinant DNA construct are also embodied.

The methods embodied by the present disclosure relate to 1) a method for transforming a host cell, including a plant cell, comprising transforming the host cell with the polynucleotide of an embodiment of the present disclosure, 2) a method for producing a plant comprising transforming a plant cell with the recombinant DNA construct of an embodiment of the present disclosure and regenerating a plant from the transformed plant cell, and 3) methods of conferring or enhancing resistance to BLB, comprising transforming a plant with the recombinant DNA construct disclosed herein, thereby conferring and/or enhancing resistance to *Xanthomonas oryzae* or BLB.

Methods of altering the level of expression of a protein capable of conferring resistance to *Xanthomonas oryzae* or BLB in a plant or plant cell comprising (a) transforming a plant cell with a recombinant DNA construct disclosed herein and (b) growing the transformed plant cell under conditions that are suitable for expression of the recombinant DNA construct wherein expression of the recombinant DNA construct results in production of altered levels of a protein capable of conferring resistance to *Xanthomonas oryzae* or BLB in the transformed host are also embodied.

Rice plants identified and/or selected using any of the methods presented above are also provided.

DESCRIPTION OF THE DRAWINGS

FIG. 1 shows (A) a 4.7 kb construct containing BLB4 (1.19 kb) and a 3.0kb genomic region from PRA1091 containing putative promoter and 500bp putative terminator region for overexpression, and (B) the target site selected to create targeted mutations in TATA box with CCA as the PAM site, and the predicted cut site has been underlined.

DETAILED DESCRIPTION

As used herein the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the protein" includes reference to one or more proteins and equivalents thereof, and so forth. All technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs unless clearly indicated otherwise.

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

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

The phrase “closely linked”, in the present application, means that recombination between two linked loci occurs with a frequency of equal to or less than about 10% (i.e., are separated on a genetic map by not more than 10 cM). Put another way, the closely linked loci co-segregate at least 90% of the time. Marker loci are especially useful with respect to the subject matter of the current disclosure when they demonstrate a significant probability of co-segregation (linkage) with a desired trait (e.g., resistance to bacterial leaf blight). Closely linked loci such as a marker locus and a second locus can display an inter-locus recombination frequency of 10% or less, preferably about 9% or less, still more preferably about 8% or less, yet more preferably about 7% or less, still more preferably about 6% or less, yet more preferably about 5% or less, still more preferably about 4% or less, yet more preferably about 3% or less, and still more preferably about 2% or less. In highly preferred embodiments, the relevant loci display a recombination a frequency of about 1% or less, e.g., about 0.75% or less, more preferably about 0.5% or less, or yet more preferably about 0.25% or less. Two

loci that are localized to the same chromosome, and at such a distance that recombination between the two loci occurs at a frequency of less than 10% (e.g., about 9 %, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25%, or less) are also said to be "proximal to" each other. In some cases, two different markers can have the same genetic map coordinates. In that case, the two markers are in such close proximity to each other that recombination occurs between them with such low frequency that it is undetectable.

When referring to the relationship between two genetic elements, such as a genetic element contributing to BLB resistance and a proximal marker, "coupling" phase linkage indicates the state where the "favorable" allele at the BLB resistance locus is physically associated on the same chromosome strand as the "favorable" allele of the respective linked marker locus. In coupling phase, both favorable alleles are inherited together by progeny that inherit that chromosome strand.

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

A plant referred to herein as a "doubled haploid" is developed by doubling the haploid set of chromosomes (i.e., half the normal number of chromosomes). A doubled haploid plant has two identical sets of chromosomes, and all loci are considered homozygous.

An "elite line" is any line that has resulted from breeding and selection for superior agronomic performance.

An "exotic rice strain" or an "exotic rice germplasm" is a strain derived from a rice plant not belonging to an available elite rice line or strain of germplasm. In the context of a cross between two rice plants or strains of germplasm, an exotic germplasm is not closely related by descent to the elite germplasm with which it is crossed. Most commonly, the exotic germplasm is not derived from any known elite line of rice, but rather is selected to introduce novel genetic elements (typically novel alleles) into a breeding program.

A "favorable allele" is the allele at a particular locus (a marker, a QTL, a gene etc.) that confers, or contributes to, an agronomically desirable phenotype, e.g., BLB resistance, and that allows the identification of plants with that agronomically desirable phenotype. A favorable allele of a marker is a marker allele that segregates with the favorable phenotype.

"Genetic markers" are nucleic acids that are polymorphic in a population and where the alleles of which can be detected and distinguished by one or more analytic methods, e.g., RFLP, AFLP, isozyme, SNP, SSR, and the like. The term also refers to nucleic acid sequences complementary to the genomic sequences, such as nucleic acids used as probes. Markers corresponding to genetic polymorphisms between members of a population can be detected by

methods well-established in the art. These include, e.g., PCR-based sequence specific amplification methods, detection of restriction fragment length polymorphisms (RFLP), detection of isozyme markers, detection of polynucleotide polymorphisms by allele specific hybridization (ASH), detection of amplified variable sequences of the plant genome, detection of self-sustained sequence replication, detection of simple sequence repeats (SSRs), detection of single nucleotide polymorphisms (SNPs), or detection of amplified fragment length polymorphisms (AFLPs). Well established methods are also known for the detection of expressed sequence tags (ESTs) and SSR markers derived from EST sequences and randomly amplified polymorphic DNA (RAPD).

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

A "haplotype" is the genotype of an individual at a plurality of genetic loci, i.e. a combination of alleles. Typically, the genetic loci described by a haplotype are physically and genetically linked, i.e., on the same chromosome segment.

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

The heterotic response of material, or "heterosis", can be defined by performance which exceeds the average of the parents (or high parent) when crossed to other dissimilar or unrelated groups.

A "heterotic group" comprises a set of genotypes that perform well when crossed with genotypes from a different heterotic group (Hallauer et al. (1998) Corn breeding, p. 463-564. In G.F. Sprague and J.W. Dudley (ed.) *Corn and corn improvement*). Inbred lines are classified into heterotic groups, and are further subdivided into families within a heterotic group, based on several criteria such as pedigree, molecular marker-based associations, and performance in hybrid combinations (Smith et al. (1990) *Theor. Appl. Gen.* 80:833-840). The two most widely used heterotic groups in the United States are referred to as "Iowa Stiff Stalk Synthetic" (also referred to herein as "stiff stalk") and "Lancaster" or "Lancaster Sure Crop" (sometimes referred to as NSS, or non-Stiff Stalk).

Some heterotic groups possess the traits needed to be a female parent, and others, traits for a male parent. For example, in maize, yield results from public inbreds released from a population

called BSSS (Iowa Stiff Stalk Synthetic population) has resulted in these inbreds and their derivatives becoming the female pool in the central Corn Belt. BSSS inbreds have been crossed with other inbreds, e.g. SD 105 and Maiz Amargo, and this general group of materials has become known as Stiff Stalk Synthetics (SSS) even though not all of the inbreds are derived from the original BSSS population (Mikel and Dudley (2006) *Crop Sci*: 46:1193-1205). By default, all other inbreds that combine well with the SSS inbreds have been assigned to the male pool, which for lack of a better name has been designated as NSS, i.e. Non-Stiff Stalk. This group includes several major heterotic groups such as Lancaster Surecrop, Iodent, and Leaming Corn.

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

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

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

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

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

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

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

As used herein, the term "linkage" is used to describe the degree with which one marker locus is associated with another marker locus or some other locus. The linkage relationship between a molecular marker and a locus affecting a phenotype is given as a "probability" or "adjusted

probability". Linkage can be expressed as a desired limit or range. For example, in some embodiments, any marker is linked (genetically and physically) to any other marker when the markers are separated by less than 50, 40, 30, 25, 20, or 15 map units (or cM) of a single meiosis map (a genetic map based on a population that has undergone one round of meiosis, such as e.g. an F_2; the IBM2 maps consist of multiple meiosis). In some aspects, it is advantageous to define a bracketed range of linkage, for example, between 10 and 20 cM, between 10 and 30 cM, or between 10 and 40 cM. The more closely a marker is linked to a second locus, the better an indicator for the second locus that marker becomes. Thus, "closely linked loci" such as a marker locus and a second locus display an inter-locus recombination frequency of 10% or less, preferably about 9% or less, still more preferably about 8% or less, yet more preferably about 7% or less, still more preferably about 6% or less, yet more preferably about 5% or less, still more preferably about 4% or less, yet more preferably about 3% or less, and still more preferably about 2% or less. In highly preferred embodiments, the relevant loci display a recombination frequency of about 1% or less, e.g., about 0.75% or less, more preferably about 0.5% or less, or yet more preferably about 0.25% or less. Two loci that are localized to the same chromosome, and at such a distance that recombination between the two loci occurs at a frequency of less than 10% (e.g., about 9 %, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25%, or less) are also said to be "in proximity to" each other. Since one cM is the distance between two markers that show a 1% recombination frequency, any marker is closely linked (genetically and physically) to any other marker that is in close proximity, e.g., at or less than 10 cM distant. Two closely linked markers on the same chromosome can be positioned 9, 8, 7, 6, 5, 4, 3, 2, 1, 0.75, 0.5 or 0.25 cM or less from each other.

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

Linkage disequilibrium is most commonly assessed using the measure r^2, which is calculated using the formula described by Hill, W.G. and Robertson, A, Theor. Appl. Genet. 38:226-231(1968). When $r^2 = 1$, complete LD exists between the two marker loci, meaning that the markers have not

been separated by recombination and have the same allele frequency. The r^2 value will be dependent on the population used. Values for r^2 above 1/3 indicate sufficiently strong LD to be useful for mapping (Ardlie et al., Nature Reviews Genetics 3:299-309 (2002)). Hence, alleles are in linkage disequilibrium when r^2 values between pairwise marker loci are greater than or equal to 0.33, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1.0.

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

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

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

"Rice" refers to a plant of *Oryza sativa.*

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

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

particular trait in human genetics (e.g. a marker for breast cancer). The term marker locus is the locus (gene, sequence or nucleotide) that the marker detects.

Markers that detect genetic polymorphisms between members of a population are well-established in the art. Markers can be defined by the type of polymorphism that they detect and also the marker technology used to detect the polymorphism. Marker types include but are not limited to, e.g., detection of restriction fragment length polymorphisms (RFLP), detection of isozyme markers, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), detection of simple sequence repeats (SSRs), detection of amplified variable sequences of the plant genome, detection of self-sustained sequence replication, or detection of single nucleotide polymorphisms (SNPs). SNPs can be detected e.g. via DNA sequencing, PCR-based sequence specific amplification methods, detection of polynucleotide polymorphisms by allele specific hybridization (ASH), dynamic allele-specific hybridization (DASH), molecular beacons, microarray hybridization, oligonucleotide ligase assays, Flap endonucleases, 5' endonucleases, primer extension, single strand conformation polymorphism (SSCP) or temperature gradient gel electrophoresis (TGGE). DNA sequencing, such as the pyrosequencing technology has the advantage of being able to detect a series of linked SNP alleles that constitute a haplotype. Haplotypes tend to be more informative (detect a higher level of polymorphism) than SNPs.

