Chapter 6
Genome Wide Association Study (GWAS) on Disease Resistance in Maize

Vivek Shrestha, Mani Awale, and Avinash Karn

6.1 Introduction

Crop losses due to disease infestation account for a major loss to farmers worldwide. Scientists and farmers have long adopted many traditional and chemical methods to control diseases in crop plants. Intensive use of chemicals, mostly fungicides and pesticides, has been able to control the disease infestation to some extent, but they come at a price. The use of chemicals does not only increase the cost of production, but it also negatively impacts the environment and health of both farmers and the consumers. Often times, the pathogen develops resistivity toward the fungicides being applied on them, which are no longer effective to control the damage that affect the crop yield and quality. Therefore, one sustainable way to reduce the impact of crop yield and quality loss due to plant diseases is to develop disease-resistant crops. Disease resistance breeding has been a major source of disease control (Hammond-Kosack and Jones 1997; Balint-Kurti and Johal 2009; Poland et al. 2009). The simplest way to define disease resistance breeding is the introgression of the disease resistance genes in the plants infected with the disease. The source of the resistance genes is either natural or induced. Disease resistance is generally categorized as qualitative and quantitative resistance. Qualitative resistance is based on a single dominant or recessive gene, race-specific and usually confers a high level of resistance, whereas the quantitative resistance is based on the oligogenic or polygenic inheritance and governed by additive or partial dominant

V. Shrestha (✉)
Division of Biological Sciences, University of Missouri, Columbia, MO, USA
e-mail: vs6d9@mail.missouri.edu

M. Awale
Division of Plant Sciences, University of Missouri, Columbia, MO, USA

A. Karn
School of Integrative Plant Science, Cornell AgriTech, Geneva, NY, USA

genes and generally race-nonspecific (Wisser et al. 2006). Quantitative disease resistance is given more importance by the plant breeders as it is more durable and has broader specificity (Parlevliet 2002; Poland et al. 2009).

One of the widely known theories regarding the disease resistance/susceptibility is H.H. Flor's gene-for-gene interaction theory (Flor 1971). Gene-for-gene interaction theory originated using flax (*Linumusitatissimum*) as host plant and fungal rust pathogen, *Melampsora lini*. The theory states that, a host plant requires a dominant or semi-dominant resistance R gene with a corresponding avirulence *(Avr)* gene in the pathogen to be disease resistant. R genes are responsible in the plants to detect the *Avr* gene-specific pathogen molecules, resulting in the downstream signal cascades to produce defensins, which trigger defense (Hammond-Kosack and Jones 1997). Hypersensitivity response is commonly seen as a defense response triggering the incompatibility reaction between host and pathogen. Modification or complete loss of the R gene or the *Avr* gene results in disease susceptibility, which is commonly seen in biotrophic pathogens, such as fungi, bacteria, viruses, and nematodes. R gene generally encodes proteins that recognize the pathogen effectors or modification of plant proteins that are the targets of the effectors (Nimchuk et al. 2003). Among the six known classes of R-genes, the most known class is the nucleotide binding, leucine-rich repeat (NB-LRR) amino acids sequence motifs, which are involved in the pathogen recognition and related functions. Unraveling the structural and functional roles of these R genes will be beneficial to improve disease resistance in plants. There were several R genes being identified, isolated, and cloned. The first R gene to be isolated was *Hm1* from maize, which is responsible for resistance against the leaf spot fungus *Cochliobolus carbonum* (Johal and Briggs 1992). *Hm1* encodes for a reductase enzyme that detoxifies the *C. carbonum* HC-toxin. On invasion by the biotrophic pathogens, R-genes confers an effective defense response, usually, by involving in a hypersensitive response, where the tissue immediately adjacent to the site of the pathogen undergoes rapid programmed cell death (Poland et al. 2009). Some other early cloned R genes were *Pto* gene (Martin et al. 1993) that encodes for serine threonine kinase, in tomato against the *Pseudomonas syringae*. Other cloned genes for *Pseudomonas syringae* are *RPS2* in Arabidopsis (Bent et al. 1994), a NBS/LRR protein family. *Xa-21* (Song et al. 1995) gene in rice for *Xanthomonas oryzae* pv. *oryzae*. The list and details of several other important cloned genes can be found in Hammond-Kosack and Jones (1997).

Although the R genes are found to be effective in biotrophic fungus, studies show that the R-genes might not work in a similar fashion against the necrotrophic pathogens. For instance, the hypersensitive response can increase the susceptibility to necrotrophic pathogens, instead of increasing resistance. Although there is immense potential of disease resistance by R genes, observation of the performance of the crop cultivars with different types of resistance have led to the conclusion that quantitative disease resistance is more durable than the typical R-gene mediated resistance (Parlevliet 2002; Poland et al. 2009).

