Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3228155 A1 2023/02/23

(21) 3 228 155

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2022/08/16
(87) **Date publication PCT/PCT Publication Date:** 2023/02/23
(85) **Entrée phase nationale/National Entry:** 2024/02/01
(86) **N° demande PCT/PCT Application No.:** US 2022/074995
(87) **N° publication PCT/PCT Publication No.:** 2023/023499
(30) **Priorité/Priority:** 2021/08/17 (US PCT/US2021/046227)

(51) **Cl.Int./Int.Cl.** ***A01H 6/46*** (2018.01),
A01H 1/04 (2006.01), *C12Q 1/6895* (2018.01),
A01H 5/10 (2018.01)

(71) **Demandeur/Applicant:**
PIONEER HI-BRED INTERNATIONAL, INC., US

(72) **Inventeurs/Inventors:**
DELEON, ALYSSA MARIE, US;
HUMBERT, SABRINA, US;
JUNG, MARK TIMOTHY, US;
MARTIN, STACY, US;
TABOR, GIRMA M., US;
THATCHER, SHAWN, US;
WOLTERS, PETRA J, US

(74) **Agent:** TORYS LLP

(54) **Titre : COMPOSITIONS ET PROCEDES DE RESISTANCE AUX TACHES GRISES DES FEUILLES**
(54) **Title: COMPOSITIONS AND METHODS FOR GRAY LEAF SPOT RESISTANCE**

(57) **Abrégé/Abstract:**
Provided are plants, cells, tissues, and germplasm thereof comprising genes and marker alleles associated with increased resistance to gray leaf spot disease. Also provided are breeding methods and methods of identifying and selecting plants having alleles associated with increased resistance to gray leaf spot. Provided are methods to identify novel genes that encode proteins providing plant resistance to gray leaf spot and uses thereof. The disclosed genes and marker alleles are useful in the production of gray leaf spot disease resistant plants through breeding, transgenic modification, or genome editing.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 February 2023 (23.02.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/023499 A1

(51) International Patent Classification:
A01H 6/46 (2018.01) *C12Q 1/6895* (2018.01)
A01H 1/04 (2006.01) *A01H 5/10* (2018.01)

(21) International Application Number:
PCT/US2022/074995

(22) International Filing Date:
16 August 2022 (16.08.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/US2021/046227
17 August 2021 (17.08.2021) US

(71) Applicant: **PIONEER HI-BRED INTERNATIONAL, INC.** [US/US]; 7100 NW 62ND AVENUE, PO BOX 1014, JOHNSTON, Iowa 50131-1014 (US).

(72) Inventors: **DELEON, Alyssa Marie**; 7250 NW 62nd Avenue, P.O. Box 552, Johnston, Iowa 50131-0552 (US). **HUMBERT, Sabrina**; 6321 NW 96th Street, Johnston, Iowa 50131 (US). **JUNG, Mark Timothy**; 7250 NW 62nd Avenue, P.O. Box 552, Johnston, Iowa 50131-0552 (US). **MARTIN, Stacy**; 2310 Helena Road, Lancaster, Pennsylvania 17603-4126 (US). **TABOR, Girma M**; 7250 NW 62nd Avenue, P.O. Box 552, JOHNSTON, Iowa 50131-0552 (US). **THATCHER, Shawn**; 7250 NW 62nd Avenue, P.O. Box 552, Johnston, Iowa 50131-0552 (US). **WOLTERS, Petra J**; 7250 NW 62nd Avenue, P.O. Box 552, Johnston, Iowa 50131-0552 (US).

(74) Agent: **RIVAS, Marcos P.**; 9330 Zionsville Road, Indianapolis, Indiana 46268 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** COMPOSITIONS AND METHODS FOR GRAY LEAF SPOT RESISTANCE

(57) **Abstract:** Provided are plants, cells, tissues, and germplasm thereof comprising genes and marker alleles associated with increased resistance to gray leaf spot disease. Also provided are breeding methods and methods of identifying and selecting plants having alleles associated with increased resistance to gray leaf spot. Provided are methods to identify novel genes that encode proteins providing plant resistance to gray leaf spot and uses thereof. The disclosed genes and marker alleles are useful in the production of gray leaf spot disease resistant plants through breeding, transgenic modification, or genome editing.

WO 2023/023499 A1

COMPOSITIONS AND METHODS FOR GRAY LEAF SPOT RESISTANCE

FIELD

The disclosure relates to plants, plant breeding and methods of identifying, selecting, and/or creating plants with genes and marker alleles associated with resistance to gray leaf spot disease. Provided are polynucleotides and constructs that encode proteins capable of providing resistance to gray leaf spot and uses thereof. The disclosed polynucleotides, constructs and marker alleles are useful in the production of disease resistant plants through breeding, transgenic modification, and/or genome editing.

CROSS-REFERENCE TO RELATED APPLICAITONS

This application claims priority to International Patent Application PCT/US2021/046227 filed on August 17, 2021, the contents of which are herein incorporated by reference in their entirety.

REFERENCE TO A SEQUENCE LISTING SUBMITTED AS A TEXT FILE VIA EFS-WEB

The official copy of the sequence listing is submitted electronically via Patent Center as an XML formatted sequence listing as a file named 8545_WO_PCT_ST26, created August 9, 2022 and having a size of 106,544 bytes. The sequence listing comprised in this XML formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND

Gray leaf spot a major disease in maize (*Zea mays*). Gray leaf spot is a major concern due to significant reduction in yield, grain weight and quality. Yield losses occur from premature plant death that interrupts filling of the grain and from stalk breakage and lodging that causes ears to be lost in the field. Gray leaf spot occurs in all corn growing areas and can result in 10 to 20% losses.

Farmers can combat infection by fungi such as gray leaf spot through the use of fungicides, but these have environmental side effects and require monitoring of fields and diagnostic techniques to determine which fungus is causing the infection so that the correct fungicide can be used. The use of corn 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. Genetic sources of resistance have been described (White et al. (1979) Annu. Corn Sorghum Res. Conf. Proc. 34:1-15; Carson. 1981.

Sources of inheritance of resistance to gray leaf spot of corn. Ph.D. Thesis, University of Illinois, Urbana-Champaign; Badu-Apraku et al. (1987) Phytopathology 77:957-959; Toman et al. 1993. Phytopathology, 83:981-986; Cowen, N et al. (1991) Maize Genetics Conference Abstracts 33; Jung et al. (1994). Theoretical and Applied Genetics, 89:413-418). However, introgression of resistance can be highly complex.

Selection through the use of molecular markers associated with the gray leaf spot 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 maize plants with a higher level of gray leaf spot. There are multiple QTL controlling resistance to gray leaf spot, with each having a different effect on the trait. Thus, it is desirable to provide compositions and methods for identifying and selecting maize plants with newly conferred or enhanced gray leaf spot resistance. There is a continuous need for disease-resistant plants and methods to find disease resistant genes.

SUMMARY

Compositions and methods useful in identifying and selecting plant markers or genes associated with increased disease resistance are provided herein. These include “R genes” which are genes associated with (and that can provide) increased resistance to a disease. The compositions and methods disclosed herein are thus useful in selecting disease resistant plants, breeding for disease resistant plants, creating transgenic disease resistant plants, and/or creating genome edits for disease resistance in plants. Also provided herein are plants and methods for making plants having the disclosed markers and/or genes associated with disease resistance that is enhanced as compared to control plants. In some embodiments, the compositions and methods are useful in selecting disease resistant plants, introgressing disease resistance into plants, creating transgenic disease resistant plants, and/or creating disease resistant genome edited plants. In some embodiments, the disease resistant marker or plant provided herein is one associated with increased resistance to gray leaf spot (“GLS”).

A plant having an R gene or associated marker allele disclosed herein may be crossed to a second plant in order to obtain one or more progeny plants having the resistant gene or marker allele. The progeny plant may have newly disease resistance or enhanced relative to a control plant that does not have the disease resistance gene or marker allele.

In an aspect, provided are methods for identifying and/or selecting one or more plant materials having an R gene or maker allele associated with increased resistance to gray leaf spot. As used herein, the term “plant materials” refers to one or more plants, plant cells, plant tissues, seeds, or germplasm thereof. In some examples, the methods for identifying and/or

selecting comprise detecting or selecting one or more plant materials having a genomic region comprising SEQ ID NO:40 or SEQ ID NO:56. In certain examples, the methods for identifying and/or selecting comprise detecting or selecting a plant material having a genomic region comprising the promoter or terminator of SEQ ID NOs:39, 55, 57 or 61. The identified or selected plant may possess disease resistance (e.g. GLS resistance) that is newly conferred or enhanced relative to a control plant that does not have a genomic region comprising one or more of SEQ ID NOs:40, 56, 39, or 55. In a further example, the identified or selected plant material comprises a gene (the "PRR gene") encoding a PRR polypeptide (PRR01 or PRR03) that is associated with resistance to gray leaf spot. The PRR polypeptide can comprise an amino acid sequence as set forth in SEQ ID NO:41 (PRR03 polypeptide) or SEQ ID NO:58 (PRR01 polypeptide).