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

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

"Marker assisted counter-selection" is a process by which marker genotypes are used to identify plants that will not be selected, allowing them to be removed from a breeding program or planting.

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

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

The term "molecular marker" may be used to refer to a genetic marker, as defined above, or an encoded product thereof (e.g., a protein) used as a point of reference when identifying a linked locus. A marker can be derived from genomic nucleotide sequences or from expressed nucleotide sequences (e.g., from a spliced RNA, a cDNA, etc.), or from an encoded polypeptide. The term also refers to nucleic acid sequences complementary to or flanking the marker sequences, such as nucleic acids used as probes or primer pairs capable of amplifying the marker sequence. A "molecular

marker probe" is a nucleic acid sequence or molecule that can be used to identify the presence of a marker locus, e.g., a nucleic acid probe that is complementary to a marker locus sequence. Alternatively, in some aspects, a marker probe refers to a probe of any type that is able to distinguish (i.e., genotype) the particular allele that is present at a marker locus. Nucleic acids are "complementary" when they specifically hybridize in solution. Some of the markers described herein are also referred to as hybridization markers when located on an indel region, such as the non-collinear region described herein. This is because the insertion region is, by definition, a polymorphism *vis a vis* a plant without the insertion. Thus, the marker need only indicate whether the indel region is present or absent. Any suitable marker detection technology may be used to identify such a hybridization marker, e.g. SNP technology is used in the examples provided herein.

An allele “negatively” correlates with a trait when it is linked to it and when presence of the allele is an indicator that a desired trait or trait form will not occur in a plant comprising the allele.

The term “phenotype”, “phenotypic trait”, or “trait” can refer to the observable expression of a gene or series of genes. The phenotype can be observable to the naked eye, or by any other means of evaluation known in the art, e.g., weighing, counting, measuring (length, width, angles, etc.), microscopy, biochemical analysis, or an electromechanical assay. In some cases, a phenotype is directly controlled by a single gene or genetic locus, i.e., a “single gene trait” or a “simply inherited trait”. In the absence of large levels of environmental variation, single gene traits can segregate in a population to give a “qualitative” or “discrete” distribution, i.e. the phenotype falls into discrete classes. In other cases, a phenotype is the result of several genes and can be considered a “multigenic trait” or a “complex trait”. Multigenic traits segregate in a population to give a “quantitative” or “continuous” distribution, i.e. the phenotype cannot be separated into discrete classes. Both single gene and multigenic traits can be affected by the environment in which they are being expressed, but multigenic traits tend to have a larger environmental component.

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

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

A “production marker” or “production SNP marker” is a marker that has been developed for high-throughput purposes. Production SNP markers are developed to detect specific polymorphisms

and are designed for use with a variety of chemistries and platforms. The marker names used here begin with a PHM prefix to denote 'Pioneer Hi-Bred Marker', followed by a number that is specific to the sequence from which it was designed, followed by a "." or a "-" and then a suffix that is specific to the DNA polymorphism. A marker version can also follow (A, B, C etc.) that denotes the version of the marker designed to that specific polymorphism.

The term "quantitative trait locus" or "QTL" refers to a region of DNA that is associated with the differential expression of a quantitative phenotypic trait in at least one genetic background, e.g., in at least one breeding population. The region of the QTL encompasses or is closely linked to the gene or genes that affect the trait in question.

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

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

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

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

The term "yield" refers to the productivity per unit area of a particular plant product of commercial value. Yield is affected by both genetic and environmental factors. "Agronomics", "agronomic traits", and "agronomic performance" refer to the traits (and underlying genetic elements) of a given plant variety that contribute to yield over the course of growing season. Individual agronomic traits include emergence vigor, vegetative vigor, stress tolerance, disease resistance or tolerance, herbicide resistance, branching, flowering, seed set, seed size, seed density, standability, threshability and the like. Yield is, therefore, the final culmination of all agronomic traits.

Rice marker loci that demonstrate statistically significant co-segregation with a resistance traits that confers broad resistance against bacterial leaf blight (BLB), and a resistance gene, BLB4,

are provided herein. Detection of these loci or additional linked loci and the resistance gene may be used in marker assisted selection as part of a rice breeding program to produce rice plants that have resistance to BLB.

Genetic mapping

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

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

Association Mapping

Understanding the extent and patterns of linkage disequilibrium (LD) in the genome is a prerequisite for developing efficient association approaches to identify and map quantitative trait loci (QTL). Linkage disequilibrium (LD) refers to the non-random association of alleles in a collection of individuals. When LD is observed among alleles at linked loci, it is measured as LD decay across a specific region of a chromosome. The extent of the LD is a reflection of the recombinational history of that region. The average rate of LD decay in a genome can help predict the number and density of markers that are required to undertake a genome-wide association study and provides an estimate of the resolution that can be expected.

Association or LD mapping aims to identify significant genotype-phenotype associations. It has been exploited as a powerful tool for fine mapping in outcrossing species such as humans (Corder et al. (1994) "Protective effect of apolipoprotein-E type-2 allele for late-onset Alzheimer-disease," *Nat Genet* 7:180-184; Hastbacka et al. (1992) "Linkage disequilibrium mapping in isolated

founder populations: diastrophic dysplasia in Finland," *Nat Genet* 2:204-211; Kerem et al. (1989) "Identification of the cystic fibrosis gene: genetic analysis," *Science* 245:1073-1080) and maize (Remington et al., (2001) "Structure of linkage disequilibrium and phenotype associations in the maize genome," *Proc Natl Acad Sci* USA 98:11479-11484; Thornsberry et al. (2001) "*Dwarf8* polymorphisms associate with variation in flowering time," *Nat Genet* 28:286-289; reviewed by Flint-Garcia et al. (2003) "Structure of linkage disequilibrium in plants," *Annu Rev Plant Biol.* 54:357-374), where recombination among heterozygotes is frequent and results in a rapid decay of LD. In inbreeding species where recombination among homozygous genotypes is not genetically detectable, the extent of LD is greater (i.e., larger blocks of linked markers are inherited together) and this dramatically enhances the detection power of association mapping (Wall and Pritchard (2003) "Haplotype blocks and linkage disequilibrium in the human genome," *Nat Rev Genet* 4:587-597).

The recombinational and mutational history of a population is a function of the mating habit as well as the effective size and age of a population. Large population sizes offer enhanced possibilities for detecting recombination, while older populations are generally associated with higher levels of polymorphism, both of which contribute to observably accelerated rates of LD decay. On the other hand, smaller effective population sizes, e.g., those that have experienced a recent genetic bottleneck, tend to show a slower rate of LD decay, resulting in more extensive haplotype conservation (Flint-Garcia et al. (2003) "Structure of linkage disequilibrium in plants," *Annu Rev Plant Biol.* 54:357-374).

Elite breeding lines provide a valuable starting point for association analyses. Association analyses use quantitative phenotypic scores (e.g., disease tolerance rated from one to nine for each line) in the analysis (as opposed to looking only at tolerant versus resistant allele frequency distributions in intergroup allele distribution types of analysis). The availability of detailed phenotypic performance data collected by breeding programs over multiple years and environments for a large number of elite lines provides a valuable dataset for genetic marker association mapping analyses. This paves the way for a seamless integration between research and application and takes advantage of historically accumulated data sets. However, an understanding of the relationship between polymorphism and recombination is useful in developing appropriate strategies for efficiently extracting maximum information from these resources.

This type of association analysis neither generates nor requires any map data, but rather is independent of map position. This analysis compares the plants' phenotypic score with the genotypes at the various loci. Subsequently, any suitable map (for example, a composite map) can optionally be used to help observe distribution of the identified QTL markers and/or QTL marker clustering using previously determined map locations of the markers.

Traditional linkage analysis

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

Rice marker loci that demonstrate statistically significant co-segregation with a BLB resistance trait, as determined by traditional linkage analysis and by whole genome association analysis, are provided herein. Detection of these loci or additional linked loci can be used in marker assisted rice breeding programs to produce plants having resistance to bacterial leaf blight.

Activities in marker assisted rice breeding programs may include but are not limited to: selecting among new breeding populations to identify which population has the highest frequency of favorable nucleic acid sequences based on historical genotype and agronomic trait associations, selecting favorable nucleic acid sequences among progeny in breeding populations, selecting among parental lines based on prediction of progeny performance, and advancing lines in germplasm improvement activities based on presence of favorable nucleic acid sequences.

Chromosomal intervals

Chromosomal intervals that correlate with the BLB resistance trait are provided. A variety of methods well known in the art are available for identifying chromosomal intervals. The boundaries of such chromosomal intervals are drawn to encompass markers that will be linked to the gene(s) controlling the trait of interest. In other words, the chromosomal interval is drawn such that any marker that lies within that interval (including the terminal markers that define the boundaries of the interval) can be used as a marker for BLB disease resistance trait. Table 2 identify markers within the chromosome 7 genomic region that were shown herein to associate with the BLB resistance trait and that are linked to a gene(s) controlling BLB resistance. Reference sequences for each of the markers are represented by SEQ ID NOs:10-26.

Conversely, e.g., if two markers in close proximity show co-segregation with the desired phenotypic trait, it is sometimes unclear if each of those markers identify the same gene or two different gene or multiple genes. Regardless, knowledge of how many genes are in a particular

physical/genomic interval is not necessary to make or practice that which is presented in the current disclosure.

The chromosome 7 interval may encompass any of the markers identified herein as being associated with the BLB resistance trait including: a "G" at R1091-7-06, a "G" at R1091-7-021, a "C" at R1091-7-022, a "G" at R1091-7-027, a "G" at R1091-7-028, a "G" at R1091-7-026, an "T" at R1091-7-42, a "A" at R1091-7-037, a "C" at R1091-7-030, a "G" at R1091-7-018, a "G" at R1091-7-053, a "C" at R1091-7-052, a "G" at R1091-7-040, a "T" at R1091-7-044, a "A" at R1091-7-039, a "C" at R1091-7-024, and a "G" at R1091-7-007. The chromosome 7 interval, for example, may be defined by markers at R1091-7-06 and R1091-7-007, a further subinterval of which can be defined by the CDS marker R1091-7-42. Any marker located within these intervals can find use as a marker for BLB resistance and can be used in the context of the methods presented herein to identify and/or select rice plants that have resistance to BLB, whether it is newly conferred or enhanced compared to a control plant. In certain embodiments, markers located upstream and downstream of BLB4 gene position are very tightly linked genetically and physically and hence may be used to select the BLB4 gene for trait introgression and products development.

Chromosomal intervals can also be defined by markers that are linked to (show linkage disequilibrium with) BLB resistant gene, and r^2 is a common measure of linkage disequilibrium (LD) in the context of association studies. If the r^2 value of LD between a chromosome 7 marker locus in an interval of interest and another chromosome 7 marker locus in close proximity is greater than 1/3 (Ardlie et al., Nature Reviews Genetics 3:299-309 (2002)), the loci are in linkage disequilibrium with one another.