Molecular mechanism of gene-for-gene interaction theory or the host-pathogen interaction were not well known until recent work by Jones and Dangl (2006),

where they studied the molecular mechanism of host-pathogen interaction mainly pertaining to the biotrophic pathogen (Jones and Dangl 2006). They elaborated on the complex multiphase host-pathogen interplay. Briefly, they explained that the plant (host), when invaded by a pathogen, initially recognizes some common feature of those pathogen, which are called microbial associated molecular patterns (MAMPs), using pattern recognizing receptors. For instance, flagellin in bacteria and chitin in fungus are the MAMPs. This recognition event triggers the innate immune response in host plant, also known as the first line of defense or host basal defense, which defends further pathogen invasion or development. In response to that MAMP-triggered immunity of the host, virulent pathogen fight back, releasing effector proteins, which destroy the host basal defense. This triggers the host plant to further use its second line of defense, using resistance genes such as NB-LRR; these recognize the pathogen secreted effector proteins and finally destroy them. This immunity is known as the effector-triggered immunity. The molecular understanding of the host pathogen interaction will definitely be beneficial for the development of disease resistance in plants; however, there are still a lot of unknowns in the field of host pathogen interaction.

Maize has been a model plant for many plant scientists. It has a long history of research in disease resistance. Maize southern leaf blight is one of the biggest epidemics ever known in history, causing loss of a billion dollars to the US economy. Hybrid seed production using the maize carrying Texas cytoplasm for male sterility (cms-T) was popular in the era of 1950s. A race of the southern corn leaf blight called race T was found to be very pathogenic on cms-T maize, causing epidemic in 1970 and 1971 (Ullstrup 1972). It was found later that the pathogen race T produces T-toxin (Pring and Lonsdale 1989) (a family of linear long chain polyketides) that binds specifically to the URF13. URF13 is a peptide of 13 kDa that resides in the inner membrane of mitochondria and acts as a ligand-gated channel (Levings and Siedow 1992). The interaction between the T-toxin and the URF13 transforms the channel to a large pore, causing the membrane to be leaky, and ultimately leading to the cell death. Since then cms-T were eliminated from the elite germplasm, and then polygenic disease resistance studies were introduced (Balint-Kurti and Carson 2006; Balint-Kurti and Johal 2009). The first gene to be cloned in maize is the *Hm1* genes, which confer specific resistance against a leaf blight and ear mold disease of corn caused by *C. carbonum* race 1 (CCR1).The pathogen produces a toxin called HC-toxin. This gene was cloned using transposon tagging and was found to be an NADPH-dependent HC-toxin reductase, which inactivates the HC-toxin by reducing the key carbonyl group on HC-toxin (Johal and Briggs 1992). Studies have shown that 228 R gene analogs have been identified in maize, using the partial sequence data derived from several different maize lines (Xiao et al. 2007). Wisser et al. (2006) studied 50 publications regarding the disease resistance gene in maize, which included 437 QTL and 17 major genes (Wisser et al. 2006). For a more comprehensive review of maize disease management, it is suggested to the readers to look in the following reviews (Pratt and Gordon 2006; Wisser et al. 2006; Balint-Kurti and Johal 2009).

6.2   Association Mapping Versus QTL Mapping

Quantitative or metric traits are those traits which can be measured and possess continuous variation. The loci that govern the genetics of these traits are called quantitative trait loci (QTL).The continuous variation is due to the polygenic inheritance of genes with mostly small additive effects, and these genes are influenced by the environment. Mendelian methods of genetic analysis are not suitable to dissect these quantitative traits and hence different quantitative methods are used to study and understand them. Sax (1923) reported the linkage between seed coat color and seed size in common bean (*Phaseolus vulgaris*) that started the physical localization/mapping of the polygenes (Sax 1923). Development of the concept of the linkage along with the ability to construct the genomic map of the given species leads to the development of the QTL mapping. The first QTL mapping was done by Paterson et al. in 1988, using the restriction fragment length polymorphism in tomato (Paterson et al. 1988). Several reviews on QTL mapping in disease resistance in maize have been published (Wisser et al. 2006; Balint-Kurti and Johal 2009). The general methods of QTL mapping involve the utilization of a mapping population, usually a bi-parental population, derived from the cross between two genetically diverse parents, a dense marker linkage map for a particular species and genotypic data (SNPs, SSRs), standard phenotypic measurement and suitable software program (Singh and Singh 2015), such as R/QTL (Broman et al. 2003), QTL Cartographer (Wang et al. 2007), and so on.