In one aspect, methods are provided to identify and/or select plant materials with a QTL or marker allele associated with increased resistance to gray leaf spot. Generally, such methods can include obtaining a nucleic acid sample from a plant, seed, tissue or germplasm thereof; and screening the sample for the presence of nucleic acid comprising (i) SEQ ID NOs:40, 56, 39, or 55, (ii) sequence encoding a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, or (iii) one or more GLS resistance marker alleles. As used herein the term a "GLS resistance marker allele" refers to any of: an "A" at C2_GLS_38 (position 56 of reference sequence SEQ ID NO:42); a "C" at C2_GLS_68 (position 55 of reference sequence SEQ ID NO:43); an "A" at C2_GLS_43 (position 51 of reference sequence SEQ ID NO:44); a "T" at C2_GLS_80 (position 62 of reference sequence SEQ ID NO:45); a "C" at C2_GLS_87 (position 201 of reference sequence SEQ ID NO:46); a "G" at C2_GLS_47 (position 61 of reference sequence SEQ ID NO:47); a "T" at C2_GLS_48 (position 61 of reference sequence SEQ ID NO:48); a "C" at C01281-1 (position 118 of reference sequence SEQ ID NO:49); a "G" at C01685-2 (position 125 of reference sequence SEQ ID NO:50); a "G" at C01267-2 (position 143 of reference sequence SEQ ID NO:51); a "T" at C00682-2 (position 244 of reference sequence SEQ ID NO:52); a "C" or "G" at C00572-3 (position 172 of reference sequence SEQ ID NO:53); an "A" at PHM2363-23 (position 55 of reference sequence SEQ ID NO:53); a "C" at C103H6N-001 (position 51 of reference sequence SEQ ID NO:19); a "C" at C103H6R-001 (position 51 of reference sequence SEQ ID NO:20); a "C" at ZmChr4v2_142171204 (position 201 of reference sequence SEQ ID NO:21); a "G" at ZmChr4v2_142171261 (position 200 of reference sequence SEQ ID NO:22); a "C" at ZmChr4v2_142192918 (position 200 of reference sequence SEQ ID NO:23); a "G" at ZmChr4v2_143270779 (position 200 of reference sequence SEQ ID NO:25); a "C" at

C103H6T-001 (position 51 of reference sequence SEQ ID NO:26); a "C" at C103HNM (position 201 of reference sequence SEQ ID NO:27); a "C" at C103HUA (position 201 of reference sequence SEQ ID NO:28); a "C" at ZmChr4v2_143572529 (position 201 of reference sequence SEQ ID NO:29); a "C" at C103H6U-001 (position 51 of reference sequence SEQ ID NO:30); a "T" at ZmChr4v2_145458082 (position 201 of reference sequence SEQ ID NO:31); a "C" at ZmChr4v2_145460303 (position 201 of reference sequence SEQ ID NO:32); a "T" at ZmChr4v2_145460875 (position 201 of reference sequence SEQ ID NO:33); a "G" at ZmChr4v2_148727396 (position 201 of reference sequence SEQ ID NO:34); a "T" at ZmChr4v2_148728091 (position 201 of reference sequence SEQ ID NO:35); a "C" at ZmChr4v2_149477191 (position 201 of reference sequence SEQ ID NO:36); a "C" at ZmChr4v2_149477479 (position 201 of reference sequence SEQ ID NO:37); a "T" at PHM521-8 (position 244 of reference sequence SEQ ID NO:1); a "T" at PHM199-23 (position 243 of reference sequence SEQ ID NO:2); a "T" at GLS_G1 (position 51 of reference sequence SEQ ID NO:3); a "C" at GLS_G7 (position 51 of reference sequence SEQ ID NO:4); a "G" at GLS_G14 (position 51 of reference sequence SEQ ID NO:5); a "C" at GLS_G19 (position 51 of reference sequence SEQ ID NO:6); a "C" at GLS_G21 (position 51 of reference sequence SEQ ID NO:7), an "A" at AMD1 (position 51 of reference sequence SEQ ID NO:14); an "A" at AMD4 (position 51 of reference sequence SEQ ID NO:15); an "A" at AMD12 (position 51 of reference sequence SEQ ID NO:16); a "T" at AMD14 (position 51 of reference sequence SEQ ID NO:17); an "A" at AMD20 (position 51 of reference sequence SEQ ID NO:18); a "T" at SYN21168 (C002T5R-001) (position 61 of reference sequence SEQ ID NO:8); a "C" at GLS_G45 (position 51 of reference sequence SEQ ID NO:9); an "A" at C001YAR-001 (position 51 of reference sequence SEQ ID NO:10); a "T" at PHM586-10 (position 114 of reference sequence SEQ ID NO:11); a "C" at PHM5013-12 (position 107 of reference sequence SEQ ID NO:12); a "C" at PHM289-20 (position 121 of reference sequence SEQ ID NO:13); an "A" at PHM6764-7 (position 150 of reference sequence SEQ ID NO:59); and a "G" at PHM16360-9 (position 121 of reference sequence SEQ ID NO:60). Alternatively, the method can include screening the sample for the presence of a marker allele linked to GLS any of the resistance marker allele, e.g., 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, 0.1 cM, or less on a single meiosis-based genetic map, and associated. The method can further include detecting one or more of (i) SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising the amino acid sequence of SEQ ID NO:41 or SEQ ID NO:58, or (iii) one or more of the foregoing GLS resistance marker alleles thereby identifying the plant material as

comprising a QTL or marker allele associated with increased GLS resistance. Additionally, the method can include selecting the plant material identified as comprising a QTL or marker allele associated with increased GLS resistance.

In a particular example, the foregoing method of identifying and/or selecting plant materials with a QTL or marker allele associated with increased resistance to gray leaf spot can include obtaining a nucleic acid sample from each of one or more plants, seeds, tissues or germplasm in a population; screening each sample for the presence of one or more of (i) SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising the amino acid sequence of SEQ ID NO:41 or SEQ ID NO:58, or (iii) the foregoing GLS resistance marker alleles; and selecting one or more of the plants, seeds, tissues or germplasm having the GLS resistance marker allele(s) or sequences associated with increased resistance to gray leaf spot.

In another example, the foregoing methods of identifying and/or selecting plant materials with gray leaf spot resistance can include obtaining a nucleic acid sample from one or more plants, seeds, tissues or germplasm, each sample being representative of a plurality (e.g., a population) of plants, seeds, tissues or germplasm; screening each sample for the presence of one or more of the foregoing GLS resistance marker alleles; and selecting one or more plurality of plants, seeds, tissues or germplasm, wherein the representative sample for the selected plurality has the GLS resistance marker allele(s) associated with increased resistance to gray leaf spot.

The foregoing methods identifying and/or selecting plants can further include crossing at least one of the selected plants to a second plant that does not have a PRR gene, thereby producing a progeny plant whose genome comprises one or more of (i) SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, or (iii) the foregoing GLS resistance marker alleles. In a further example, the second plant is one of a plant line (a “recurrent parent line”) and the method further includes crossing the progeny plant with another plant of the recurrent parent line to produce a second generation progeny whose genome comprises one or more of (i) SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, or (iii) the foregoing GLS resistance marker alleles. Optionally, the second generation progeny can be crossed with the recurrent parent line to produce a third generation progeny whose genome comprises one or more of (i) SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, or (iii) the foregoing GLS resistance marker alleles. This process can be repeated three, four, five, six, seven, or more times, such

that each subsequent generation progeny is crossed with the recurrent parent line, thereby introgressing the PRR gene into the recurrent parent line.

In an alternative method, a plant having increased resistance to gray leaf spot is crossed with a second plant to produce progeny plants. The progeny plants are screened for a QTL or marker allele associated with increased resistance to gray leaf spot in accordance with the methods disclosed herein. Generally, such screening includes obtaining a nucleic acid sample from each of the progeny plants and screening the sample for the presence of nucleic acid comprising (i) SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, or (iii) one or more GLS resistance marker alleles. As used herein the term a "GLS resistance marker allele" refers to any of: an "A" at C2_GLS_38 (position 56 of reference sequence SEQ ID NO:42); a "C" at C2_GLS_68 (position 55 of reference sequence SEQ ID NO:43); an "A" at C2_GLS_43 (position 51 of reference sequence SEQ ID NO:44); a "T" at C2_GLS_80 (position 62 of reference sequence SEQ ID NO:45); a "C" at C2_GLS_87 (position 201 of reference sequence SEQ ID NO:46); a "G" at C2_GLS_47 (position 61 of reference sequence SEQ ID NO:47); a "T" at C2_GLS_48 (position 61 of reference sequence SEQ ID NO:48); a "C" at C01281-1 (position 118 of reference sequence SEQ ID NO:49); a "G" at C01685-2 (position 125 of reference sequence SEQ ID NO:50); a "G" at C01267-2 (position 143 of reference sequence SEQ ID NO:51); a "T" at C00682-2 (position 244 of reference sequence SEQ ID NO:52); a "C" or "G" at C00572-3 (position 172 of reference sequence SEQ ID NO:53); an "A" at PHM2363-23 (position 55 of reference sequence SEQ ID NO:53); a "C" at C103H6N-001 (position 51 of reference sequence SEQ ID NO:19); a "C" at C103H6R-001 (position 51 of reference sequence SEQ ID NO:20); a "C" at ZmChr4v2_142171204 (position 201 of reference sequence SEQ ID NO:21); a "G" at ZmChr4v2_142171261 (position 200 of reference sequence SEQ ID NO:22); a "C" at ZmChr4v2_142192918 (position 200 of reference sequence SEQ ID NO:23); a "G" at ZmChr4v2_143270779 (position 200 of reference sequence SEQ ID NO:25); a "C" at C103H6T-001 (position 51 of reference sequence SEQ ID NO:26); a "C" at C103HNM (position 201 of reference sequence SEQ ID NO:27); a "C" at C103HUA (position 201 of reference sequence SEQ ID NO:28); a "C" at ZmChr4v2_143572529 (position 201 of reference sequence SEQ ID NO:29); a "C" at C103H6U-001 (position 51 of reference sequence SEQ ID NO:30); a "T" at ZmChr4v2_145458082 (position 201 of reference sequence SEQ ID NO:31); a "C" at ZmChr4v2_145460303 (position 201 of reference sequence SEQ ID NO:32); a "T" at ZmChr4v2_145460875 (position 201 of reference sequence SEQ ID NO:33); a "G" at ZmChr4v2_148727396 (position 201 of reference sequence SEQ ID NO:34); a "T" at