Markers and linkage relationships

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

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

The closer a marker is to a gene controlling a trait of interest, the more effective and advantageous that marker is as an indicator for the desired trait. Closely linked loci display an inter-locus cross-over frequency of about 10% or less, preferably about 9% or less, still more preferably

about 8% or less, yet more preferably about 7% or less, still more preferably about 6% or less, yet more preferably about 5% or less, still more preferably about 4% or less, yet more preferably about 3% or less, and still more preferably about 2% or less. In highly preferred embodiments, the relevant loci (e.g., a marker locus and a target locus) display a recombination frequency of about 1% or less, e.g., about 0.75% or less, more preferably about 0.5% or less, or yet more preferably about 0.25% or less. Thus, the loci are about 10 cM, 9 cM, 8 cM, 7 cM, 6 cM, 5 cM, 4 cM, 3 cM, 2 cM, 1 cM, 0.75 cM, 0.5 cM or 0.25 cM or less apart. Put another way, two loci that are localized to the same chromosome, and at such a distance that recombination between the two loci occurs at a frequency of less than 10% (e.g., about 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25%, or less) are said to be "proximal to" each other.

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

Methods presented herein include detecting the presence of one or more marker alleles associated with BLB resistance in a plant and then identifying and/or selecting plants that have favorable alleles at those marker loci. Markers have been identified herein as being associated with the BLB resistance trait and hence can be used to predict BLB in a plant. Any marker within 50 cM, 40 cM, 30 cM, 20 cM, 15 cM, 10 cM, 9 cM, 8 cM, 7 cM, 6 cM, 5 cM, 4 cM, 3 cM, 2 cM, 1 cM, 0.75 cM, 0.5 cM or 0.25 cM (based on a single meiosis based genetic map) of any of the markers in Table 2 could also be used to predict BLB resistance in a rice plant.

Marker assisted selection

Molecular markers can be used in a variety of plant breeding applications (e.g. see Staub et al. (1996) *Hortscience* 31: 729-741; Tanksley (1983) *Plant Molecular Biology Reporter.* 1: 3-8). One of the main areas of interest is to increase the efficiency of backcrossing and introgressing genes using marker-assisted selection (MAS). A molecular marker that demonstrates linkage with a locus affecting a desired phenotypic trait provides a useful tool for the selection of the trait in a plant

population. This is particularly true where the phenotype is hard to assay. Since DNA marker assays are less laborious and take up less physical space than field phenotyping, much larger populations can be assayed, increasing the chances of finding a recombinant with the target segment from the donor line moved to the recipient line. The closer the linkage, the more useful the marker, as recombination is less likely to occur between the marker and the gene causing the trait, which can result in false positives. Having flanking markers decreases the chances that false positive selection will occur as a double recombination event would be needed. The ideal situation is to have a marker in the gene itself, so that recombination cannot occur between the marker and the gene. Such a marker is called a `perfect marker`.

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

The key components to the implementation of MAS are: (i) Defining the population within which the marker-trait association will be determined, which can be a segregating population, or a

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

SSRs can be defined as relatively short runs of tandemly repeated DNA with lengths of 6 bp or less (Tautz (1989) *Nucleic Acid Research* 17: 6463-6471; Wang et al. (1994) *Theoretical and Applied Genetics*, 88:1-6) Polymorphisms arise due to variation in the number of repeat units, probably caused by slippage during DNA replication (Levinson and Gutman (1987) *Mol Biol Evol* 4: 203-221). The variation in repeat length may be detected by designing PCR primers to the conserved non-repetitive flanking regions (Weber and May (1989) *Am J Hum Genet.* 44:388-396). SSRs are highly suited to mapping and MAS as they are multi-allelic, codominant, reproducible and amenable to high throughput automation (Rafalski et al. (1996) Generating and using DNA markers in plants. In: *Non-mammalian genomic analysis: a practical guide.* Academic press. pp 75-135).

Various types of SSR markers can be generated, and SSR profiles can be obtained by gel electrophoresis of the amplification products. Scoring of marker genotype is based on the size of the amplified fragment.

Various types of FLP markers can also be generated. Most commonly, amplification primers are used to generate fragment length polymorphisms. Such FLP markers are in many ways similar to SSR markers, except that the region amplified by the primers is not typically a highly repetitive region. Still, the amplified region, or amplicon, will have sufficient variability among germplasm, often due to insertions or deletions, such that the fragments generated by the amplification primers can be distinguished among polymorphic individuals, and such indels are known to occur frequently in maize (Bhattramakki et al. (2002). *Plant Mol Biol* 48, 539-547; Rafalski (2002b), supra).

SNP markers detect single base pair nucleotide substitutions. Of all the molecular marker types, SNPs are the most abundant, thus having the potential to provide the highest genetic map resolution (Bhattramakki et al. 2002 *Plant Molecular Biology* 48:539-547). SNPs can be assayed at an even higher level of throughput than SSRs, in a so-called `ultra-high-throughput` fashion, as SNPs do not require large amounts of DNA and automation of the assay may be straight-forward. SNPs also have the promise of being relatively low-cost systems. These three factors together make SNPs highly attractive for use in MAS. Several methods are available for SNP genotyping, including but not limited to, hybridization, primer extension, oligonucleotide ligation, nuclease cleavage, minisequencing, and coded spheres. Such methods have been reviewed in: Gut (2001) *Hum Mutat* 17 pp. 475-492; Shi (2001) *Clin Chem* 47, pp. 164-172; Kwok (2000) *Pharmacogenomics* 1, pp. 95-100;

and Bhattramakki and Rafalski (2001) Discovery and application of single nucleotide polymorphism markers in plants. In: R. J. Henry, Ed, *Plant Genotyping: The DNA Fingerprinting of Plants*, CABI Publishing, Wallingford. A wide range of commercially available technologies utilize these and other methods to interrogate SNPs including Masscode.TM. (Qiagen), INVADER®. (Third Wave Technologies) and Invader PLUS®, SNAPSHOT®. (Applied Biosystems), TAQMAN®. (Applied Biosystems) and BEADARRAYS®. (Illumina).

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

Many of the PHM markers presented herein can readily be used as single nucleotide polymorphic (SNP) markers to select for the gene loci on chromosome 7. Using PCR, the primers are used to amplify DNA segments from individuals (preferably inbred) that represent the diversity in the population of interest. The PCR products are sequenced directly in one or both directions. The resulting sequences are aligned and polymorphisms are identified. The polymorphisms are not limited to single nucleotide polymorphisms (SNPs), but also include indels, CAPS, SSRs, and VNTRs (variable number of tandem repeats). Specifically, with respect to the fine map information described herein, one can readily use the information provided herein to obtain additional polymorphic SNPs (and other markers) within the region amplified by the primers disclosed herein. Markers within the described map region can be hybridized to BACs or other genomic libraries, or electronically aligned with genome sequences, to find new sequences in the same approximate location as the described markers.

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

Isozyme profiles and linked morphological characteristics can, in some cases, also be indirectly used as markers. Even though they do not directly detect DNA differences, they are often

influenced by specific genetic differences. However, markers that detect DNA variation are far more numerous and polymorphic than isozyme or morphological markers (Tanksley (1983) *Plant Molecular Biology Reporter* 1:3-8).

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

In general, MAS uses polymorphic markers that have been identified as having a significant likelihood of co-segregation with a trait such as the BLB resistance trait. Such markers are presumed to map near a gene or genes that give the plant its BLB resistant phenotype, and are considered indicators for the desired trait, or markers. Plants are tested for the presence of a desired allele in the marker, and plants containing a desired genotype at one or more loci are expected to transfer the desired genotype, along with a desired phenotype, to their progeny. Thus, plants with BLB resistance can be selected for by detecting one or more marker alleles, and in addition, progeny plants derived from those plants can also be selected. Hence, a plant containing a desired genotype in a given chromosomal region (i.e. a genotype associated with bacterial leaf blight resistance) is obtained and then crossed to another plant. The progeny of such a cross would then be evaluated genotypically using one or more markers and the progeny plants with the same genotype in a given chromosomal region would then be selected as having BLB resistance.

Markers were identified from linkage mapping as being associated with the bacterial leaf blight resistance trait. Reference sequences for the markers are represented by SEQ ID NOs: 5 and 10-26. SNP positions are identified within the marker sequences.

The SNPs could be used alone or in combination (i.e. a SNP haplotype) to select for a favorable resistant gene allele associated with BLB resistance. For example, a SNP haplotype at the chromosome 10 QTL disclosed herein can comprise: a "G" at R1091-7-06, a "G" at R1091-7-021, a "C" at R1091-7-022, a "G" at R1091-7-027, a "G" at R1091-7-028, a "G" at R1091-7-026, an "T" at R1091-7-42, a "A" at R1091-7-037, a "C" at R1091-7-030, a "G" at R1091-7-018, a "G" at R1091-7-053, a "C" at R1091-7-052, a "G" at R1091-7-040, a "T" at R1091-7-044, a "A" at R1091-7-039, a "C" at R1091-7-024, and a "G" at R1091-7-007, or any combination thereof.

The skilled artisan would expect that there might be additional polymorphic sites at marker loci in and around the chromosome 10markers identified herein, wherein one or more polymorphic sites is in linkage disequilibrium (LD) with an allele at one or more of the polymorphic sites in the haplotype and thus could be used in a marker assisted selection program to introgress a gene allele or

genomic fragment of interest. Two particular alleles at different polymorphic sites are said to be in LD if the presence of the allele at one of the sites tends to predict the presence of the allele at the other site on the same chromosome (Stevens, *Mol. Diag.* 4:309-17 (1999)). The marker loci can be located within 5 cM, 2 cM, or 1 cM (on a single meiosis based genetic map) of the BLB resistance trait QTL.

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

Plant compositions

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

Proteins and Variants and Fragments Thereof

BLB4 polypeptides are encompassed by the disclosure. "BLB4 polypeptide" and "BLB4 protein" as used herein interchangeably refers to a polypeptide(s) having BLB resistance activity, and is sufficiently homologous to the BLB4 polypeptide of SEQ ID NO: 3. A variety of BLB4 polypeptides are contemplated.

"Sufficiently identical" is used herein to refer to an amino acid sequence that has at least about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater sequence identity. In one embodiment the BLB4 polypeptide has at least about 40%, 45%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater sequence identity compared to SEQ ID NO: 3. In some embodiments the sequence identity is against the full length sequence of an BLB4 polypeptide. The term "about" when used herein in context with percent sequence identity means +/- 1.0%.

A "recombinant protein" is used herein to refer to a protein that is no longer in its natural environment, for example in vitro or in a recombinant bacterial or plant host cell; a protein that is expressed from a polynucleotide that has been edited from its native version; or a protein that is expressed from a polynucleotide in a different genomic position relative to the native sequence.

"Substantially free of cellular material" as used herein refers to a polypeptide including preparations of protein having less than about 30%, 20%, 10% or 5% (by dry weight) of non-target protein (also referred to herein as a "contaminating protein").