The QTL mapping has been widely used in several crops, such as rice (Ray et al. 1996; Tan et al. 2001; Tian et al. 2015), maize (Lübberstedt et al. 1999; Balint-Kurti and Carson 2006; Li et al. 2008; Park et al. 2014), wheat (Quarrie et al. 1994; Castro et al. 2008; Acuna et al. 2014), tomato (Paterson et al. 1988; Foolad 1999; Causse et al. 2002; Causse et al. 2004), and others. QTL mapping provides the QTL effect size, the additive and dominance effect that are helpful for the trait introgression and improve breeding scheme. QTL mapping, also known as linkage mapping, possesses high power to detect the QTL and has the potential to identify or map the rare functional alleles of genes compared to the association mapping. With all these merits, QTL mapping also possesses multiple demerits, such as genetic variation is limited in the bi-parental mapping, as the mapping population is initiated with just two parents with limited recombination events. Low resolution power is another most challenging issue of QTL mapping. A QTL location may span from a few to tens of centimorgan; usually from 5 to 20 cM, encompassing several hundred genes, which will be time-consuming and difficult to analyze and further validate the identified QTLs (Doerge 2002). Hence, there are only a limited number of known QTLs that has been cloned or tagged at the gene level (Price 2006).

Association mapping, also known as linkage disequilibrium (LD), has emerged as a popular tool to dissect the complex traits at the sequence level. Initially, association mapping had been used extensively in medical genetics, but was limited in plant genetics due to the structured population often found in plants, which may lead to nonfunctional associations. In 2001, Thornsberry et al. introduced association mapping in plants (maize) by using statistical methods to account for the

variation due to population structure (Thornsberry et al. 2001). Since its introduction, association mapping has gained wide popularity in dissecting the complex traits in plants because of the advances in high-throughput genomic technologies, interests in identifying novel and superior allele, and improvements in statistical methods (Zhu et al. 2008).

Association mapping is different to QTL mapping in the following aspects: association mapping usually involves the use of unstructured or natural populations, consisting of diverse sets of individual or taxa. For instance, the Goodman association panel of maize consists of 282 diverse sets of maize inbreds, which collectively include tropical, subtropical, temperate, popcorn, and sweet corn lines, drawn from different environments and different locations (Flint-Garcia et al. 2005). The merit of using such natural population is that it exploits the linkage disequilibrium (LD) from the ancestral recombination present between or among them (Nordborg and Tavaré 2002), unlike linkage in the QTL mapping, which is only from the hybridization between the bi-parental lines. Association mapping utilizes the LD (nonrandom association of the alleles, which is a property of a population unlike linkage, which is a property of individual) between the SNPs and the associated genes or QTLs for detecting the marker-traits association. One of the important aspects of using the association mapping over QTL mapping is its high resolution that can detect the causative variants or causal genes.

Nested Association Mapping (NAM) population is another most widely used population both for the QTL mapping and association studies in Maize (McMullen et al. 2009). NAM is designed in such a way that it can harbor the advantage of both QTL and association mapping. The population was constructed to enable both high power and high resolution through the joint-linkage association analysis. Briefly, the population was designed by crossing a common parent, B73, with the other 24 diverse founder parents, and 200 RILs per family were created using the subsequent selfing for 5 generations of the resulting F1s. The diverse lines consist of mostly the tropical lines, a few temperate, sweet corn, and a popcorn inbred line. The NAM genetic map is a composite map created using 4699 RILs combined across the 25 families representing 1106 loci, with the average marker density of one marker on every 1.3 cM (McMullen et al. 2009).

6.2.1 GWAS Working Models

Similar to QTL mapping, association mapping also requires the phenotypic and genotypic data along with the genome map. However, due to use of unstructured population, one needs to be careful to consider those variations generated by the unstructured population while running the GWAS. The GWAS model needs to account for population structure in order to avoid getting spurious SNP hits in the analysis. This can be accounted using the Q matrix or the principal components (PCs) in the GWAS model, which can be obtained from the marker information. The next covariate used in the model is the relationship matrix or the kinship matrix

(K), which will account for the variation from the related alleles by identical by descent in the population. K matrix can be calculated either from the pedigree data or from the marker data. With the development of the genotyping platform and the statistical methods, the methodology of GWAS has improved drastically in the recent years.

Association mapping, simply, is a genome-wide scan of the tested molecular markers with the phenotype of interest. The association test idea was brought up on using the simple linear model, where the model is fitted using the SNPs as the predictor variable as fixed effect, and estimates the markers effect for the particular trait of interest using t-statistics. Since, the association mapping involves mostly the use of diverse natural or unstructured panel, the variation due to the diversity of the subpopulation within the population needs to be accounted in the model. The general linear model (GLM) takes into account the population structure in the form of Q matrix or principal components (PCs) and uses it as covariates in the model, which helps to control the spurious association (Price et al. 2006). The very popular and widely used mixed linear model (MLM) is the enhancement of the GLM, in the sense that it also accounts for the genetic relatedness, i.e., the K matrix fitted as random effect in the model. This Q + K matrix strongly helps to control the false positives (Yu et al. 2006). The development of tools and methods for doing association studies continues to grow using EMMA (Kang et al. 2008), CMLM (Zhang et al. 2010), and MLMM (Segura et al. 2012), which were developed especially to reduce the computational time for doing the GWAS. Recently, a new method was developed, FARMCPU, which is reported to completely remove the confounding between the testing markers and both K and Q, by combining MLMM and Fast-LMM-Select, allowing a fixed and random effect model to perform separately, and is also reported to be computationally efficient (Liu et al. 2016). FARMCPU claims to reduce the false positives without losing the true positives.