ZmChr4v2_148728091 (position 201 of reference sequence SEQ ID NO:35); a "C" at ZmChr4v2_149477191 (position 201 of reference sequence SEQ ID NO:36); a "C" at ZmChr4v2_149477479 (position 201 of reference sequence SEQ ID NO:37); a "T" at PHM521-8 (position 244 of reference sequence SEQ ID NO:1); a "T" at PHM199-23 (position 243 of reference sequence SEQ ID NO:2); a "T" at GLS_G1 (position 51 of reference sequence SEQ ID NO:3); a "C" at GLS_G7 (position 51 of reference sequence SEQ ID NO:4); a "G" at GLS_G14 (position 51 of reference sequence SEQ ID NO:5); a "C" at GLS_G19 (position 51 of reference sequence SEQ ID NO:6); a "C" at GLS_G21 (position 51 of reference sequence SEQ ID NO:7), an "A" at AMD1 (position 51 of reference sequence SEQ ID NO:14); an "A" at AMD4 (position 51 of reference sequence SEQ ID NO:15); an "A" at AMD12 (position 51 of reference sequence SEQ ID NO:16); a "T" at AMD14 (position 51 of reference sequence SEQ ID NO:17); an "A" at AMD20 (position 51 of reference sequence SEQ ID NO:18); a "T" at SYN21168 (C002T5R-001) (position 61 of reference sequence SEQ ID NO:8); a "C" at GLS_G45 (position 51 of reference sequence SEQ ID NO:9); an "A" at C001YAR-001 (position 51 of reference sequence SEQ ID NO:10); a "T" at PHM586-10 (position 114 of reference sequence SEQ ID NO:11); a "C" at PHM5013-12 (position 107 of reference sequence SEQ ID NO:12); a "C" at PHM289-20 (position 121 of reference sequence SEQ ID NO:13); an "A" at PHM6764-7 (position 150 of reference sequence SEQ ID NO:59); and a "G" at PHM16360-9 (position 121 of reference sequence SEQ ID NO:60). Alternatively the method can include screening the sample for the presence of a marker allele linked to any of the GLS resistance marker alleles, e.g., 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 method can further include detecting and selecting one or more of the progeny plants whose nucleic acid sample comprises (i) SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, or (iii) one or more of the foregoing GLS resistance marker alleles, thereby identifying novel progeny plants comprising a QTL or marker allele associated with increased GLS resistance.

In another aspect, methods are provided that include expressing in a plant material a heterologous nucleic acid capable increasing resistance to gray leaf spot disease. The method can include introducing into the plant material a nucleic acid sequence or marker allele associated with increased gray leaf spot resistance, e.g., by transgenic modification or genome editing, approaches. In some examples, the plant material is susceptible to gray leaf spot disease prior to introducing the heterologous nucleic acid. For example, the genome of a plant (e.g., a

plant that is that is susceptible to gray leaf spot disease) is altered by transgenic modification or gene editing to include one or more of (i) at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% nucleic acid sequence identity to one of SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% amino acid sequence identity to SEQ ID NO:41 or SEQ ID NO:58, or (iii) one or more GLS resistance marker alleles disclosed herein. Also provided are the plant materials whose genome has been trangenically modified or gene edited to include (i) SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, or (iii) one or more GLS resistance marker alleles disclosed herein. In some examples, the genome edited plant materials provide increased resistance to GLS disease relative to an isogenic plant lacking the gene edit.

Provided herein is a method of introducing a construct into a plant material that does not comprise a PRR gene associated with increased GLS resistance. The introduced construct comprises a nucleic acid that is heterologous to the plant material, and the heterologous nucleic acid comprises (i) at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% nucleic acid sequence identity to one of SEQ ID NOs:40, 56, 39, or 55, (ii) a sequence encoding a PRR polypeptide comprising at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% amino acid sequence identity to SEQ ID NO:41 or SEQ ID NO:58, or (iii) one or more GLS resistance marker alleles disclosed herein. For example, the construct introduced into the plant can comprise one or more of (i) SEQ ID NOs:40, 56, 39, or 55 or (ii) a sequence encoding a polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58. In some examples, the introduced construct is integrated into a different (non-native) genomic locus. Thus, a construct comprising a sequence that (i) has at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% nucleic acid sequence identity to SEQ ID NO:40 or (ii) encodes a PRR polypeptide having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% amino acid sequence identity to SEQ ID NO:41 can be inserted at a chromosomal locus other than the native PRR03 locus (between 90 and 115 cM) on maize chromosome 04 (e.g. on maize chromosome 1, 2, 3, 5, 6, 7, 8, 9, or 10). A construct comprising a sequence that (i) has at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least

97%, at least 98%, at least 99% or 100% nucleic acid sequence identity to SEQ ID NO:56 or (ii) encodes a PRR polypeptide having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% amino acid sequence identity to SEQ ID NO:58 can be inserted at a chromosomal locus other than the native PRR01 locus (220 and 255 cm) on maize chromosome 02 (e.g. on maize chromosome 1, 3, 4, 5, 6, 7, 8, 9, or 10).

Also disclosed herein is an isolated polynucleotide construct comprising a nucleotide sequence encoding a PRR polypeptide capable of conferring resistance to gray leaf spot, wherein the isolated polynucleotide construct comprises a sequence (i) having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% nucleic acid sequence identity to one of SEQ ID NOs:40, 56, 39, or 55, (ii) encoding a polypeptide having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99%, or 100% amino acid sequence identity to a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, or (iii) that includes one or more GLS resistance marker alleles disclosed herein. In particular examples, the construct is a recombinant construct that includes a sequence encoding a polypeptide having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99%, or 100% amino acid sequence identity to a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58, wherein the coding sequence is operably linked to at least one heterologous regulatory sequence. Also provided are plant materials, such as a plant, plant cell, plant tissue, seed, or germplasm thereof, comprising the isolated polynucleotide construct disclosed herein.

The methods embodied by the present disclosure relate to a method for transforming a host cell, which can be a plant cell. The method comprises transforming the host or plant cell with the isolated polynucleotide construct disclosed herein. The method can further include producing a plant by transforming a plant cell with a construct of the present disclosure and regenerating a plant from the transformed plant cell, thereby producing a plant having a PRR gene or marker allele associated with increased gray leaf spot resistance. In some examples, the regenerated plant has improved GLS disease resistance, as compared to an isogenic plant lacking the PRR gene or marker allele associated with increased gray leaf spot resistance.

In some embodiments, the compositions and methods relate to modified plant material having increased resistance to a disease, wherein prior to modification the plant material lacked a PRR gene associated with GLS resistance. The plant material is modified, e.g., by

mutagenesis or gene editing, to include a nucleotide sequence (i) having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% nucleic acid sequence identity to one of SEQ ID NOs:40, 56, 39, or 55 or (ii) encoding a polypeptide having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99%, or 100% amino acid sequence identity to a PRR polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58.

In another aspect, provided herein is a method of generating a variant of a PRR gene, by gene shuffling one or more nucleotide sequences encoding (i) SEQ ID NO:41 or SEQ ID NO:58, (ii) a protein that is at least 90% identical to any one of SEQ ID NO:41 or SEQ ID NO:58, or (iii) a fragment of (i) or (ii) to thereby generate variants of the PRR gene. Variants are then transiently or stably expressed in plant material and tested for whether they provide increased resistance to gray leaf spot. For example, one or more variants can be incorporated into construct(s), and the construct(s) can be introduced into a regenerable plant cell; and a plant comprising the variant(s) construct can be regenerated from the plant cell. Plants containing the variant(s) can be evaluated for their tolerance/susceptibility to gray leaf spot. Plants having a variant that provides increased tolerance to gray leaf spot, relative to an isogenic plant that lacks the variant construct, can be selected. The plant can be maize, or the plant can be *Arabidopsis*, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, millet, sugar cane, or switchgrass.