"Fragments" or "biologically active portions" include polypeptide or polynucleotide fragments comprising sequences sufficiently identical to an BLB4 polypeptide or polynucleotide, respectively, and that exhibit BLB resistance when expressed in a plant.

"Variants" as used herein refers to proteins or polypeptides having an amino acid sequence that is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater identical to the parental amino acid sequence.

In some embodiments an BLB4 polypeptide comprises an amino acid sequence having at least about 40%, 45%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater identity to the full length or a fragment of the amino acid sequence of SEQ ID NO: 3, wherein the BLB4 polypeptide has BLB resistance when expressed in a plant.

In some embodiments an BLB4 polypeptide comprises an amino acid sequence of SEQ ID NO: 3 having 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70,71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95 or more amino acid substitutions compared to the amino acid at the corresponding position of the respective SEQ ID NO: 3.

Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of a BLB4 polypeptide can be prepared by mutations in the DNA. This may also be accomplished by one of several forms of mutagenesis, such as for example site-specific double strand break technology, and/or in directed evolution. In some aspects, the changes encoded in the amino acid sequence will not substantially affect the function of the protein. Such variants will possess the desired activity. However, it is understood that the ability of a BLB4 polypeptide to confer BLB resistance may be improved by the use of such techniques upon the compositions of this disclosure.

Conservative amino acid substitutions may be made at one or more predicted nonessential amino acid residues. A "nonessential" amino acid residue is a residue that can be altered from the wild-type sequence of a BLB4 polypeptide without altering the biological activity. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include: amino acids with basic side chains (e.g., lysine, arginine, histidine); acidic side chains (e.g., aspartic acid, glutamic acid); polar, negatively charged residues and their amides (e.g., aspartic acid, asparagine, glutamic, acid, glutamine; uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine); small aliphatic, nonpolar or

slightly polar residues (e.g., Alanine, serine, threonine, proline, glycine); nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan); large aliphatic, nonpolar residues (e.g., methionine, leucine, isoleucine, valine, cystine); beta-branched side chains (e.g., threonine, valine, isoleucine); aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine); large aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan).

Amino acid substitutions may be made in nonconserved regions that retain function. In general, such substitutions would not be made for conserved amino acid residues or for amino acid residues residing within a conserved motif, where such residues are essential for protein activity. Examples of residues that are conserved and that may be essential for protein activity include, for example, residues that are identical between all proteins contained in an alignment of similar or related toxins (or protein class) to the sequences of the embodiments (e.g., residues that are identical in an alignment of homologous proteins). Examples of residues that are conserved but that may allow conservative amino acid substitutions and still retain activity include, for example, residues that have only conservative substitutions between all proteins contained in an alignment of similar or related toxins (or protein class) to the sequences of the embodiments (e.g., residues that have only conservative substitutions between all proteins contained in the alignment homologous proteins). However, one of skill in the art would understand that functional variants may have minor conserved or nonconserved alterations in the conserved residues.

Alternatively, alterations may be made to the protein sequence of many proteins at the amino or carboxy terminus without substantially affecting activity. This can include insertions, deletions or alterations introduced by modern molecular methods, such as PCR, including PCR amplifications that alter or extend the protein coding sequence by virtue of inclusion of amino acid encoding sequences in the oligonucleotides utilized in the PCR amplification. Alternatively, the protein sequences added can include entire protein-coding sequences, such as those used commonly in the art to generate protein fusions. Such fusion proteins are often used to (1) increase expression of a protein of interest (2) introduce a binding domain, enzymatic activity or epitope to facilitate either protein purification, protein detection or other experimental uses known in the art (3) target secretion or translation of a protein to a subcellular organelle, such as the periplasmic space of Gram-negative bacteria, mitochondria or chloroplasts of plants or the endoplasmic reticulum of eukaryotic cells, the latter of which often results in glycosylation of the protein.

Variant nucleotide and amino acid sequences of the disclosure also encompass sequences derived from mutagenic and recombinogenic procedures such as DNA shuffling. With such a procedure, one or more different BLB4 polypeptide coding regions can be used to create a new BLB4 polypeptide possessing the desired properties. In this manner, libraries of recombinant polynucleotides

are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined in vitro or in vivo.

In another embodiment fusion proteins are provided that include within its amino acid sequence an amino acid sequence comprising an BLB4 polypeptide of the disclosure. Methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art. Polynucleotides encoding an BLB4 polypeptide may be fused to signal sequences which will direct the localization of the BLB4 polypeptide to particular compartments of a prokaryotic or eukaryotic cell and/or direct the secretion of the BLB4 polypeptide of the embodiments from a prokaryotic or eukaryotic cell.

Nucleic Acid Molecules and Variants and Fragments Thereof

Isolated or recombinant nucleic acid molecules comprising nucleic acid sequences encoding BLB4 polypeptides or biologically active portions thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding proteins with regions of sequence homology are provided. As used herein, the term "nucleic acid molecule" refers to DNA molecules (e.g., recombinant DNA, cDNA, genomic DNA, plastid DNA, mitochondrial DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule (or DNA) is used herein to refer to a nucleic acid sequence (or DNA) that is no longer in its natural environment, for example in vitro. A "recombinant" nucleic acid molecule (or DNA) is used herein to refer to a nucleic acid sequence (or DNA) that is in a recombinant bacterial or plant host cell; has been edited from its native sequence; or is located in a different location than the native sequence. In some embodiments, an "isolated" or "recombinant" nucleic acid is free of sequences (preferably protein encoding sequences) that naturally flank the nucleic acid (i.e., sequences located at the 5′ and 3′ ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For purposes of the disclosure, "isolated" or "recombinant" when used to refer to nucleic acid molecules excludes isolated chromosomes. For example, in various embodiments, the recombinant nucleic acid molecules encoding BLB4 polypeptides can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleic acid sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived.

In some embodiments an isolated nucleic acid molecule encoding BLB4 polypeptides has one or more change in the nucleic acid sequence compared to the native or genomic nucleic acid sequence. In some embodiments the change in the native or genomic nucleic acid sequence includes but is not limited to: changes in the nucleic acid sequence due to the degeneracy of the genetic code; changes in

the nucleic acid sequence due to the amino acid substitution, insertion, deletion and/or addition compared to the native or genomic sequence; removal of one or more intron; deletion of one or more upstream or downstream regulatory regions; and deletion of the 5' and/or 3' untranslated region associated with the genomic nucleic acid sequence. In some embodiments the nucleic acid molecule encoding an BLB4 polypeptide is a non-genomic sequence.

A variety of polynucleotides that encode BLB4 polypeptides or related proteins are contemplated. Such polynucleotides are useful for production of BLB4 polypeptides in host cells when operably linked to a suitable promoter, transcription termination and/or polyadenylation sequences. Such polynucleotides are also useful as probes for isolating homologous or substantially homologous polynucleotides that encode BLB4 polypeptides or related proteins.

In some embodiments the nucleic acid molecule encoding an BLB4 polypeptide is a polynucleotide having the sequence set forth in SEQ ID NO: 2, and variants, fragments and complements thereof. "Complement" is used herein to refer to a nucleic acid sequence that is sufficiently complementary to a given nucleic acid sequence such that it can hybridize to the given nucleic acid sequence to thereby form a stable duplex. "Polynucleotide sequence variants" is used herein to refer to a nucleic acid sequence that except for the degeneracy of the genetic code encodes the same polypeptide.

In some embodiments the nucleic acid molecule encoding the BLB4 polypeptide is a non-genomic nucleic acid sequence. As used herein a "non-genomic nucleic acid sequence" or "non-genomic nucleic acid molecule" or "non-genomic polynucleotide" refers to a nucleic acid molecule that has one or more change in the nucleic acid sequence compared to a native or genomic nucleic acid sequence. In some embodiments the change to a native or genomic nucleic acid molecule includes but is not limited to: changes in the nucleic acid sequence due to the degeneracy of the genetic code; optimization of the nucleic acid sequence for expression in plants; changes in the nucleic acid sequence to introduce at least one amino acid substitution, insertion, deletion and/or addition compared to the native or genomic sequence; removal of one or more intron associated with the genomic nucleic acid sequence; insertion of one or more heterologous introns; deletion of one or more upstream or downstream regulatory regions associated with the genomic nucleic acid sequence; insertion of one or more heterologous upstream or downstream regulatory regions; deletion of the 5' and/or 3' untranslated region associated with the genomic nucleic acid sequence; insertion of a heterologous 5' and/or 3' untranslated region; and modification of a polyadenylation site. In some embodiments the non-genomic nucleic acid molecule is a synthetic nucleic acid sequence.

In some embodiments the nucleic acid molecule encoding an BLB4 polypeptide disclosed herein is a non-genomic polynucleotide having a nucleotide sequence having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%,

70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater identity, to the nucleic acid sequence of SEQ ID NO: 2, wherein the BLB4 polypeptide has BLB resistance activity when expressed in a plant.

In some embodiments the nucleic acid molecule encodes an BLB4 polypeptide variant comprising one or more amino acid substitutions to the amino acid sequence of SEQ ID NOS: 3.

Nucleic acid molecules that are fragments of these nucleic acid sequences encoding BLB4 polypeptides are also encompassed by the embodiments. "Fragment" as used herein refers to a portion of the nucleic acid sequence encoding an BLB4 polypeptide. A fragment of a nucleic acid sequence may encode a biologically active portion of an BLB4 polypeptide or it may be a fragment that can be used as a hybridization probe or PCR primer using methods disclosed below. Nucleic acid molecules that are fragments of a nucleic acid sequence encoding an BLB4 polypeptide comprise at least about 150, 180, 210, 240, 270, 300, 330, 360, 400, 450, or 500 contiguous nucleotides or up to the number of nucleotides present in a full-length nucleic acid sequence encoding a BLB4 polypeptide disclosed herein, depending upon the intended use. "Contiguous nucleotides" is used herein to refer to nucleotide residues that are immediately adjacent to one another. Fragments of the nucleic acid sequences of the embodiments will encode protein fragments that retain the biological activity of the BLB4 polypeptide and, hence, retain BLB resesistance. "Retains BLB resistance" is used herein to refer to a polypeptide having at least about 10%, at least about 30%, at least about 50%, at least about 70%, 80%, 90%, 95% or higher of the BLB resistance of the full-length BLB4 polypeptide set forth in SEQ ID NO: 3.

"Percent (%) sequence identity" with respect to a reference sequence (subject) is determined as the percentage of amino acid residues or nucleotides in a candidate sequence (query) that are identical with the respective amino acid residues or nucleotides in the reference sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any amino acid conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2. Those skilled in the art can determine appropriate parameters for aligning sequences, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (e.g., percent identity of query sequence = number of identical positions between query and subject sequences/total number of positions of query sequence $\times 100$).

In some embodiments a BLB4 polynucleotide encodes a BLB4 polypeptide comprising an amino acid sequence having at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%,

90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater identity across the entire length of the amino acid sequence of SEQ ID NO: 3.