6.3   Disease Resistance Studies in Maize Using GWAS

6.3.1   *Fusarium Ear Rot*

Zila et al. (2013) revealed some of the important QTLs in the Fusarium ear rot resistance in maize (Zila et al. 2013), using the maize core diversity panel (Flint-Garcia et al. 2005). Fusarium ear rot, caused by *Fusarium verticillioides* (Sacc) Nirenberg, a common disease of maize affects both the quality of food and feed. The fungus is a hemibiotrophic fungus and endemic to the maize growing regions in the world. The fungus produces mycotoxin fumonisin, a suspected carcinogen associated with the various diseases in livestock and humans. It has been reported that a high genotypic correlation exists between the ear rot resistance and the fumonisin accumulation, indicating effective negative selection on fumonisin in the resistance cultivars. The disease has caused a huge loss of grains and the quality of grains. The best strategy to control the disease is to develop the disease resistance

maize cultivars. Fusarium ear rot resistance is mostly governed by polygenes and strongly influenced by the environment. Hence, there has been no report of the fully immune cultivars being discovered (Clements et al. 2004; Zila et al. 2013). Previous QTL studies have shown that the resistance QTLs for the Fusarium ear rot have small effect size and are not consistent between the populations (Pérez Brito et al. 2001; Ding et al. 2008).

Disease resistance itself is a quantitative trait, governed with multiples genes and affected by the environment. Hence, it is a great challenge to a plant scientist to incorporate the disease resistance genes without any growth or yield penalty. Briefly, the study was conducted using the 267 inbred lines evaluated in two sets of environment and the association mapping was done using 47,445 SNPs (Olukolu et al. 2013), using a mixed model. Three SNPs were found to be significantly associated with disease resistance in at least one subset of environment (Zila et al. 2013). Two of the three identified SNPs were found to be co-localized with the genes related with the programmed cell death. The chromosome 9 SNP explained the largest proportion of the variation in line mean values for ear rot resistance ($R^2 = 11.5$), whereas the SNPs in chromosome 1 and 5 explained 8.8 and 9.6% variation for the Fusarium ear rot resistance; collectively, 26% of the variation is explained by all the 3 SNPs.

Chromosome 9 gene was identified as GRMZM2G178880, which belongs to the cellulose synthase-like family A (*CslA*) protein family. Expression of this gene is found to be highest in the endosperm of the developing seed kernel between 20 and 24 days after flowering (Sekhon et al. 2011). Genes in the *CslA* protein family encode for the noncellulose polysaccharides, such as mannan polymers that form part of the wall matrix in plant cells (Dhugga 2005). Degradation of the mannan-rich cell walls might play an important role in the programmed cell death in the host-pathogen interaction (Rodríguez-Gacio et al. 2012) and may play a role in the disease resistance.

The SNP on chromosome 5 was located downstream of a Heat Shock Protein (HSP60) gene, GRMZM2G111477 (Zila et al. 2013). HSP60s are the chaperonins and are involved in the protein folding when the plants are in stressed condition. In Rice and Arabidopsis, the role of HSP60s is reported to be involved in the programmed cell death (Ishikawa et al. 2003). SNP on chromosome 1 is found within the coding region of the GRMZM2G703598, but has neither gene function predicted nor orthologs with other grass species (Zila et al. 2013).

A major limitation of association mapping in maize is its low linkage disequilibrium (LD) state which requires large number of genetic markers to detect marker-trait associations. Romay et al. (2013) reported that the use of approximately 680,000 GBS markers were sufficient to detect most of the known candidate genes associated with flowering time in maize (Romay et al. 2013). Polymorphism that strongly associated with the lower LD in tropical or subtropical population was more difficult to detect compared to polymorphism that more frequently associated with greater LD in temperate subpopulations. Hence, it indicates that although increased marker coverage and association panel size improves the power of the GWAS, consideration needs to be taken while doing GWAS with low LD subpopulation (tropical/subtropical population), in order to capture the rare allele variants

associated with those subpopulations (Romay et al. 2013). Rapid LD decay along the chromosomes in the maize core diversity panel suggests the use of larger SNP density and large association panels needed to identify the novel loci associated with the ear rot resistance.

6.3.2   Northern Leaf Blight

Poland et al. (2011) studied northern leaf blight in maize (Poland et al. 2011) using GWAS. They evaluated 5000 inbred lines from the nested association mapping lines for the resistance to the northern leaf blight and identified 29 QTLs, and most of them possessed multiple alleles.