Also provided is a method of identifying an allelic variant of a PRR gene disclosed herein. The method comprises obtaining a plurality of plants each of which exhibits differing levels of gray leaf spot resistance; screening nucleic acid samples from each plant for the presence of allelic variations in a polynucleotide sequence encoding a sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO:41 or SEQ ID NO:58; evaluating the variations for genetic linkage to altered tolerance/susceptible to gray leaf spot; and identifying one or more allelic variations associated with increased resistance to gray leaf spot.

The disclosure also provides plants identified, selected, or created using any of the methods presented herein.

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.

The NBS-LRR ("NLR") group of R-genes is the largest class of R-genes discovered to date. In Arabidopsis thaliana, over 150 are predicted to be present in the genome (Meyers et al. (2003), Plant Cell, 15:809-834; Monosi et al. (2004), Theoretical and Applied Genetics, 109:1434-1447), while in rice, approximately 500 NLR genes have been predicted (Monosi (2004) supra). The NBS-LRR class of R genes is comprised of two subclasses. Class 1 NLR genes contain a TIR-Toll/Interleukin-1 like domain at their N' terminus; which to date have only been found in dicots (Meyers (2003) supra; Monosi (2004) supra). The second class of NBS-LRR contain either a coiled-coil domain or an (nt) domain at their N terminus (Baiet et al. (2002) Genome Research, 12:1871-1884; Monosi (2004) supra; Pan et al. (2000), Journal of Molecular Evolution, 50:203-213). Class 2 NBS-LRR have been found in both dicot and monocot species. (Bai (2002) supra; Meyers (2003) supra; Monosi (2004) supra; Pan (2000) supra).

The NBS domain of the gene appears to have a role in signaling in plant defense mechanisms (van der Biezen et al. (1998), Current Biology: CB, 8:R226-R227). The LRR region appears to be the region that interacts with the pathogen AVR products (Michelmore et al. (1998), Genome Res., 8:1113-1130; Meyers (2003) supra). This LRR region in comparison with the NB-ARC (NBS) domain is under a much greater selection pressure to diversify (Michelmore (1998) supra; Meyers (2003) supra; Palomino et al. (2002), Genome Research, 12:1305-1315). LRR domains are found in other contexts as well; these 20-29-residue motifs are present in tandem arrays in a number of proteins with diverse functions, such as hormone – receptor interactions, enzyme inhibition, cell adhesion and cellular trafficking. A number of recent studies revealed the involvement of LRR proteins in early mammalian development, neural development, cell polarization, regulation of gene expression and apoptosis signaling.

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 to herein, "disease resistant" or "have resistance to a disease" refers to a plant showing increase resistance to a disease compared to a control plant, e.g., a control plant can be one that lacks the QTL or PRR gene that provides disease resistance but is otherwise isogenic to the disease resistant plant. Disease resistance may manifest in fewer and/or smaller

lesions, increased plant health, increased yield, increased root mass, increased plant vigor, less or no discoloration, increased growth, reduced necrotic area, or reduced wilting. In some embodiments, an allele may show resistance one or more diseases.

Diseases affecting maize plants include, but are not limited to, bacterial leaf blight and stalk rot; bacterial leaf spot; bacterial stripe; chocolate spot; goss's bacterial wilt and blight; holcus spot; purple leaf sheath; seed rot-seedling blight; bacterial wilt; corn stunt; anthracnose leaf blight; gray leaf spot; aspergillus ear and kernel rot; banded leaf and sheath spot; black bundle disease; black kernel rot; borde blanco; brown spot; black spot; stalk rot; cephalosporium kernel rot; charcoal rot; corticium ear rot; curvularia leaf spot; didymella leaf spot; diplodia ear rot and stalk rot; seed rot; corn seedling blight; diplodia leaf spot or leaf streak; downy mildews; brown stripe downy mildew; crazy top downy mildew; green ear downy mildew; graminicola downy mildew; java downy mildew; philippine downy mildew; sorghum downy mildew; spontaneum downy mildew; sugarcane downy mildew; dry ear rot; ergot; horse's tooth; corn eyespot; fusarium ear and stalk rot; fusarium blight; seedling root rot; gibberella ear and stalk rot; gray ear rot; gray leaf spot; cercospora leaf spot; helminthosporium root rot; hormodendrum ear rot; cladosporium rot; hyalothyridium leaf spot; late wilt; northern leaf blight; white blast; crown stalk rot; corn stripe; northern leaf spot; helminthosporium ear rot; penicillium ear rot; corn blue eye; blue mold; phaeocytostroma stalk rot and root rot; phaeosphaeria leaf spot; physalospora ear rot; botryosphaeria ear rot; pyrenochaeta stalk rot and root rot; pythium root rot; pythium stalk rot; red kernel disease; rhizoctonia ear rot; sclerotial rot; rhizoctonia root rot and stalk rot; rostratum leaf spot; common corn rust; southern corn rust; tropical corn rust; sclerotium ear rot; southern blight; selenophoma leaf spot; sheath rot; shuck rot; silage mold; common smut; false smut; head smut; southern corn leaf blight and stalk rot; southern leaf spot; tar spot; trichoderma ear rot and root rot; white ear rot, root and stalk rot; yellow leaf blight; zonate leaf spot; american wheat striate (wheat striate mosaic); barley stripe mosaic; barley yellow dwarf; brome mosaic; cereal chlorotic mottle; lethal necrosis (maize lethal necrosis disease); cucumber mosaic; johnsongrass mosaic; maize bushy stunt; maize chlorotic dwarf; maize chlorotic mottle; maize dwarf mosaic; maize leaf fleck; maize pellucid ringspot; maize rayado fino; maize red leaf and red stripe; maize red stripe; maize ring mottle; maize rough dwarf; maize sterile stunt; maize streak; maize stripe; maize tassel abortion; maize vein enation; maize wallaby ear; maize white leaf; maize white line mosaic; millet red leaf; and northern cereal mosaic.

Diseases affecting plants include, but are not limited to, bacterial blight; bacterial leaf streak; foot rot; grain rot; sheath brown rot; blast; brown spot; crown sheath rot; downy mildew;

eyespot; false smut; kernel smut; leaf smut; leaf scald; narrow brown leaf spot; root rot; seedling blight; sheath blight; sheath rot; sheath spot; alternaria leaf spot; and stem rot.

A plant having disease resistance may have 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100% increased resistance to a disease compared to a control plant. In some embodiments, a plant may have 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100% increased plant health in the presence of a disease compared to a control plant. In some embodiments, a plant comprising

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 less than or equal to 10%.

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).

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

An "exotic strain," a "tropical line," or an "exotic germplasm" is a strain derived from a plant not belonging to an available elite line or strain of germplasm. In the context of a cross between two 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, 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., disease 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., maize germplasm collection or Andean germplasm collection). The germplasm can be part of an organism, 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. Offspring comprising the desired allele may 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 "linked" or "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 rounds of 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. 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 GLS). Thus, "closely linked" loci such as a marker locus and a second locus can display an inter-locus recombination frequency of 10% or less, preferably about 9% or less, still more preferably about 8% or less, yet more preferably about 7% or less, still more preferably about 6% or less, yet more preferably about 5% or less, still more preferably about 4% or less, yet more preferably about 3% or less, and still more preferably about 2% or less. In highly preferred embodiments, the relevant loci display a recombination frequency of about 1% or less, e.g., about 0.75% or less, more preferably about 0.5% or less, or yet more preferably about 0.25% or less. Two loci that are localized to the same chromosome, and at such a distance that recombination between the two loci occurs at a frequency of less than 10% (e.g., about 9 %, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25%, or less) are also said to be "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. In some cases, two different markers can have the same genetic map coordinates. In that case, the two markers are in such close proximity to each other that recombination occurs between them with such low frequency that it is undetectable.

The term "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. (2002) Nature Reviews Genetics 3:299-309). 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(5046):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.

The term "plant" includes whole plants, plant cells, plant protoplast, plant cell or tissue culture from which plants can be regenerated, plant calli, plant clumps and plant cells that are intact in plants, or parts of plants, such as seeds, flowers, cotyledons, leaves, stems, buds, roots, root tips and the like. As used herein, a "modified plant" means any plant that has a genetic change due to human intervention. A modified plant may have genetic changes introduced through plant transformation, genome editing, mutagenesis, or conventional plant breeding.

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 may consist of primers complementary to sequence flanking the locus and/or probes that hybridize to polymorphic alleles at the locus. A DNA marker, or a genetic marker, may 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 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 molecular 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, 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 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 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.

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.

Marker loci that demonstrate statistically significant co-segregation with a disease resistance trait that confers broad resistance against a specified disease or diseases 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 breeding program to produce plants that have resistance to a disease or diseases.

Genetic mapping

It has been recognized for quite some time that specific genetic loci correlating with particular phenotypes, such as disease resistance, 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 are available for detecting molecular markers or clusters of molecular markers that co-segregate with a trait of interest, such as a disease 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.