The embodiments also encompass nucleic acid molecules encoding BLB4 polypeptide variants. "Variants" of the BLB4 polypeptide encoding nucleic acid sequences include those sequences that encode the BLB4 polypeptides disclosed herein but that differ conservatively because of the degeneracy of the genetic code as well as those that are sufficiently identical as discussed above. Naturally occurring allelic variants can be identified with the use of well-known molecular biology techniques, such as polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant nucleic acid sequences also include synthetically derived nucleic acid sequences that have been generated, for example, by using site-directed mutagenesis but which still encode the BLB4 polypeptides disclosed herein.

The skilled artisan will further appreciate that changes can be introduced by mutation of the nucleic acid sequences thereby leading to changes in the amino acid sequence of the encoded BLB4 polypeptides, without altering the biological activity of the proteins. Thus, variant nucleic acid molecules can be created by introducing one or more nucleotide substitutions, additions and/or deletions into the corresponding nucleic acid sequence disclosed herein, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Such variant nucleic acid sequences are also encompassed by the present disclosure.

Alternatively, variant nucleic acid sequences can be made by introducing mutations randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for ability to confer activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly, and the activity of the protein can be determined using standard assay techniques.

The polynucleotides of the disclosure and fragments thereof are optionally used as substrates for a variety of recombination and recursive recombination reactions, in addition to standard cloning methods as set forth in, e.g., Ausubel, Berger and Sambrook, i.e., to produce additional polypeptide homologues and fragments thereof with desired properties. A variety of such reactions are known. Methods for producing a variant of any nucleic acid listed herein comprising recursively recombining such polynucleotide with a second (or more) polynucleotide, thus forming a library of variant polynucleotides are also embodiments of the disclosure, as are the libraries produced, the cells comprising the libraries and any recombinant polynucleotide produced by such methods. Additionally, such methods optionally comprise selecting a variant polynucleotide from such libraries based on activity, as is wherein such recursive recombination is done in vitro or in vivo.

A variety of diversity generating protocols, including nucleic acid recursive recombination protocols are available and fully described in the art. The procedures can be used separately, and/or in combination to produce one or more variants of a nucleic acid or set of nucleic acids, as well as variants of encoded proteins. Individually and collectively, these procedures provide robust, widely applicable ways of generating diversified nucleic acids and sets of nucleic acids (including, e.g., nucleic acid libraries) useful, e.g., for the engineering or rapid evolution of nucleic acids, proteins, pathways, cells and/or organisms with new and/or improved characteristics.

While distinctions and classifications are made in the course of the ensuing discussion for clarity, it will be appreciated that the techniques are often not mutually exclusive. Indeed, the various methods can be used singly or in combination, in parallel or in series, to access diverse sequence variants.

The result of any of the diversity generating procedures described herein can be the generation of one or more nucleic acids, which can be selected or screened for nucleic acids with or which confer desirable properties or that encode proteins with or which confer desirable properties. Following diversification by one or more of the methods herein or otherwise available to one of skill, any nucleic acids that are produced can be selected for a desired activity or property, or such activity at a desired pH, etc. This can include identifying any activity that can be detected, for example, in an automated or automatable format, by any of the assays in the art. A variety of related (or even unrelated) properties can be evaluated, in serial or in parallel, at the discretion of the practitioner.

The nucleotide sequences of the embodiments can also be used to isolate corresponding sequences from a different source. In this manner, methods such as PCR, hybridization, and the like can be used to identify such sequences based on their sequence homology to the sequences set forth herein. Sequences that are selected based on their sequence identity to the entire sequences set forth herein or to fragments thereof are encompassed by the embodiments. Such sequences include sequences that are orthologs of the disclosed sequences. The term "orthologs" refers to genes derived from a common ancestral gene and which are found in different species as a result of speciation. Genes found in different species are considered orthologs when their nucleotide sequences and/or their encoded protein sequences share substantial identity as defined elsewhere herein. Functions of orthologs are often highly conserved among species.

In a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any organism of interest. Methods for designing PCR primers and PCR cloning are generally known in the art and are disclosed in Sambrook, *et al.*, (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York), hereinafter "Sambrook". See also, Innis, *et al.*, eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York);

Innis and Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); and Innis and Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, New York). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially-mismatched primers, and the like.

In hybridization methods, all or part of the nucleic acid sequence can be used to screen cDNA or genomic libraries. Methods for construction of such cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook and Russell, (2001), *supra*. The so-called hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments or other oligonucleotides and may be labeled with a detectable group such as 32P or any other detectable marker, such as other radioisotopes, a fluorescent compound, an enzyme or an enzyme co-factor. Probes for hybridization can be made by labeling synthetic oligonucleotides based on the known BLB4 polypeptide-encoding nucleic acid sequences disclosed herein. Degenerate primers designed on the basis of conserved nucleotides or amino acid residues in the nucleic acid sequence or encoded amino acid sequence can additionally be used. The probe typically comprises a region of nucleic acid sequence that hybridizes under stringent conditions to at least about 12, at least about 25, at least about 50, 75, 100, 125, 150, 175 or 200 consecutive nucleotides of nucleic acid sequences encoding BLB4 polypeptides of the disclosure or a fragment or variant thereof. Methods for the preparation of probes for hybridization and stringency conditions are generally known in the art and are disclosed in Sambrook and Russell, (2001), *supra*.

Nucleotide Constructs, Expression Cassettes and Vectors

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

A further embodiment relates to a transformed organism such as an organism selected from plant cells, bacteria, yeast, baculovirus, protozoa, nematodes and algae. The transformed organism comprises a DNA molecule of the embodiments, an expression cassette comprising the DNA molecule or a vector comprising the expression cassette, which may be stably incorporated into the genome of the transformed organism.

The sequences of the embodiments are provided in DNA constructs for expression in the organism of interest. The construct will include 5' and 3' regulatory sequences operably linked to a sequence of the embodiments. The term "operably linked" as used herein refers to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and where necessary to join two protein coding regions in the same reading frame. The construct may additionally contain at least one additional gene to be cotransformed into the organism. Alternatively, the additional gene(s) can be provided on multiple DNA constructs.

Such a DNA construct is provided with a plurality of restriction sites for insertion of the BLB4 polypeptide gene sequence of the disclosure to be under the transcriptional regulation of the regulatory regions. The DNA construct may additionally contain selectable marker genes.

The DNA construct will generally include in the 5' to 3' direction of transcription: a transcriptional and translational initiation region (i.e., a promoter), a DNA sequence of the embodiments, and a transcriptional and translational termination region (i.e., termination region) functional in the organism serving as a host. The transcriptional initiation region (i.e., the promoter) may be native, analogous, foreign or heterologous to the host organism and/or to the sequence of the embodiments. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. The term "foreign" as used herein indicates that the promoter is not found in the native organism into which the promoter is introduced. Where the promoter is "foreign" or "heterologous" to the sequence of the embodiments, it is intended that the promoter is not the native or naturally occurring promoter for the operably linked sequence of the embodiments. As used herein, a chimeric gene comprises a coding sequence operably linked to a transcription initiation region that is heterologous to the coding sequence. Where the promoter is a native or natural sequence, the expression of the operably linked sequence is altered from the wild-type expression, which results in an alteration in phenotype.

In some embodiments the DNA construct comprises a polynucleotide encoding an BLB4 polypeptide of the embodiments. In some embodiments the DNA construct comprises a polynucleotide encoding a fusion protein comprising an BLB4 polypeptide of the embodiments.

In some embodiments the DNA construct may also include a transcriptional enhancer sequence. As used herein, the term an "enhancer" refers to a DNA sequence which can stimulate promoter activity, and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Various enhancers are known in the art including for example, introns with gene expression enhancing properties in plants (US Patent Application Publication Number 2009/0144863, the ubiquitin intron (i.e., the maize ubiquitin intron 1 (see, for example, NCBI sequence S94464)), the omega enhancer or the omega prime enhancer (Gallie, *et al.*, (1989) *Molecular Biology of RNA* ed. Cech (Liss, New York) 237-256 and Gallie, *et al.*, (1987) *Gene* 60:217-25), the CaMV 35S enhancer (see, e.g., Benfey, *et al.*, (1990) *EMBO J.* 9:1685-96) and the enhancers of US Patent Number 7,803,992 may also be used. The above list of transcriptional enhancers is not meant to be limiting. Any appropriate transcriptional enhancer can be used in the embodiments.

The termination region may be native with the transcriptional initiation region, may be native with the operably linked DNA sequence of interest, may be native with the plant host or may be derived from another source (i.e., foreign or heterologous to the promoter, the sequence of interest, the plant host or any combination thereof).

Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also, Guerineau, *et al.*, (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot, (1991) *Cell* 64:671-674; Sanfacon, *et al.*, (1991) *Genes Dev.* 5:141-149; Mogen, *et al.*, (1990) *Plant Cell* 2:1261-1272; Munroe, *et al.*, (1990) *Gene* 91:151-158; Ballas, *et al.*, (1989) *Nucleic Acids Res.* 17:7891-7903 and Joshi, *et al.*, (1987) *Nucleic Acid Res.* 15:9627-9639.

Where appropriate, a nucleic acid may be optimized for increased expression in the host organism. Thus, where the host organism is a plant, the synthetic nucleic acids can be synthesized using plant-preferred codons for improved expression. See, for example, Campbell and Gowri, (1990) *Plant Physiol.* 92:1-11 for a discussion of host-preferred usage. For example, although nucleic acid sequences of the embodiments may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498). Thus, the rice-preferred for a particular amino acid may be derived from known gene sequences from rice.

Additional sequence modifications are known to enhance gene expression in a cellular host. These include elimination of sequences encoding spurious polyadenylation signals, exon-intron splice site signals, transposon-like repeats, and other well-characterized sequences that may be deleterious to gene expression. The GC content of the sequence may be adjusted to levels average for a given cellular

host, as calculated by reference to known genes expressed in the host cell. The term "host cell" as used herein refers to a cell which contains a vector and supports the replication and/or expression of the expression vector is intended. Host cells may be prokaryotic cells such as *E. coli* or eukaryotic cells such as yeast, insect, amphibian or mammalian cells or monocotyledonous or dicotyledonous plant cells. An example of a monocotyledonous host cell is a rice host cell. When possible, the sequence is modified to avoid predicted hairpin secondary mRNA structures.

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

A number of promoters can be used in the practice of the embodiments. The promoters can be selected based on the desired outcome. The nucleic acids can be combined with constitutive, tissue-preferred, inducible or other promoters for expression in the host organism.

Plant Transformation

The methods of the embodiments involve introducing a polypeptide or polynucleotide into a plant. "Introducing" is as used herein means presenting to the plant the polynucleotide or polypeptide in such a manner that the sequence gains access to the interior of a cell of the plant. The methods of the embodiments do not depend on a particular method for introducing a polynucleotide or polypeptide into a plant, only that the polynucleotide(s) or polypeptide(s) gains access to the interior of at least one cell of the plant. Methods for introducing polynucleotide(s) or polypeptide(s) into plants are known in the art including, but not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

"Stable transformation" as used herein means that the nucleotide construct introduced into a plant integrates into the genome of the plant and is capable of being inherited by the progeny thereof. "Transient transformation" as used herein means that a polynucleotide is introduced into the plant and does not integrate into the genome of the plant or a polypeptide is introduced into a plant. “Plant” as used herein refers to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds, plant cells, propagules, embryos and progeny of the same. Plant cells can be differentiated or undifferentiated (e.g. callus, suspension culture cells, protoplasts, leaf cells, root cells, phloem cells and pollen).