Quantitative disease resistance (QDR) is reported to be associated with durable resistance as pathogen that overcomes a single allele of small effect does not gain a large selective advantage, and loss of the allele with small effect does not leave the host completely susceptible (Poland et al. 2009). Northern leaf blight (NLB) is an endemic disease in the maize growing areas of the world causing moderate to severe yield losses (Perkins and Pedersen 1987). NLB is caused by a fungal pathogen *Setosphaeriaturcica* (anamorph *Exserohilumturcicum*), a hemibiotrophic fungus, and is commonly found in the tropical highlands. Previous studies on the NLB have shed light on several QTLs. Among these, three genes confer incomplete race-specific resistance. *Ht1* (Bentolila et al. 1991) located in maize bin 2.08 and *Ht2* (Yin et al. 2003) and *Htn1* (Simcox and Bennetzen 1993) located in the maize bin 8.06. However, as stated earlier, due to the low resolution of the QTL mapping, the positional cloning of these genes was difficult and not widely used in the breeding programs. Hence, this study combined the positive aspect of both the association mapping and the linkage study to unravel the genetic architecture of the NLB.

Briefly, a large NAM population created with 5000 recombinant inbred lines (RILs) was used for the dissection of the complex traits (Yu et al. 2008). Apart from the 25 NAM RILS families, RILs from the intermated B73 X MO17 (IBM) population was included as a 26th family in the study. The NAM RILs were genotyped with 1106 SNP markers and the data are also publicly available in *www.panzea.org*. The study was conducted over three seasons in nurseries artificially inoculated with the single isolate of *S. turcica* race 1. The NAM parent showed the extensive variation on the resistance of the NLB, where the common parent B73 showed 34% of the diseased leaf area being moderately susceptible. The study reported that the GXE interaction of the NLB resistance was minimal; however, the study was carried out only in one location. The author also mentioned that there exists a strong negative correlation between the flowering time (days to anthesis (DTA)) and the NLB resistance in the founder lines (Poland et al. 2011). Joint linkage study was done using stepwise model selection and DTA as a covariate in the model, resulting in the 29 QTLs accounting for 77% of the total variance. Most of the QTLs have a small effect and only few have a large effect. Large effect QTL was identified on chromosome 8 at 152.2 MB segregating in multiple families which were likely to

be *Ht2*, the position was consistent with the physical location identified by fine mapping. The broad sense heritability for the NLB index for the NAM founders was 0.74 (Poland et al. 2011). Most tropical maize lines possess higher level of NLB and other disease resistance than the temperate lines reflecting the favorable conditions for disease development in tropics and thus, useful for a breeder to select for resistance in these environments (Poland et al. 2011). GWAS was done using 1.6 million SNPs, identified 208 significant SNPs association, and 28 of 29 QTLs had one or more SNP associations.

A small subset of the 208 SNP loci found to be associated with the resistance to NLB is shown in Table 6.1, which is adapted from the Poland et al. (2011). The study showed five SNPs associated with the receptor-like kinase (RLK) genes and one additional association with the sixth LRR-related gene. LRR domains have been reported to be associated with the plant disease resistance. Several SNPs detected showed candidate genes with antifreeze domains which has high similarity to the pathogenesis related proteins and were reported to enhance disease resistance. Several serine/threonine protein kinases were identified, and they are also involved in the plant defense responses.

6.3.3 Southern Leaf Blight

Kump et al. (2011) conducted GWAS study on southern leaf blight (SLB) of maize, using nested association mapping population (Kump et al. 2011). SLB is caused by the fungus *Cochliobolus heterostrophus*, a necrotrophic fungus, which tends to occur usually after anthesis. There are limited QTLs and markers identified as the disease resistance QTLs for SLB. They performed joint linkage analysis and identified 32 QTLs, with mostly small additive effects on the SLB resistance. Most of the SNPs detected were previously reported to be near or within the sequence homologues to the genes previously identified in the disease resistance.

Maize NAM represents 135,000 recombination events and hence, good for association mapping as well as linkage study. The study was done across the three environments. GWAS was run using 1.6 M HapMap SNPs that were identified among the founder lines and imputed on the complete NAM panel for the study (Kump et al. 2011). In the study, they used the SLB index values as a phenotypic measurement, which represent the mean of SLB resistance measured across time points and environments. Measurement was done using a standard nine-point rating scale. The B73, common NAM parent, was the most susceptible among all parents. Heritability of the SLB index score was found high, around 87%. The identified SNPs and QTLs and their position in the chromosome were shown in the study (Kump et al. 2011). The 32 QTLs jointly explained 80% of the phenotypic variation of the SLB resistance, as well as 93% of the genotypic variation of the SLB resistance. Additive epistatic interaction between the QTLs was not detected. The QTL with the largest effect estimate was mapped to the bin 3.04, which is known as the previously identified region for the SLB resistance (Balint-Kurti et al. 2007). With the above studies,