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).

Marker loci that demonstrate statistically significant co-segregation with a disease 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 breeding programs to produce plants having disease resistance.

Activities in marker assisted 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 disease resistance trait are provided. A variety of methods 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 a disease resistance trait.

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.

Thus disclosed herein is an interval on chromosome 2 and an interval on chromosome 4. The disclosed interval on chromosome 2 can encompass 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 of the markers for GLS resistance disclosed in Table 10 of this disclosure. The disclosed interval on chromosome 4 can encompass 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, or 39 of the markers disclosed in Table 11 of this disclosure. Any marker located within these intervals can find use as a marker for GLS resistance and can be used in the context of the methods presented

herein to identify and/or select plants that have resistance to GLS, whether it is newly conferred or enhanced compared to a control plant. In certain embodiments, markers located upstream and downstream of PRR gene position are closely linked genetically and physically and hence may be used to select the PRR gene for trait introgression and products development.

Chromosomal intervals can also be defined by markers that are linked to (show linkage disequilibrium with) a disease 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. (2002) *supra*), 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 disease resistance trait, it is important to note that the marker locus is not necessarily responsible for the expression of the disease resistance phenotype. For example, it is not a requirement that the marker polynucleotide sequence be part of a gene that is responsible for the disease resistant phenotype (for example, is part of the gene open reading frame). The association between a specific marker allele and the disease resistance trait is due to the original "coupling" linkage phase between the marker allele and the allele in the ancestral 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 the 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 disease 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 disease resistance trait and hence can be used to predict disease resistance 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) could also be used to predict disease resistance in a 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. In the most preferred case, a marker is located within the gene itself, so that recombination cannot occur between the marker and the gene. In some embodiments, the methods disclosed herein produce a marker in a disease resistance gene, wherein the gene was identified by inferring genomic location from clustering of conserved domains or a clustering analysis.

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. Linkage drag may also result in reduced yield or other negative agronomic characteristics even after multiple cycles of backcrossing into the elite 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 disease 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. Using automated high throughput marker detection platforms makes this process highly efficient and effective.

Many of the markers presented herein can readily be used as single nucleotide polymorphic (SNP) markers to select for the PRR gene. 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 GLS disease resistance trait. Such markers are presumed to map near a gene or genes that give the plant its disease 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 GLS disease resistance may 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 disease 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 disease resistance.

The SNPs could be used alone or in combination (i.e. a SNP haplotype) to select for a favorable resistant gene allele associated with GLS disease resistance. For example, a SNP haplotype can include a combination of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 of the markers for GLS resistance in Table 10 of this disclosure. A SNP haplotype can also include a combination of 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, or 39 of the markers in Table 11 of this disclosure.

The skilled artisan would expect that there might be additional polymorphic sites at marker loci in and around a chromosome marker identified by the methods disclosed 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 disease resistance trait QTL.

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.

Proteins and Variants and Fragments Thereof

PRR polypeptides are encompassed by the disclosure. "PRR polypeptide" and "PRR protein" as used herein interchangeably refers to a polypeptide(s) having GLS resistance activity, and is sufficiently identical to the PRR polypeptide of SEQ ID NO:41 or SEQ ID NO:58. A variety of PRR 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 some embodiments the sequence identity is against the full-length sequence of a polypeptide. The term "about" when used herein in context with percent sequence identity means +/- 1.0 percentage point, relative to the recited percentage.

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 PRR polypeptide or polynucleotide, respectively, and that exhibit disease 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 a PRR 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:41 or SEQ ID NO:58, wherein the PRR polypeptide has GLS resistance when expressed in a plant.

Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of a PRR polypeptide may 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 an PRR polypeptide to confer disease resistance may, in some cases, be improved by the use of such techniques upon the compositions of this disclosure.

Nucleic Acid Molecules and Variants and Fragments Thereof

Isolated or recombinant nucleic acid molecules comprising nucleic acid sequences encoding PRR 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. In some examples, the nucleic acid molecule can be single-stranded. In some examples, the nucleic acid molecule can be double-stranded.

An "isolated" nucleic acid molecule (e.g., RNA or DNA) is used herein to refer to a nucleic acid sequence (e.g., RNA or DNA) that is no longer in its natural environment, for example *in vitro*. A "recombinant" nucleic acid molecule (e.g., RNA or DNA) is used herein

to refer to a nucleic acid sequence (e.g., RNA 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 PRR 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 PRR 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 PRR polypeptide is a non-genomic sequence.

A variety of polynucleotides that encode PRR polypeptides or related proteins are contemplated. Such polynucleotides are useful for production of PRR 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 PRR polypeptides or related proteins.

Provided herein are nucleic acid molecules encoding an PRR polypeptide. Such a polynucleotide cam have the sequence set forth in SEQ ID NO:41 or SEQ ID NO:58, 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. A reverse complement is a complement formed by exchanging each A with T, T with A, C with G, and G with C in a sequence and then reversing the 5' to 3' order of the exchanged sequence,

such that the reverse complement of 5'-ACCTGAG-3' is 5'-CTCAGGT-3'. "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 examples, the nucleic acid molecule encoding the PRR 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 examples, 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 examples, the non-genomic nucleic acid molecule is a synthetic nucleic acid sequence.

In some examples, the nucleic acid molecule encoding an PRR 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:41 or SEQ ID NO:58, wherein the PRR polypeptide has GLS resistance activity when expressed in a plant.

In some examples, the nucleic acid molecule encodes an PRR polypeptide variant comprising one or more amino acid substitutions relative to the amino acid sequence of SEQ ID NO:41 or SEQ ID NO:58.

Nucleic acid molecules that are fragments of these nucleic acid sequences encoding PRR polypeptides are also encompassed by the disclosure. "Fragment" as used herein refers to a portion of the nucleic acid sequence encoding an PRR polypeptide. A fragment of a nucleic acid sequence may encode a biologically active portion of an PRR 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 PRR 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 PRR polypeptide identified by the methods 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 will encode protein fragments that retain the biological activity of the PRR polypeptide and, hence, retain disease resistance. "Retains disease 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 disease resistance of the full-length PRR polypeptide as set forth in SEQ ID NO:41 or SEQ ID NO:58.

"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, 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 examples, a PRR polynucleotide encodes a PRR 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:41 or SEQ ID NO:58. In some examples, a PRR polynucleotide comprises genomic sequence, including introns, regulatory elements, and untranslated regions.

The disclosure also provides nucleic acid molecules encoding PRR polypeptide variants. "Variants" of PRR polypeptide encoding nucleic acid sequences include those sequences that encode the PRR polypeptides identified by the methods 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 PRR 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 PRR 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 examples 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. 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, e.g. 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 disclosed herein 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 identified by the methods disclosed 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 disclosure. Such sequences include sequences that are orthologs of the 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.

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 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 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 ^{32}P 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 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 polypeptides or a fragment or variant thereof. Methods for the preparation of probes for hybridization and stringency conditions are disclosed in Sambrook and Russell (2001), *supra*.

Nucleotide Constructs, Expression Cassettes and Vectors

The use of the term "construct" in connection with isolated and/or heterologous polynucleotides herein is not intended to limit the disclosure to constructs comprising DNA. Polynucleotide constructs, particularly polynucleotides and oligonucleotides composed of ribonucleotides and combinations of ribonucleotides and deoxyribonucleotides, may also be employed in the methods disclosed herein. The isolated polynucleotide constructs, nucleic acids, and nucleotide sequences disclosed herein additionally encompass all complementary forms (e.g., the reverse complement) of each sequence disclosed for such a construct. Further, polynucleotide constructs and nucleotide sequences disclosed herein can encompass any such constructs, molecules, and sequences suitable for use in a method for transforming plant material disclosed herein. Such constructs can include naturally occurring molecules and/or synthetic analogues. The disclosed nucleotide constructs, nucleic acids, and nucleotide sequences 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.

Transformed organisms disclosed herein include plant cells, bacteria, yeast, baculovirus, protozoa, nematodes and algae. The transformed organism comprises a disclosed sequence (e.g., as part of a construct, expression cassette, or vector comprising the nucleotide sequence disclosed herein which are associated with GLS disease resistance.

The disclosed sequences can be used in constructs for expression in the organism of interest. Constructs can include 5' and 3'; regulatory sequences operably linked to a coding sequence for a PRR polypeptide disclosed herein. The term "operably linked" as used herein refers to a functional linkage between a promoter and/or a regulatory sequence and a second sequence, wherein the promoter and/or regulatory sequence initiates, mediates, and/or affects 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 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 (e.g., a promoter), a DNA sequence of the embodiments, and a transcriptional and translational termination region (e.g., termination region) functional in the organism serving as a host. The transcriptional initiation region (e.g., the promoter) may be native, analogous, foreign or heterologous to the host organism and/or to the sequence of the embodiments. Additionally, the promoter or regulatory sequence 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. As used herein, the term "heterologous" in reference to a sequence means a sequence that originates from a foreign species or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. 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 PRR polypeptide of the embodiments. In some embodiments the DNA construct comprises a polynucleotide encoding a fusion protein comprising an PRR 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 include, 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 plant-preferred for a particular amino acid may be derived from known gene sequences from plants.