Transformation protocols as well as protocols for introducing nucleotide sequences into plants may vary depending on the type of plant or plant cell, i.e., monocot or dicot, targeted for transformation. Suitable methods of introducing nucleotide sequences into plant cells and subsequent insertion into the plant genome include microinjection (Crossway, *et al*., (1986) *Biotechniques* 4:320-

334), electroporation (Riggs, *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606), *Agrobacterium*-mediated transformation (US Patent Numbers 5,563,055 and 5,981,840), direct gene transfer (Paszkowski, *et al.*, (1984) *EMBO J.* 3:2717-2722) and ballistic particle acceleration (see, for example, US Patent Numbers 4,945,050; 5,879,918; 5,886,244 and 5,932,782; Tomes, *et al.*, (1995) in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg and Phillips, (Springer-Verlag, Berlin) and McCabe, *et al.*, (1988) *Biotechnology* 6:923-926) and Lecl transformation (WO 00/28058). For potato transformation see, Tu, *et al.*, (1998) *Plant Molecular Biology* 37:829-838 and Chong, *et al.*, (2000) *Transgenic Research* 9:71-78. Additional transformation procedures can be found in Weissinger, *et al.*, (1988) *Ann. Rev. Genet.* 22:421-477; Sanford, *et al.*, (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou, *et al.*, (1988) *Plant Physiol.* 87:671-674 (soybean); McCabe, *et al.*, (1988) *Bio/Technology* 6:923-926 (soybean); Finer and McMullen, (1991) *In Vitro Cell Dev. Biol.* 27P:175-182 (soybean); Singh, *et al.*, (1998) *Theor. Appl. Genet.* 96:319-324 (soybean); Datta, *et al.*, (1990) *Biotechnology* 8:736-740 (rice); Klein, *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein, *et al.*, (1988) *Biotechnology* 6:559-563 (maize); US Patent Numbers 5,240,855; 5,322,783 and 5,324,646; Klein, *et al.*, (1988) *Plant Physiol.* 91:440-444 (maize); Fromm, *et al.*, (1990) *Biotechnology* 8:833-839 (maize); Hooykaas-Van Slogteren, *et al.*, (1984) *Nature (London)* 311:763-764; US Patent Number 5,736,369 (cereals); Bytebier, *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet, *et al.*, (1985) in *The Experimental Manipulation of Ovule Tissues,* ed. Chapman, *et al.*, (Longman, New York), pp. 197-209 (pollen); Kaeppler, *et al.*, (1990) *Plant Cell Reports* 9:415-418 and Kaeppler, *et al.*, (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); D'Halluin, *et al.*, (1992) *Plant Cell* 4:1495-1505 (electroporation); Li, *et al.*, (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford, (1995) *Annals of Botany* 75:407-413 (rice); Osjoda, *et al.*, (1996) *Nature Biotechnology* 14:745-750 (maize via *Agrobacterium tumefaciens*).

Methods to Introduce Genome Editing Technologies into Plants

In some embodiments, the disclosed BLB4 polynucleotide compositions can be introduced into the genome of a plant using genome editing technologies, or previously introduced BLB4 polynucleotides in the genome of a plant may be edited using genome editing technologies. For example, the disclosed polynucleotides can be introduced into a desired location in the genome of a plant through the use of double-stranded break technologies such as TALENs, meganucleases, zinc finger nucleases, CRISPR-Cas, and the like. For example, the disclosed polynucleotides can be introduced into a desired location in a genome using a CRISPR-Cas system, for the purpose of site-specific insertion. The desired location in a plant genome can be any desired target site for insertion, such as a genomic region amenable for breeding or may be a target site located in a genomic window with an existing trait of interest. Existing traits of interest could be either an endogenous trait or a previously introduced trait.

In some embodiments, where the disclosed BLB4 polynucleotide has previously been introduced into a genome, genome editing technologies may be used to alter or modify the introduced polynucleotide sequence. Site specific modifications that can be introduced into the disclosed BLB4 polynucleotide compositions include those produced using any method for introducing site specific modification, including, but not limited to, through the use of gene repair oligonucleotides (e.g. US Publication 2013/0019349), or through the use of double-stranded break technologies such as TALENs, meganucleases, zinc finger nucleases, CRISPR-Cas, and the like. Such technologies can be used to modify the previously introduced polynucleotide through the insertion, deletion or substitution of nucleotides within the introduced polynucleotide. Alternatively, double-stranded break technologies can be used to add additional nucleotide sequences to the introduced polynucleotide. Additional sequences that may be added include, additional expression elements, such as enhancer and promoter sequences. In another embodiment, genome editing technologies may be used to position additional BLB resistant proteins in close proximity to the disclosed BLB4 polynucleotide compositions disclosed herein within the genome of a plant, in order to generate molecular stacks of BLB resistant proteins.

An “altered target site,” “altered target sequence.” “modified target site,” and “modified target sequence” are used interchangeably herein and refer to a target sequence as disclosed herein that comprises at least one alteration when compared to non-altered target sequence. Such "alterations" include, for example: (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, or (iv) any combination of (i) - (iii).

EXAMPLES

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

Example 1

BLB phenotyping, and genetics analysis of PRA1091 donor line

By screening a large rice germplasm, we have identified a plant line, PRA1091, which confers broad-spectrum resistance against BLB across locations.

BLB inoculations were conducted under controlled conditions at maximum tillering stage (45-50 days after sowing) using standard leaf-clipping method. Reaction to the pathogen/disease intensity was evaluated at 14 days post inoculation. Evaluation of leaf damage level by the pathogen was carried out by measuring the average lesion length of four leaves per plant (Resistant <3 cm and susceptible >12cm). BLB phenotype reaction of PRA1091 (<3 cm) discriminated from susceptible

lines (<3=resistant and >12 susceptible). Genetics of PRA1091 BLB resistance were identified in F1 and F2 populations developed from a KCP79G7 x PRA1091 cross. F1 showed dominant BLB reaction and was further confirmed by 3 (resistant):1 (susceptible) F2 segregation ratio (single gene inheritance). Three F2 mapping populations were developed to identify the resistant gene(/s) in PRA1091 donor plant (KCP79G7 x PRA1091, VHR27M5 x PRA1091, and V9526M7 x PRA1091). All three recipient parents were highly susceptible to BLB (lesion length >12cm).

Table 1. BLB phenotype of mapping population's parents

Genotype	Average Lesion Length (cm)	BLB phenotype
PRA1091	1	Resistant
KCP79G7	23	Susceptible
V9526M7	18	Susceptible
VHR27M5	16	Susceptible

Example 2

Coarse-mapping and fine-mapping and trait linked markers

Coarse mapping was conducted with genome wide SNP markers using 150 F2 lines derived from KCP79G7 x PRA1091 cross. SNP variations across the chromosomes were used to build marker haplotypes. Marker haplotypes across all chromosomes were correlated with phenotype data to find marker-trait associations and genomic regions contributing for BLB resistance. Joinmap 4.1 and Map-QTL6 were used to conduct genome wide marker-trait associations for potential genes/QTLs. Two different genes (one on chromosome 7 and other on Chromosome 11) were identified in PRA1091. A resistant gene, which is located on Chromosome 7 (between 2.14-4.8 Mb) was further prioritized for fine-mapping and map-based cloning due to its broad-spectrum resistance. To further fine-map the gene on chromosome 7, 45 new SNP variants between PRA1091 and three susceptible parents were identified from in-house sequence data.

The physical interval identified by coarse-mapping was narrowed down using informative recombinants, genotyping, and phenotyping of next generation progenies. By using multiple batches of three independent mapping populations, F2/F3/F4/F5 informative recombinants, the region was fine-mapped to 26 kb interval (MSU7 Osa1 release 7: 4082457-4108842). Informative recombinants identified in each generation through marker genotypes were advanced to next generation and confirmed by phenotypic segregation (3:1 in case of heterozygous class or inbred type reaction in case of parental type allele). Since the novel resistance gene in PRA1091 was dominant, it showed 3:1 segregation in heterozygous class and no segregation in parental homozygous classes. From F3 onwards, about 20 segregant plants in each F2/F3/F4/F5 family were considered for genotyping and

phenotyping. Based on fine-mapping data, a new SNP marker was developed and validated (See Table 2).

Table 2. Marker for BLB1091 candidate region

Use / Trait	Chromosome	Rice Public Genetic Map (cM)	Position (MSUv7)	KCP79G7 Susceptible	PRA1091 Resistant	Reference SEQ ID NO:
BLB CDS	7	30.01	4089141	C	T	5
Flanking marker for 26kb interval-Left	7	30.94	4083456	A	G	10
genic region (g1)	7	30.95	4084277	A	G	11
genic region (G2)	7	30.99	4088814	T	C	12
Intergenic region	7	30.96	4085557	A	G	13
Intergenic region	7	30.97	4086372	A	G	14
genic region (G3)	7	31	4089494	A	G	15
genic region (G4/BLB4)	7	31.04	4093339	A	T	16
Intergenic region between G 4 & 5	7	31.05	4094416	T	A	17
Intergenic region between G 4 & 5	7	31.06	4095525	T	C	18
Intergenic region between G 4 & 5	7	31.07	4096196	A	G	19
Intergenic region between G 4 & 5	7	31.07	4096327	A	G	20
Intergenic region between G 4 & 5	7	31.07	4096351	T	C	21
Intergenic region between G 4 & 5	7	31.07	4096657	A	G	22
Intergenic region between G 4 & 5	7	31.09	4099115	C	T	23
genic region (G5)	7	31.14	4103911	G	A	24
Intergenic region	7	31.17	4106558	A	C	25
Flanking marker for 26kb interval-Right	7	31.2	4109841	T	G	26