Table 6.1 A summary of important SNPs and QTLs of the five major diseases of maize

Disease	Chromosome	Physical position bp	AGP Version	Candidate gene/ (QTLs)	Annotation	Reference
Fusarium Ear Rot	9	151,295,233	AGPv2	GRMZM2G178880	Cellulose synthase-like family A (CslA) protein family	Zila et al. (2013)
	5	30,997,717	AGPv2	GRMZM2G111477	Heat Shock Protein	
	1	63,540,590	AGPv2	GRMZM2G703598		
Northern Leaf Blight	1	12,136,678	AGPv1		Serine/threonine protein kinase	Poland et al. (2011)
	2	9,394,756	AGPv1		RLK	
	2	160,834,095	AGPv1		RLK	
	3	3,382,179	AGPv1		Antifreeze	
	5	203,735,206	AGPv1		RLK	
	1	88,927,678	AGPv1		Serine/threonine protein kinase	
	7	125,153,323	AGPv1		Peptidase/serine/threonine protein kinase	
Southern Leaf Blight	1	45,565,372	AGPv1		Mitochondrial carrier protein (programmed cell death); Ran GTPase (plant defense response)	Kump et al. (2011)
	1	80,360,348	AGPv1		Glutathione S-transferase (plant defense)	
	1	210,676,683	AGPv1		LRR receptor kinase (disease resistance25)	
	2	10,687,858	AGPv1		AP2 transcription factor (disease resistance)	
	3	22,604,327			NRR (defense response)	
	4	240,050,394			Pti4, Pti5, and Pti6 ERF transcription factors; ABC transporter (disease resistance)	
	10	1,221,166			NPR1 (disease resistance)	
Head Smut	1	278,884,507	AGPv1	GRMZM2G300990	Serine/threonine protein kinase	Wang et al. (2012)
	2	201,359,446	AGPv1	GRMZM2G140231	Serine/threonine protein kinase	
	2	219,834,173	AGPv1	GRMZM2G166566	Basic leucine zipper transcription factor	

	3	124,139,795	AGPv1	GRMZM2G137289	MADS-box	
	5	188,373,740	AGPv1	GRMZM2G312274	Auxin	
	9	84,924,940	AGPv1	GRMZM2G443953	WD40 repeat	
	10	137,723,733	AGPv1	GRMZM2G117667	Lipase	
Gray Leaf Spot	1	234,309,200	AGPv2	GLS 1.2 (QTLGLSchr1)		Mammadov et al. (2015)
	6	107,475,955	AGPv2	GLS 6.1 (QTLGLSchr6)		
	7	19,500,572	AGPv2	GLS 7.2 (QTLGLSchr7)		
	8	19,550,800	AGPv2	GLS 8.2 (QTLGLSchr8a)		
	8	79,142,282	AGPv2	GLS 8.3 (QTLGLSchr8b)		

it should be noted that plant scientists working in GWAS of diseases that correlated with the flowering time or days to anthesis (DTA) should account for DTA variation as covariate in their GWAS model. In the study, the author found the 30 flowering QTLs, explaining around 85% of the phenotypic variation for the DTA, where 8 pairs of the QTLs for SLB and DTA were found to have overlapping support intervals. The RILs subfamily B73 X CML247 possesses the highest of 15 QTLs with significant allelic effect, whereas the B73 X CML52 possesses 2 QTLs.

A list of the important candidate gene annotation for SLB resistance is shown in Table 6.1, adapted from Kump et al. (2011). Two genes with leucine-rich repeat (LRR) domains were found. Another important gene found was a gene with strong similarity to *NPR1*, which is related to defense response. An SNP was found adjacent to a homolog of the rice gene NRR (negative regulator of the resistance) that encodes a protein which interacts with the NPR1 protein during the defense response (Chern et al. 2005).

6.3.4   *Head Smut*

A GWAS study on the head smut conducted by Wang et al. (2012) identified 18 novel candidate genes, which were further categorized into resistance genes, disease response genes, and other disease resistance function genes. The author used 45,658 SNPs with an association panel of 144 inbred lines and ran the GWAS in mixed linear model (Q + K) in Tassel V2.1.

Head smut is caused by fungus *Sphacelothecareiliana* (Kühn) Clint, a global maize disease causing moderate to severe loss of both quality and quantity. The study was conducted in different environments with different replications. Artificial inoculation was performed using the previously collected teliospores of *S. reiliana* at a ratio of 1000:1 (teliospores: seed). The percentages of the completely infected plants per plots were scored in either tassels or ears at the mature plant stage. Extensive variation of the susceptibility to head smut was found among the lines, which range from 0% to 83% susceptibility range and the broad sense heritability was high at 88.7% (Wang et al. 2012). The genotype by environment interaction was also significant.