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 maize 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 include, but are 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, polynucleotides encoding a PRR polypeptide can be introduced into the genome of a plant using genome editing technologies. For example, the identified 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, a PRR gene 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. Thus, for example, a PRR gene can be altered though gene editing in its native site to encode a PRR polypeptide having the amino acid sequence set forth in SEQ ID NO:41 or SEQ ID NO:58. Alternatively or additionally, a PRR gene can be introduced by genome editing at a different genomic location. For example, a nucleotide construct encoding a polypeptide having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% amino acid sequence identity to SEQ ID NO:58 can be inserted at a genomic locus other than chromosome 02, or other than the chromosome 02 interval between 220 and 255 cM. In another example, a nucleotide construct encoding a polypeptide having at least 50%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% amino acid sequence identity to SEQ ID NO:41 can be inserted at a genomic locus other than chromosome 04, or other than the chromosome 04 interval between 90 and 115 cM.

In some embodiments, where an GLS resistance PRR gene allele has been identified in a genome, genome editing technologies may be used to alter or modify the polynucleotide sequence. Site specific modifications that can be introduced into the desired PRR gene allele polynucleotide 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 disease resistant proteins in close proximity to the PRR polynucleotide compositions within the genome of a plant, in order to generate molecular stacks disease 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. QTL mapping

Gray Leaf Spot (GLS), caused by the fungal pathogen *Cercospora zeae-maydis* is a destructive foliar disease of maize, responsible for consistent and significant yield losses. (Wiebold et al. 2020). In order to identify native maize genes that can confer resistance to GLS, mapping populations were created by crossing the GLS-resistant line Inbred A with the susceptible inbreds as described in WO2018013323. The plant material was backcrossed to the susceptible inbreds to create mapping populations. In a separate project, mapping populations were created by making an F1 between the resistant inbred Inbred B and a susceptible inbred (Inbred C). First generation back-cross (BC1) plants were initially screened to determine segregation of the phenotype indicating a major, dominant QTL. Plants were inoculated with *C. zeae-maydis* spores on a Sorghum seed carrier 3-4 times between the V10 and V15 growth stages. Scoring was done 4-6 weeks after flowering using a visual score for GLS severity (GLFSPT) based on the leaf area affected by lesions caused by *C. zeae-maydis*. Severity was determined based on a scale of 1-9, with 9 being the most resistant and 1 being the most susceptible. Scores of 1-3 are considered susceptible, scores between 4-6 are intermediate, and scores 7-9 are classified as resistant. Individual plants

were genotyped with SNP markers for marker trait association analysis. Mapping data was analyzed with the TIBCO® Spotfire® software package (Version 10.3.3) using the Kruskal-Wallis method for comparing numeric and categorical variables. Initial QTL mapping from the Inbred A source (Inbred A Table 2) placed the quantitative trait locus (QTL) for GLS resistance between 94.78 and 113.94 cM on Chr04 with a peak correlation at 111.72 cM (PHM586-10). Initial QTL Mapping from the Inbred B source (Inbred B Table 1) placed the QTL on Chr02 between 220.4 and 252.78 with a peak correlation at 252.78 (PHM2363-23). Table 1 provides a p-value representing the correlation of the genotype at a given genetic and physical position with the gray leaf spot phenotype in Inbred A. Genetic positions are based on an internal proprietary the B73 maize line gene map and physical genome positions are based on the published genome sequence of B73 maize line (Version 2). See Schnable et al. (2009) *Science* 326(5956): 1112-5. Table 2 provides a p-value representing the correlation of the genotype at a given genetic position with the gray leaf spot phenotype in Inbred B.

Table 1

Name	Map Position (B73 v2)	Physical Position (Public B73 v2)	p-value	Trait Value RP Allele	Trait Value Het	Trait Value Donor Allele
PHM6764-7	94.78	36901556	2.85E-16	4.88	5.28	5.34
PHM16360-9	104.12	66732567	3.80E-29	4.77	5.32	5.37
PHM521-8	108.38	91395407	8.76E-37	4.74	5.34	5.39
PHM586-10	111.72	141279810	2.57E-43	4.69	5.37	5.37
PHM289-20	113.94	144840683	1.57E-29	5.1	5.25	5.36

Table 2

Name	Map Position (B73 v2)	Physical Position (Public B73 V2)	p-value	Trait Value RP Allele	Trait Value Donor Allele
C01281-1	220.4	225162286	1.14E-01	5.00	5.50
C01685-2	224.69	228044954	8.98E-02	4.95	5.15
C01267-2	230.9	231332773	2.89E-02	4.82	5.64
C00682-2	235.38	232372317	1.99E-02	4.82	5.69
C00572-3	247.31	234537676	1.78E-02	4.89	5.56
PHM2363-23	252.78	235441327	6.90E-04	4.68	5.69

Example 2. QTL fine mapping and candidate gene identification

To fine map the resistance gene(s), Inbred A segregating material was further backcrossed to the susceptible parents to generate BC2, BC3 or BC4 segregating populations, screen for recombination events in the region and characterize these plants using the previously described approach. Additional SNP markers were generated for fine mapping using a combination of 56K SNPs as well as exome capture data.

Results from phenotyping selected recombination events in Inbred A are presented in Tables 3-5. The p-value represents the correlation of the genotype at a given genetic position with the GLS phenotype. Genetic positions are based on the Corteva B73 genome (Version 2). Peak correlation between genotype and phenotype occurred at 111.49-111.74 cM on Chr04 and the associated markers are shown in bold. Flanking markers listed in Tables 3-5 are indicated by asterisk (*). Table 3 shows marker trait analysis for chromosome 4 of Inbred A GLS QTL based on results of an experiment in 2016. Table 4 shows marker trait analysis for chromosome 4 GLS QTL of an Inbred A based on a first experiment done in 2017. Table 5 shows marker trait analysis for chromosome 4 GLS QTL of Inbred A based on a second experiment in 2017.

Table 3

Name	GEN_POS	PHYS_POS (Public B73 v2)	p-value	Trait Value RP Allele	Trait Value Donor Allele
PHM521-8	108.38	91395407	5.19E-02	4.56	4.84
PHM199-23*	110.50	124636222	3.24E-02	4.57	4.86
GLS_G1	111.34	135312626	1.16E-05	4.40	4.97
GLS_G14	111.38	135912644	1.50E-05	4.39	4.95
GLS_G19	111.44	136668804	2.97E-04	4.52	4.98
GLS_G21	**111.49**	**138159108**	**2.86E-05**	**4.47**	**5.02**
SYN21168	111.68	139440592	9.26E-05	4.46	4.98
GLS_G45	111.68	139441329	3.28E-05	4.44	4.98
C001YAR-001*	111.72	139986905	2.57E-02	4.56	4.86
PHM586-10	111.72	141279810	1.99E-04	4.32	4.93
PHM5013-12	112.65	141270992	1.99E-04	4.32	4.93
PHM289-20	113.94	144840683	6.79E-04	4.46	5.10

Table 4

Name	GEN_POS	PHYS_POS (Public B73v2)	p-value	Trait Value RP Allele	Trait Value Donor Allele
PHM199-23	110.5	124636222	1.50E-01	5.49	5.32
GLS_G1	111.34	135312626	1.00E-02	5.29	5.81
GLS_G7	111.38	135909419	8.04E-03	5.29	5.98
GLS_G19 *	111.44	136668804	1.57E-01	5.29	5.98

GLS_G21	111.49	138159108	4.80E-01	4.38	5.40
AMD1	111.54	137892875	7.28E-07	4.90	5.78
AMD4	111.55	137984468	7.28E-07	4.90	5.78
AMD14	111.57	138186349	7.28E-07	4.90	5.78
AMD20	111.58	138187436	7.28E-07	4.90	5.78
GLS_G45	111.68	139441329	7.28E-07	4.90	5.78
C001YAR-001	111.72	139986905	9.93E-07	4.97	5.78
PHM5013-12*	112.6500015	141270992	7.73E-04	5.11	5.59
PHM289-20	113.9400024	144840683	1.34E-01	4.93	5.45

Table 5

Name	GEN_POS	PHYS_POS (Public B73 v2)	p-value	Trait Value RP Allele	Trait Value Donor Allele
C103H6N-001	111.38	135909419	2.93E-01	5.38	5.45
C103H6R-001	111.44	136668804	5.06E-03	5.16	5.62
ZmChr4v2_142171204	111.47	136935833	3.06E-03	5.26	5.73
ZmChr4v2_142171261	111.47	136935890	1.01E-02	5.16	5.57
ZmChr4v2_142192918	111.47	136960918	1.24E-02	5.19	5.64
ZmChr4v2_142192934 *	111.47	136960934	7.29E-03	5.16	5.60
ZmChr4v2_143270779	111.55	137984468	2.42E-05	4.96	5.83
C103H6T-001	111.57	138159108	1.14E-05	4.96	5.83
C103HUA	111.57	138186063	1.76E-04	4.99	5.82
ZmChr4v2_143572529	111.57	138186349	2.42E-05	4.96	5.82
C103H6U-001	111.68	139441329	1.41E-06	4.82	5.82
ZmChr4v2_145458082	111.73	140005890	9.67E-07	4.82	5.83
ZmChr4v2_145460875	111.74	140008683	9.67E-07	4.82	5.83
ZmChr4v2_148727396 *	113.5	143087377	6.56E-03	5.06	5.69
ZmChr4v2_148728091	113.5	143088072	9.15E-05	4.93	5.7
ZmChr4v2_149477191	113.83	143804372	4.58E-04	5.06	5.68
ZmChr4v2_149477479	113.83	143804660	5.63E-01	5.22	5.65

To fine map the region comprising the resistance gene(s) from Inbred B, further backcrossing of selected recombinant individuals to the susceptible parent (Inbred C) followed by selfing of key recombinant individuals were performed to generate additional populations containing new recombinant events that were characterized phenotypically as well as

genotypically with markers developed within the QTL region to further fine-map the region. Additional SNP markers were generated for fine mapping using a combination of 56K SNPs as well as additional genomic sequence data from Inbred B and Inbred C.