Example 3

High-resolution mapping and map-based cloning

The 26 kb fine-mapped region on chromosome 7 (Chr 7:4082457-4108842) in PRA1091 showed five annotated genes Loc_Os07g08050, Loc_Os07g08060, Loc_Os07g08070, Loc_Os07g08080 and Loc_Os07g08090 (public rice genome project, MSU v7). These five genes were sequenced in PRA1091 and susceptible parents to identify potential sequence variations and to develop new SNP markers. Additional segregation populations were developed from F2/F3/4/5/6 to identify informative recombinants within 26 kb region. Phenotyping and genotyping of informative

recombinants identified from multiple population batches further narrowed down the candidate region to 6.7 kb interval (Chr7:4094124-4100865), which is an intergenic region as per the rice MSU v7 genome version and no genes are annotated. Altogether, for fine-mapping and map-based cloning of the putative candidate gene in PRA1091, approximately 6,000 segregating lines were used. This 6.7kb genomic fragment showed complete co-segregation with BLB phenotype in advanced populations and no further recombinants were identified. Part of the 5' intergenic region consists of putative promoter of MSU v7 gene annotation Loc_Os07g08080 gene (the "BLB4" gene). The 6.7kb intergenic region was sequenced by sanger sequencing and deep sequencing (Next generation sequencing) in PRA1091 and susceptible lines to understand potential sequence variations and possibility of any unknown genes which were not annotated by public rice genome annotation project MSU v7. Multiple SNP variations and in-dels were found within 6.7 kb region of PRA1091 when compared to susceptible lines. Only one SNP variation was found in the exon of BLB4 between resistant and susceptible lines, but it showed no association with phenotype in advanced mapping populations. The TATA signal and Transcription start site (TSS) of BLB4 was predicted to be in the 6.7kb intergenic region from PRA1091 and KCP79G7. The 4kb sequence upstream to translational start (ATG) from PRA1091 was analyzed in order to identify the TATA signal and Transcription start site (TSS). A predicted prominent 7bp TATA signal was identified at 1595bp upstream to ATG and a low intense TATA signal which is 32bp downstream to the former. The analysis of the putative promoter region of BLB4 from KCP79G7 identified a strong TATA signal at 1559bp upstream to ATG. The TATA signal identified from KCP79G7 was the low intense TATA signal identified from PRA1091. This prediction shows that the strong TATA signal identified from PRA1091 was unique to that line and the strong TATA signal predicted for KCP79G7 is common to both the lines. The BLB4 coding region (SEQ ID NO: 2) along with the 3.0 kb intergenic region (SEQ ID NO: 4) containing its promoter was cloned for further transgenic validation. The expression profile of BLB4 in leaf samples was analyzed from the infected and control plants of PRA1091 (resistant), KCP79G7 (susceptible), V9526M7 (susceptible) and VHR27M5 (susceptible) at different time intervals. RT-PCR results showed up regulation of BLB4 at 48 hours and 72 hours post inoculation and mocking. Data showed BLB4 expression was induced by both pathogen inoculation and wounding only in the resistant line PRA1091. No up regulation of BLB4 was observed at 0, 24, 48 and 72 hours post inoculation and mocking in all three susceptible lines. RT-PCR results were also confirmed with real time quantitative PCR (qPCR).

Example 4

Rice Transformation

Gene gun mediated particle bombardment was used to generate rice stable transgenic events or genome edited variants.

Seeds from two rice inbred lines; PRA1091 and KCP79G7 were sterilized in 75% ethanol for 2-3 minutes and washed thoroughly with water and incubated in 4% sodium hypochlorite for 10 minutes. The seeds were then washed 5 times with water and dried completely at room temperature. The dried seeds were inoculated on callus inducing media and the plates were incubated at 28 °C in light for 5-7 days. After that the proliferating calli obtained from rice seeds were placed on osmotic media for 4 hours before being bombarded with DNA:gold particles.

Sufficient amount of gold particles (amount of gold particles depends on the number of bombardments) were weighed and placed in 2.0 ml eppendorf tubes. One ml of 100% ethanol was added to the tube and sonicated for 30 seconds before centrifuging for 1 min. The pellet containing the gold particles was resuspended in 1 ml of 100% ethanol, vortexed for 30 seconds and centrifuged again. This step was repeated twice before resuspending the pellet in 1 ml of sterile water. Fifty µl of gold particle suspension was aliquoted into eppendorf tubes and stored at 4 °C. Five µg of DNA, 50 µl of 2.5 mM $CaCl_2$ and 20 µl of 0.1 M spermidine were added to the 50 µl of gold particle suspension; vortexed for 1-2 minutes and allowed to settle down for 5 minutes. The tubes were centrifuged for 2 minutes before discarding the supernatant. The pellet was resuspended in 40 µl of 100% ethanol and mixed gently by vortexing and 5 µl of sample was quickly dispensed onto macro carrier disks and dried completely. The macro carrier disk carrying the DNA:gold particle prep was loaded onto a macro carrier disk holder and a stopping screen was placed on top of the disk. The manufacturer's instructions were followed to deliver DNA:gold particles onto tissue samples which were placed on osmotic medium using a Bio-Rad gene gun (PDS 1000). After bombardment, the tissue samples were kept on the same osmotic medium for 24 h at 32 °C in the dark.

After 24 hours post bombardment, the samples were sub-cultured onto resting media and kept in the dark at 28 °C for 5 days. The cultures were then transferred to selection media containing Hygromycin as selectable agent. After 3-4 selection cycles, proliferating, hygromycin resistant and Zs-Yellow positive callus events or variants were sub-cultured onto regeneration media and then onto rooting and hardening media to obtain stable transgenic events or genome edited variants. Each independent line was transferred to an individual pot in the greenhouse and samples were collected to perform molecular and phenotypic analysis.

Example 5

Validation of BLB4 gene by transgenic overexpression/ complementation and by disruption of unique TATA box using CRISPR-Cas9 system

In order to validate the putative candidate region, KCP79G7 (susceptible line) was transformed with 4.7 kb construct (SEQ ID NO: 6) containing BLB4 genomic region (SEQ ID NO: 1, 1.19 kb) and a 3.0kb genomic region from PRA1091 containing putative promoter (SEQ ID NO: 4) and 500bp putative terminator region for overexpression (See FIG. 1(A)). Phenotyping of 15 T0 transgenic KCP79G7 events showed resistant levels similar to PRA1091 donor plant using the assay as described in Example 1, thus confirming the genetic complementation (Table 3). Single copy T0 events showed expected phenotype and genetic segregation in T1 generation.

Table 3. Transgenic validation of PRA1091 BLB4 driven by its native promoter in KCP79G7 (susceptible line):

ID	Line	Average Lesion Length(cm)	BLB phenotype
4040-4042	Event 1	2.4	Resistant
4045-4047	Event 2	1.8	Resistant
5424-5425	Event 3	1.1	Resistant
7783-7785	PRA1091 (Wild type-resistant)	0.5	Resistant
7812-7816	KCP79G7 (Wild type-susceptible)	21.2	Susceptible

To generate targeted mutations in the CDS and promoter region of BLB4, CRISPR/Cas9 constructs were made to target the unique SNP in the CDS and TATA box in the promoter region of BLB4 in PRA1091 background. The gRNA, (SEQ ID NO: 7) was designed at the juncture of unique SNP position in CDS and the gRNA (SEQ ID NO: 8) was designed to target TATA box of BLB4 promoter. The target sequence and cut sites have been annotated in FIG. 2. Entry plasmid vectors were made for both the gRNA's and sequence verified. Finally, binary vectors were constructed for each gRNA by LR reaction (following the manufacturer's instruction, Invitrogen). The binary vector constitutes gRNA cassette, Cas9 cassette, and plant selection marker. The binary vector (PHP85342) targeting CDS and the binary vector (PHP85017) targeting the TATA box region in the promoter were transformed to rice by using gene-gun mediated in-house rice transformation protocol.

The unique TATA box in PRA1091 was edited by CRISPR-Cas9 system to check the significance of unique TATA box element. The target site selected to create targeted mutations in TATA box is shown in FIG. 1(B) (SEQ ID NO: 9, and annotated as OS-BLB4-CR12 in the FIG. 1(A)). T0 stable variants were analyzed to check for any mutations in the targeted TATA box region. The variants which had deletions of the TATA box showed susceptible reaction to Xanthomonas oryzae and those variants which had intact TATA box showed resistant reaction (Table 4). First generation of CRISPR variants were advanced to second generation (by selfing) to confirm the phenotype and genotype. Detailed analysis of second generation plants confirmed the role of 5bp (TATAAAA)

unique TATA box element in conferring BLB resistance. The results indicate a correlation between variable deletion length and disease phenotype in variants.

Table 4. Validation of BLB4 by disrupting unique TATA box in PRA1091 line using CRISPR-Cas9 system

Variant ID	Average lesion length (cm)	BLB phenotype
321127218	12.38	Susceptible
321127240	16.25	Susceptible
321127244	13.0	Susceptible
321127245	13.0	Susceptible
321127246	14.75	Susceptible
321127247	14.75	Susceptible
Wild type PRA1091	1.5	Resistant

The CDS of BLB4 was also edited by CRISPR-Cas9 in PRA1091 background. The gRNA, SEQ ID NO: 7, was designed at the juncture of the unique SNP position in the CDS. T0 variants which had disruption in the CDS showed susceptible reaction and the variants having wild type CDS showed resistance to BLB. Correlation of sequencing and phenotypic results showed that the T0 variants having deletions in both alleles (Bi-allelic) exhibited susceptibility and those variants having a deletion only in one allele (Mono-allelic) showed resistance to BLB. This result indicates that BLB4 confers resistance even in heterozygous condition in PRA1091, and this data is in agreement with Example 2 on F1 mapping population (KCP79G7xPRA1091) which showed resistance and exhibited dominance. The guide sequence that was designed to target the CDS has been annotated as OS-BLB4-CR13 in FIG. 1.

Table 5. Mutation and phenotypic data of T0 variants generated by targeting CDS of BLB4 in PRA1091gene

Allele change	Mutation type	Phenotype	Avg Lesion length (in cm)
Wild type	Wild type	Resistant	0.5
Wild type	Wild type	Resistant	0.5
[-GTA]ACA	Monoallelic	Resistant	1
GTAACA			
GT[+A]AACA	Monoallelic	Resistant	0.5
GTAACA			
GTAACA	Monoallelic	Resistant	0.5
GT[-A]ACA			
GT[-A]ACA	Biallelic	Moderately susceptible/Susceptible	9
ACA			
[-TGG]TAG	Biallelic	Moderately susceptible	6

GT[-A]ACA			
GT[+A]AACA	Biallelic	Moderately susceptible/Susceptible	10
GTA[+T]ACA			
GTA[+T]ACA	Biallelic	Moderately susceptible/Susceptible	9
GT[+A]AACA			
GT[-A]ACA	Biallelic	Moderately susceptible	7
GTA[+G]ACA	Biallelic	Moderately susceptible	6
GT[+A]AACA			
GTA[+T]ACA			

CLAIMS

What is claimed:

1. A method of identifying a rice plant with increased resistance to bacterial leaf blight, said method comprising:
    a. detecting in the rice plant a resistant gene allele associated with increased resistance to bacterial leaf blight, wherein said resistant gene allele comprises a "T" in the BLB4 resistance gene at nucleotide 51 of SEQ ID NO:5; and
    b. identifying the rice plant as having the QTL allele, wherein said rice plant has increased resistance to BLB4.

2. The method of claim 1, further comprising counter-selecting the rice plant from a breeding program.

3. A method of identifying a rice plant with increased resistance to bacterial leaf blight, said method comprising:
    a. detecting in the genome of a rice plant any of the following:
        i. a polynucleotide encoding a polypeptide having the amino acid sequence set forth in SEQ ID NO:3;
        ii. a polynucleotide encoding a polypeptide having an amino acid sequence that is at least 80% identical to SEQ ID NO:3 that has antiporter/sodium ion transporter activity;
        iii. a polynucleotide comprising a sequence set forth in SEQ ID NO:2;
        iv. a polynucleotide comprising a sequence set forth in SEQ ID NO:4; or
        v. one or more marker alleles within 5 cM of (i) or (ii) that are linked to and associated with (i) or (ii); and
    b. identifying a rice plant as having increased resistance to bacterial leaf blight, if any of (i), (ii), or (iii) is detected.