Several QTL mapping studies have been done and have found several QTLs for the head smut (Lu and Brewbaker 1999; Lübberstedt et al. 1999; Li et al. 2008). However, the QTLs were not very consistent among the studies. The QTLs found were reported mainly on chromosome 1, 2, 3, 8, 9, and 10. The major QTL for the head smut resistance, *qHSR1*, has been fine mapped in bin 2.09 using the 68 BC2 recombinants from the cross of Ji1037 and Huangza04 (Chen et al. 2008).

The mixed linear model outputs 19 significant SNPs, which collectively explained 86.5% of the total phenotypic variation ranging from 3.5% to 9.2%. Defense-related gene families such as serine/threonine protein kinases, leucine-rich repeat protein, MADS-box protein (bin 3.05), Auxin (bin 5.05), and WD40 repeat containing protein (bin 9.03) were identified. Two nucleotide-binding sites (NBS) encoding protein were detected on chromosome 8. Bin 2.09 was previously identified as a head

smut resistance region. The SNP on chromosome 2, with physical position 219834173 on AGPv1, was found to be overlapped in the region and was later molecularly validated as resistance gene. This gene GRMZM2G166566 is annotated as a basic leucine zipper transcription factor and has R^2 value of 9.3%.

6.3.5   Gray Leaf Spot

Gray Leaf Spot (GLS) (causal agents *Cercosporazeae-maydis* and *Cercosporazeina*) is one of the most important foliar diseases of maize. In the United States, *C. zeae-maydis* occurs everywhere where corn is being cultivated, whereas *C. zeina* is mainly found on the East coast (Wang et al. 1998). Disease is prevalent in the areas where dewy mornings are followed by a hot humid afternoon and relatively cool nights. Breeding the disease resistance cultivars is the most prominent strategy to control the disease. Mammadov et al. (2015) combined the high QTL detection power of genetic linkage mapping with high resolution power of GWAS to study the resistance of GLS, which is beneficial for the marker-assisted QTLs introgression (Mammadov et al. 2015).

The association study used the 300 maize association panel and was replicated in four different environments. The Association Panel comprised 215 DAS proprietary lines of North and South American origin, 27 ex-PVP lines, 37 CYMMIT lines, and 21 lines from the National Plant Germplasm system. All lines in the association panel were chosen based on their previously known reaction to GLS, which were categorized into four major categories: GLS susceptible, moderately GLS susceptible, moderately GLS resistant, and GLS resistant. For any bi-allelic SNP [A/B], a positive effect suggests that the allele contributing to GLS severity comes from allele A and a negative effect suggests that the allele contributing to GLS severity comes from allele B. GWAS was conducted using ~25,000 SNP markers with minor allele frequencies (MAF) > 0.1.

Besides, they used 72 lines for the biparental QTL mapping developed from the cross between DAS-001 (GLS resistant) and DAS-002 (GLS susceptible). Both DAS are the proprietary maize inbred lines of Dow AgroSciences. The map of 1985 SNP markers was evenly distributed across ten maize chromosomes. Extended composite interval mapping (ECIM) model was used for QTL mapping.

They identified four GLS resistance QTL on the chromosome 1, 6, 7, and 8, which was further validated by GWAS. The genetic linkage – GWAS hybrid mapping system in the study identified one novel GLS resistance QTL (*QTLGLSchr8a)* and confirmed four previously mapped QTL (*QTLGLSchr1*, *QTLGLSchr6*, *QTLGLSchr7*, and *QTLGLSchr8b*) with more refined position. Three minor and one major QTL were detected on chromosomes 1, 6, 7, and 8, respectively. The QTL on chromosome 8 (*QTLGLSchr8*) explained about 26.5% of the variation, while the QTL on chromosomes 1 (*QTLGLSchr1*), chromosome 6 (*QTLGLSchr6*), and chromosome 7 (*QTLGLSchr7*) were responsible for 4.55%, 6.85%, and 5.23% of GLS resistance, respectively. In total, all four identified QTL explained 43.13% of GLS resistance in the DAS-001 inbred line (Mammadov et al. 2015).

6.4   Future Perspective

Genome-wide association studies in crops have tremendously benefitted the farmers, cooperatives, agriculture companies, and the scientific community. It has already been proven that the GWAS studies are beneficial in finding the causal variants of the disease and can be effectively used in developing disease resistance cultivars. However, one needs to be careful in using the GWAS model, as it has a high chance to provide false positive SNPs, given the incorrect model, and we already know that it is affected by several factors such as population structure, kinship, and selection history, hence, it is always useful to have positive control traits/SNPs in running the GWAS model or one can do simulation of the SNPs in the absence of the positive control to avoid those spurious hits. With the decrease in the cost of genotyping, GWAS using high density markers, high population sample size replicated in different environments and years will provide high power to detect the causal variants.