Results from phenotyping selected recombination events from Inbred B are shown in Tables 6-9. The p-value presented represents the correlation of the genotype at a given genetic position with the GLS phenotype. Genetic positions are based on the Corteva B73 genome (Version 2). Peak correlation between genotype and phenotype occurred at 240.31 cM on Chr02 and this marker, InbB_C2_GLS_80 is shown in bold. Flanking markers listed in Tables 6 are indicated by asterisk (*). Table 6 provides marker trait analysis for Chr02 Inbred B GLS QTL in 2019. Table 7 provides marker trait analysis for Chr02 Inbred B GLS QTL in 2020. Tables 8 and 9 provide marker trait analysis for Chr02 Inbred B GLS QTL from hybrid data in 2019 and 2020, respectively.

Table 6

Name	Map Position (B73 v2)	Physical Position (Public B73 V2)	p-value	Trait Value RP Allele	Trait Value Donor Allele
PHM15493-24	230.55	231209991	1.07E-02	4.88	5.85
C00682-2	235.38	232372317	3.47E-03	4.59	6.08
C2_GLS_38	237.04	232594769	5.66E-05	4.47	6.33
C2_GLS_68	239.27	232878324	2.41E-05	4.25	6.37
C2_GLS_43*	239.97	232966424	3.20E-06	4.11	6.40
C2_GLS_80	**240.31**	**233012091**	**4.96E-07**	**4.11**	**6.40**
C2_GLS_47*	243.37	233426438	5.10E-06	4.21	6.31
C2_GLS_48	245.84	233693108	9.10E-05	4.24	6.11
PHM2363-23	252.78	235441327	1.41E-02	5.14	5.76

Table 7

Name	Map Position (B73 v2)	Physical Position (Public B73 V2)	p-value	Trait Value RP Allele	Trait Value Donor Allele
C2_GLS_38	237.04	232594769	4.15E-02	4.03	4.88
C2_GLS_68	239.27	232878324	2.55E-03	3.84	5.02
C2_GLS_43	239.97	232966424	2.61E-07	3.36	5.39
C2_GLS_80	**240.31**	**233012091**	**1.58E-10**	**3.14**	**5.67**
C2_GLS_87	241.08	233118574	1.05E-09	3.26	5.70
C2_GLS_47	243.37	233426438	7.38E-06	3.63	5.48
PHM2363-23	252.78	235441327	8.67E-02	4.12	4.53

Table 8

Name	Map Position (B73 v2)	Physical Position (Public B73 V2)	p-value	Trait Value RP Allele	Trait Value Donor Allele
C00682-2	235.38	232372317	1.50E-04	3.083	5.15
C2_GLS_38	237.04	232594769	2.24E-06	3.80	5.67
C2_GLS_68	239.27	232878324	8.33E-08	3.80	5.73
C2_GLS_43	239.97	232966424	1.47E-06	3.81	6.00
C2_GLS_80	240.31	233012091	8.33E-08	3.80	5.73
C2_GLS_47	**243.37**	**233426438**	**7.16E-08**	**3.80**	**5.80**
C2_GLS_48	245.84	233693108	1.95E-06	3.80	5.30
PHM2363-23	252.78	235441327	6.88E-02	4.03	4.45

Table 9

Name	Map Position (B73 v2)	Physical Position (Public B73 V2)	p-value	Trait Value RP Allele	Trait Value Donor Allele
C2_GLS_38	237.04	232594769	2.49E-05	3.56	4.38
C2_GLS_68	239.27	232878324	7.07E-05	3.52	4.35
C2_GLS_43	239.97	232966424	7.94E-06	3.45	4.39
C2_GLS_80	**240.31**	**233012091**	**1.44E-06**	**3.45**	**4.37**
C2_GLS_87	241.08	233118574	1.44E-06	3.45	4.37
C2_GLS_47	243.37	233426438	9.02E-06	3.50	4.37
PHM2363-23	252.78	235441327	7.73E-01	3.85	3.97

To obtain candidate genes in mapping intervals, finished whole genome sequences were generated for Inbred A and Inbred B. RNAseq reads from the NILs were used to refine gene models for the Chr04 region and RNAseq reads from Inbred B were used to refine gene models for the Chr02 region. Based on FGENESH gene model predictions, there are 7 pattern recognition receptor (PRR) genes in the Inbred A Chr04 region and 2 PRR genes in the Inbred B Chr02 region. These are the only genes in the intervals with known roles in host defense against pathogens. Due to the large physical but small genetic size of the Chr04 fine mapping interval, further fine mapping would have required screening many thousands of plants with no guarantee of success, so the genes were tested after use of proprietary bioinformatic identification methods rather than attempting to identify specific genes causing the effect via recombination. Table 10 shows SNP and Position of GLS resistance markers for Inbred B. Table 11 shows SNP and Position of GLS resistance markers for Inbred A.

Table 10

Name	Susceptible	Inbred B	SNP Pos in Seq	SEQ ID NO:
C2_GLS_38	G	A	56	42
C2_GLS_68	C	G	55	43
C2_GLS_43	G	A	51	44
C2_GLS_80	C	T	62	45
C2_GLS_87	G	C	201	46
C2_GLS_47	A	G	61	47
C2_GLS_48	C	T	61	48
C01281-1	T	C	118	49
C01685-2	A	G	125	50
C01267-2	A	G	143	51
C00682-2	C	T	244	52
C00572-3	C	C or G	172	53
PHM2363-23	G	A	64	54

Table 11

Name	Susceptible allele	Inbred A allele	SNP Pos in Seq	SEQ ID NO:
C103H6N-001	T	C	51	19
C103H6R-001	T	C	51	20
ZmChr4v2_142171204	T	C	201	21
ZmChr4v2_142171261	A	G	200	22
ZmChr4v2_142192918	T	C	200	23
ZmChr4v2_142192934	T	G	200	24
ZmChr4v2_143270779	A	G	200	25
C103H6T-001	T	C	51	26
C103HNM	A	C	201	27
C103HUA	T	C	201	28
ZmChr4v2_143572529	T	C	201	29
C103H6U-001	T	C	51	30
ZmChr4v2_145458082	C	T	201	31
ZmChr4v2_145460303	T	C	201	32
ZmChr4v2_145460875	C	T	201	33
ZmChr4v2_148727396	A	G	201	34
ZmChr4v2_148728091	C	T	201	35
ZmChr4v2_149477191	G	C	201	36
ZmChr4v2_149477479	G	C	201	37
PHM521-8	C	T	244	1
PHM199-23	A	T	243	2
GLS_G1	A	T	51	3
GLS_G7	T	C	51	4
GLS_G14	A	G	51	5

GLS_G19	T	C	51	6
GLS_G21	T	C	51	7
AMD1	G	A	51	14
AMD4	G	A	51	15
AMD12	C	A	51	16
AMD14	C	T	51	17
AMD20	G	A	51	18
SYN21168 (C002T5R-001)	G	T	61	8
GLS_G45	T	C	51	9
C001YAR-001	G	A	51	10
PHM586-10	C	T	114	11
PHM5013-12	T	C	107	12
PHM289-20	T	C	121	13
PHM6764-7	G	A	150	59
PHM16360-9	C	G	121	60

Example 4. Transgenic validation of PRR (PRR03 and PRR01) candidate genes

All 7 PRR genes from the Inbred A interval and the PRR01 gene from Inbred B were tested for efficacy in susceptible backgrounds. A transgenic construct for PRR03 contained the native promoter (1500bp) (SEQ ID NO:39), native coding sequence (3903bp) (SEQ ID NO:40) and the native maize terminator (473bp) (SEQ ID NO:61). PRR03 (SEQ ID NO:40 encoding SEQ ID NO:41) from Inbred A showed field efficacy with T1 segregating seed with an effect size of 2.2-2.4 on the gray leaf spot scale. Table 12 shows the enhanced resistance to gray leaf spot disease (larger values indicate better resistance) provided by this PRR03 construct.