4. A method of increasing resistance to BLB disease in plant, comprising:
    a. expressing in a plant a recombinant polynucleotide that encodes a polypeptide having an amino acid sequence of at least 80% sequence identity, when compared to SEQ ID NO:3; and
    b. increased resistance to bacterial leaf blight in the plant when compared to a control plant not comprising the recombinant polynucleotide.

5. The method of claim 4, further comprising a promoter region comprising a nucleic acid as set forth in SEQ ID NO: 4.

6. The method of claim 4, further comprising obtaining a progeny plant derived from the plant expressing the recombinant polynucleotide, wherein said progeny plant comprises in its genome the recombinant polynucleotide and exhibits increased resistance to bacterial leaf blight when compared to a control plant not comprising the recombinant polynucleotide.

7. The method of claim 4, wherein said plant is selected from the group consisting of: *Arabidopsis*, maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, millet, sugar cane, and switchgrass.

8. The method of claim 4, wherein said plant is a monocot.

9. A method of identifying a variant of the rice BLB4 gene that gives plants increased resistance to bacterial leaf blight, the method comprising the steps of:

   a. combining through gene shuffling one or more nucleotide sequences encoding one or more fragments of SEQ ID NO:3, a protein that is at least 80% identical to SEQ ID NO:3, or a fragment thereof, to generate variants of the BLB4 gene; and

   b. identifying a variant that exhibits the increased resistance to bacterial leaf blight.

10. The method of claim 9 wherein the method further comprises the steps of:
    a. introducing into a regenerable plant cell a recombinant construct comprising the variant of the BLB4 gene identified by the method of claim 9;
    b. regenerating a transgenic plant from the regenerable plant cell after step (a), wherein the transgenic plant comprises in its genome the recombinant DNA construct; and
    c. selecting a transgenic plant of (b), wherein the transgenic plant comprises the recombinant DNA construct and exhibits increased resistance to bacterial leaf blight, when compared to a control plant not comprising the recombinant DNA construct.

11. The method of claim 9 or 10, wherein said plant is selected from the group consisting of: *Arabidopsis*, maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, millet, sugar cane, and switchgrass.

12. The method of claim 9 or 10, wherein said plant is a monocot.

13. The method of claim 12, wherein said monocot is rice.

14. A method of identifying an allelic variant of the BLB4 gene wherein said allelic variant is associated with increased tolerance to salt stress and/or drought, the method comprising the steps of:

    a. obtaining a population of rice plants, wherein said rice plants exhibit differing levels of BLB4 resistance;

    b. evaluating allelic variations with respect to the polynucleotide sequence encoding a protein comprising SEQ ID NO:3, or in the genomic region that regulates the expression of the polynucleotide encoding the protein;

    c. ssociating allelic variations with increased resistance to bacterial leaf blight; and

    d. identifying an allelic variant that is associated with increased resistance to bacterial leaf blight.

15. The method of claim 14, further comprising detecting said allelic variant associated with increased resistance to bacterial leaf blight and selecting a rice plant if said allelic variant is detected.

16. A method of introducing an allelic variant of a BLB4 gene wherein said allelic variant is associated with increased resistance to bacterial leaf blight, the method comprising introducing a mutation in the endogenous BLB4 gene such that the allelic variant comprises a polynucleotide sequence encoding a protein that is at least 80% identical to SEQ ID NO:3, using zinc finger nuclease, Transcription Activator-like Effector Nuclease (TALEN), the CRISPR/Cas system, or meganuclease.

17. A recombinant DNA construct comprising a polynucleotide operably linked to at least one regulatory sequence wherein said polynucleotide comprises a nucleic acid sequence encoding an amino acid sequence of at least 80% sequence identity, when compared to SEQ ID NO:3.

18. The recombinant DNA construct of claim 17, wherein said at least one regulatory sequence is a promoter functional in a plant cell.

19. A transgenic plant cell comprising the recombinant DNA construct of claim 17.

20. A transgenic plant comprising the transgenic plant cell of claim 19.

21. The transgenic plant of claim 20, wherein said transgenic plant is selected from the group consisting of: Arabidopsis, maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, millet, sugar cane, and switchgrass.

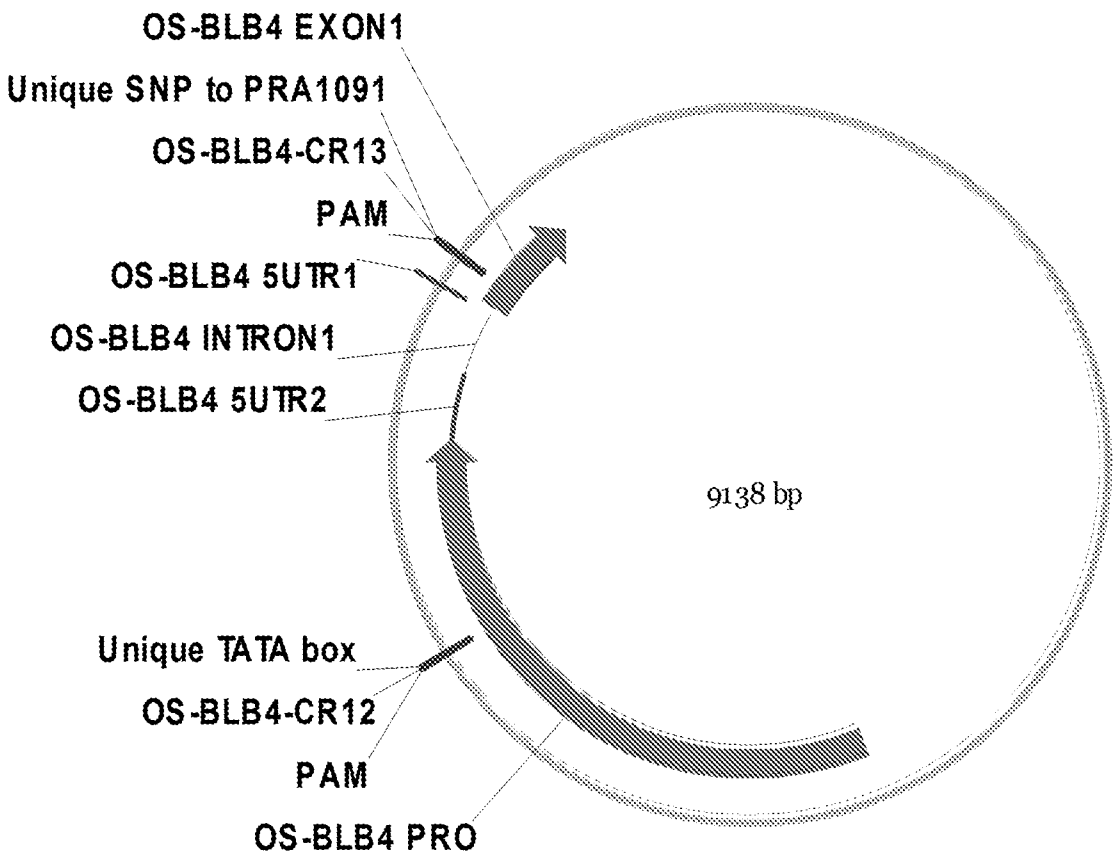
22. Transgenic seed produced from the transgenic plant of claim 21.

23. A method of identifying and/or selecting a rice plant having increased resistance bacterial leaf blight, said method comprising:
    a. screening a population with a marker located within an interval on chromosome 7 comprising and flanked by R1091-7-06 and R1091-7-007 to determine if one or more rice plants from the population comprises a gene allele comprising a "T" in the BLB4 resistance gene at nucleotide 51 of SEQ ID NO:5 and a member of the group consisting of:
        i. a "G" at R1091-7-06;
        ii. a "G" at R1091-7-021;
        iii. a "C" at R1091-7-022;
        iv. a "G" at R1091-7-027;
        v. a "G" at R1091-7-028;
        vi. a "G" at R1091-7-026;
        vii. a "T" at R1091-7-42;
        viii. a "A" at R1091-7-037;
        ix. a "C" at R1091-7-030;
        x. a "G" at R1091-7-018;
        xi. a "G" at R1091-7-053;
        xii. a "C" at R1091-7-052;
        xiii. a "G" at R1091-7-040;
        xiv. a "T" at R1091-7-044;
        xv. a "A" at R1091-7-039;
        xvi. a "C" at R1091-7-024; and
        xvii. a "G" at R1091-7-007; and
    b. selecting from said population at least one rice plant comprising the gene allele.

24. The method of claim 23, further comprising:
    a. crossing the rice plant of (b) to a second rice plant; and
    b. obtaining a progeny plant that has the gene allele.

FIG. 1

(A)



(B)

CCATCTAATATAAAAGACAGAGAT

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/020898

A. CLASSIFICATION OF SUBJECT MATTER
INV. A01H1/04 A01H5/10 C07K14/415 C12N15/82 A01H6/46
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
A01H C07K C12N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, Sequence Search, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 2 295 582 A2 (MONSANTO TECHNOLOGY LLC [US]) 16 March 2011 (2011-03-16) abstract sequence 20043 -----	17-22
X	DATABASE UniProt [Online] 5 July 2004 (2004-07-05), "SubName: Full=Os07g0177400 protein {ECO:0000313¦EMBL:BAH93805.1};", XP002792855, retrieved from EBI accession no. UNIPROT:Q6ZEY5 Database accession no. Q6ZEY5 sequence ----- -/--	19-22

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

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 July 2019	30/07/2019

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

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/020898

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DJEDATIN GUSTAVE ET AL: "Identification of novel major and minor QTLs associated withXanthomonas oryzaepv.oryzae(African strains) resistance in rice (Oryza sativaL.)", RICE, SPRINGER US, BOSTON, vol. 9, no. 1, 23 April 2016 (2016-04-23), pages 1-10, XP035672420, ISSN: 1939-8425, DOI: 10.1186/S12284-016-0090-9 [retrieved on 2016-04-23] abstract table 2 table 3b -----	1-24
X	LI Z K ET AL: "Complex genetic networks underlying the defensive system of rice (Oryza sativa L.) to Xanthomonas oryzae pv. oryzae", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 103, no. 21, 23 May 2006 (2006-05-23) , pages 7994-7999, XP002792856, ISSN: 0027-8424 abstract figure 1 -----	1-24

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2019/020898

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 2295582 A2	16-03-2011	EP 1621632 A1	01-02-2006
		EP 2295582 A2	16-03-2011
		US 2006075522 A1	06-04-2006
		US 2008090998 A1	17-04-2008
		US 2015135367 A1	14-05-2015
		US 2017292131 A1	12-10-2017