Table 12

Allele state	Average of GLS
Hemizygous	6.0
Homozygous	6.8
Null	4.6

A transgenic construct for PRR01 from Inbred B contained the native promoter (2000bp) (SEQ ID NO:55), native coding sequence (7418bp) (SEQ ID NO:56), and native terminator (1000bp) (SEQ ID NO:57). PRR01 (SEQ ID NO:56 encoding SEQ ID NO:58) from Inbred B showed greenhouse assay efficacy in 2020 using T1 segregating seed with an effect size of 6.1 on the GLS scale. Table 13 shows the enhanced resistance to gray leaf spot disease (larger values indicate better resistance) provided by this PRR01 construct.

Table 13

Allele State	Average GLS
Homozygote	7.9
Null	1.8

CLAIMS

WHAT IS CLAIMED:

1. A method of identifying a plant comprising one or more marker alleles associated with increased resistance to gray leaf spot, said method comprising:
    a. obtaining a nucleic acid sample from a plant, seed, tissue or germplasm thereof;
    b. screening the sample for a marker allele sequence associated with increased resistance to gray leaf spot, wherein said marker allele comprises an "A" at position 56 of SEQ ID NO:42; a "C" at position 55 of SEQ ID NO:43; an "A" at position 51 of SEQ ID NO:44; a "T" at position 62 of SEQ ID NO:45; a "C" at position 201 of SEQ ID NO:46; a "G" at position 61 of SEQ ID NO:47; a "T" at position 61 of SEQ ID NO:48; a "C" at position 118 of SEQ ID NO:49; a "G" at position 125 of SEQ ID NO:50; a "G" at position 143 of SEQ ID NO:51; a "T" at position 244 of SEQ ID NO:52; a "C" or "G" at position 172 of SEQ ID NO:53; an "A" at position 55 of SEQ ID NO:53; a "C" at position 51 of SEQ ID NO:19; a "C" at position 51 of SEQ ID NO:20; a "C" at position 201 of SEQ ID NO:21; a "G" at position 200 of SEQ ID NO:22; a "C" at position 200 of SEQ ID NO:23; a "G" at position 200 of SEQ ID NO:25; a "C" at position 51 of SEQ ID NO:26; a "C" position 201 of SEQ ID NO:27; a "C" at position 201 of SEQ ID NO:28; a "C" at position 201 of SEQ ID NO:29; a "C" at position 51 of SEQ ID NO:30; a "T" position 201 of SEQ ID NO:31; a "C" at position 201 of SEQ ID NO:32; a "T" at position 201 of SEQ ID NO:33; a "G" at position 201 of SEQ ID NO:34; a "T" at position 201 of SEQ ID NO:35; a "C" at position 201 of SEQ ID NO:36; a "C" at position 201 of SEQ ID NO:37; a "T" at position 244 of SEQ ID NO:1; a "T" position 243 of SEQ ID NO:2; a "T" position 51 of SEQ ID NO:3; a "C" at position 51 of SEQ ID NO:4) a "G" position 51 of SEQ ID NO:5; a "C" at position 51 of SEQ ID NO:6; a "C" at position 51 of SEQ ID NO:7, an "A" at position 51 of SEQ ID NO:14; an "A" at position 51 of SEQ ID NO:15; an "A" at position 51 of SEQ ID NO:16; a "T" at position 51 of SEQ ID NO:17; an "A" at position 51 of SEQ ID NO:18; a "T" at position 61 of SEQ ID NO:8; a "C" at position 51 of SEQ ID NO:9; an "A" at position 51 of SEQ ID NO:10; a "T" at position 114 of SEQ ID NO:11; a "C" at position 107 of SEQ ID NO:12; a "C" at position 121 of SEQ ID NO:13; an "A" at position 150 of SEQ ID NO:59; and a "G" at position 121 of SEQ ID

NO:60); wherein the presence of the marker allele is associated with increased resistance to gray leaf spot.

2. The method of claim 1, further comprising
   a. obtaining a nucleic acid sample from one or more plants, seeds, tissues or germplasm in a population;
   b. screening each sample in accordance with claim 1; and
   c. selecting one or more of the plants, seeds, tissues or germplasm having the marker allele sequence associated with increased resistance to gray leaf spot.

3. The method of claim 1, further comprising
   a. obtaining a nucleic acid sample from one or more plants, seeds, tissues or germplasm, each sample being representative of a plurality of plants, seeds, tissues or germplasm;
   b. screening each sample in accordance with claim 1; and
   c. selecting one or more plurality of plants, seeds, tissues or germplasm, wherein the representative sample for each selected plurality has the marker allele sequence associated with increased resistance to gray leaf spot.

4. A method of identifying a plant comprising a qualitative trait locus (QTL) associated with increased resistance to gray leaf spot, said method comprising:
   a. obtaining a sample comprising nucleic acid from a plant;
   b. screening the sample for any of the following:
      i. a polynucleotide encoding a polypeptide having the amino acid sequence set forth in SEQ ID NO:41 or SEQ ID NO:58;
      ii. a polynucleotide comprising a sequence having at least 90% nucleotide sequence identity to SEQ ID NO:40 or SEQ ID NO:56; or
      iii. one or more marker alleles within 5 cM of (i) or (ii) that are linked to and associated with (i) or (ii); and
   c. detecting any of (i), (ii), or (iii) in the sample and thereby identifying the plant as having the QTL associated with increased resistance to gray leaf spot.

5. A method of increasing resistance to gray leaf spot in plant material comprising introducing into the genome of the plant material a heterologous nucleic acid sequence

or expressing the heterologous nucleic acid sequence in the plant material, wherein the heterologous nucleic acid sequence (i) comprises a sequence having at least 90% nucleotide sequence identity to SEQ ID NO:40 or SEQ ID NO:56 or (ii) having at least 95% encodes a polypeptide comprising SEQ ID NO:41 or SEQ ID NO:58; wherein the plant expressing the heterologous polynucleotide has increased resistance to gray leaf spot in the plant, as compared to a control plant that does not express the heterologous polynucleotide.

6. The method of claim 5, wherein the heterologous polynucleotide further comprises a heterologous promoter.

7. The method of claim 5, further comprising obtaining a progeny plant derived from the plant material expressing the heterologous polynucleotide, wherein said progeny plant comprises in its genome the heterologous polynucleotide and exhibits increased resistance to gray leaf spot as compared to a control plant that does not express the heterologous polynucleotide.

8. The method of claim 5, wherein the plant material's genome is altered by gene editing or by transgenic modification to include the heterologous nucleic acid sequence.

9. The method of claim 5, wherein said plant is a monocot.

10. A method of generating a variant of a PRR gene, the method comprising the steps of:
    a. gene shuffling one or more nucleotide sequences encoding one or more fragments of SEQ ID NO:41 or SEQ ID NO:58, a protein that is at least 90% identical to any one of SEQ ID NO:41 or SEQ ID NO:58, or a fragment thereof, to generate variants of the PRR gene; and
    b. testing the variants for resistance to gray leaf spot.

11. The method of claim 10 wherein the method further comprises the steps of:
    a. introducing into a regenerable plant cell a recombinant construct comprising a variant of the PRR gene generated 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 gray leaf spot, as compared to a control plant that does not comprise the recombinant DNA construct.

12. The method of claim 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.

13. The method of claim 10, wherein said plant is a monocot.

14. The method of claim 13, wherein said monocot is maize.

15. A method of identifying an allelic variant of the PRR gene wherein said allelic variant is associated with increased tolerance to gray leaf spot, the method comprising the steps of:

a. obtaining a population of plants, wherein said plants exhibit differing levels of gray leaf spot resistance;
b. evaluating allelic variations with respect to a polynucleotide sequence encoding a protein comprising SEQ ID NO:41 or SEQ ID NO:58, or in the genomic region that regulates the expression of the polynucleotide encoding the protein;
c. associating allelic variations for variations in resistance to gray leaf spot; and
d. identifying allelic variations associated with increased resistance to gray leaf spot.

16. The method of claim 15, further comprising detecting said allelic variant associated with increased resistance to gray leaf spot and selecting a plant if said allelic variant is detected.

17. A method of introducing an allelic variant of a PRR gene wherein said allelic variant is associated with increased resistance to gray leaf spot, the method comprising introducing a mutation in the endogenous PRR gene such that the allelic variant comprises a polynucleotide sequence encoding SEQ ID NO:41 or SEQ ID NO:58.

18. A recombinant DNA construct comprising a polynucleotide operably linked to at least one heterologous regulatory sequence wherein said polynucleotide comprises a nucleic acid sequence encoding SEQ ID NO:41 or SEQ ID NO:58.

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

20. The recombinant DNA construct of claim 18, wherein the polynucleotide comprises a sequence having at least 95% nucleotide sequence identity to any one of SEQ ID NO:40 or SEQ ID NO:56.

21. A transgenic plant or transgenic plant cell comprising the recombinant DNA construct of any one of claims 18-20.

22. The transgenic plant or transgenic plant cell of claim 21, 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.

23. Transgenic seed produced from a transgenic plant in accordance with claim 21 or claim 22.