Quantitative traits are governed by polygenes of mostly small effects. Interaction of genes and their associative role in the phenotype is proven to be highly important. However, detecting those epistatic QTLs are still a major challenge for plant scientists. Now it is time to contemplate on designing the mapping population that can dissect the epistatic variation over the additive variation in studying quantitative traits. The other major limitation of GWAS is that it is not capable of detecting the rare allelic variants. The power of detection of marker-trait association depends on allele frequency of the particular QTL. Rare/low frequency alleles having either small or large effect are not detected by GWAS. The functional role of those rare variants has started to shed light in human as well as in plant disease. Hence, the next few decades will be important to understand the functional role of the rare variants/alleles in disease resistance.

The other limitation of GWAS includes the missing heritability concept, where the high heritable traits on the phenotypic variation remain unexplained. One of the reasons is that we tend to ignore the effect of thousands of SNPs under the threshold, which might possess good biological information. Hence, the concept of genomic prediction and selection has evolved in the recent years in the field of plant disease resistance. Genomic prediction utilizes the genomic breeding values of the genotypes obtained from genotypic and phenotypic information from the training set population and used that to predict the phenotype of the breeding set. This is useful, as the genotypic cost is decreasing dramatically, whereas the phenotypic cost is still high. Genomic prediction has already begun in a few diseases in maize (Technow et al. 2013; Gowda et al. 2015) and shown to have good prediction, which helps to reduce the cycle of selection and ease the breeding effort for developing disease-resistant cultivars. Exploring and mitigating the disease resistance challenge using multi-omics integration and system genetics approach is another interesting modern day concept. With all these fascinating developments in tools and concepts, the breeding for the disease-resistant cultivars in the coming decades will be another revolution in mitigating the poverty and malnutrition and for the sustainable agriculture across the globe.

References

Acuna TB, Rebetzke G, He X, Maynol E, Wade L (2014) Mapping quantitative trait loci associated with root penetration ability of wheat in contrasting environments. Mol Breed 34(2):631–642

Balint-Kurti P, Carson M (2006) Analysis of quantitative trait loci for resistance to southern leaf blight in juvenile maize. Phytopathology 96(3):221–225

Balint-Kurti PJ, Johal GS (2009) Maize disease resistance. In: Handbook of maize: its biology. Springer, New York, pp 229–250

Balint-Kurti P, Zwonitzer JC, Wisser RJ, Carson M, Oropeza-Rosas MA, Holland JB, Szalma SJ (2007) Precise mapping of quantitative trait loci for resistance to southern leaf blight, caused by Cochliobolus heterostrophus race O, and flowering time using advanced intercross maize lines. Genetics 176(1):645–657

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ (1994) RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant disease resistance genes. Science 265(5180):1856–1860

Bentolila S, Guitton C, Bouvet N, Sailland A, Nykaza S, Freyssinet G (1991) Identification of an RFLP marker tightly linked to theHt1 gene in maize. Theor Appl Genet 82(4):393–398

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19(7):889–890

Castro A, Tacaliti M, Giménez D, Tocho E, Dobrovolskaya O, Vasicek A, Collado M, Snape J, Börner A (2008) Mapping quantitative trait loci for growth responses to exogenously applied stress induced hormones in wheat. Euphytica 164(3):719

Causse M, Saliba-Colombani V, Lecomte L, Duffe P, Rousselle P, Buret M (2002) QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. J Exp Bot 53(377):2089–2098

Causse M, Duffe P, Gomez M, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot 55(403):1671–1685

Chen Y, Chao Q, Tan G, Zhao J, Zhang M, Ji Q, Xu M (2008) Identification and fine-mapping of a major QTL conferring resistance against head smut in maize. Theor Appl Genet 117(8):1241

Chern M, Canlas PE, Fitzgerald HA, Ronald PC (2005) Rice NRR, a negative regulator of disease resistance, interacts with Arabidopsis NPR1 and rice NH1. Plant J 43(5):623–635

Clements M, Maragos C, Pataky J, White D (2004) Sources of resistance to fumonisin accumulation in grain and Fusarium ear and kernel rot of corn. Phytopathology 94(3):251–260

Dhugga KS (2005) Plant Golgi cell wall synthesis: from genes to enzyme activities. Proc Natl Acad Sci U S A 102(6):1815–1816

Ding J-Q, Wang X-M, Chander S, Yan J-B, Li J-S (2008) QTL mapping of resistance to Fusarium ear rot using a RIL population in maize. Mol Breed 22(3):395–403

Doerge RW (2002) Multifactorial genetics: mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3(1):43

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44(6):1054–1064

Flor HH (1971) Current status of the gene-for-gene concept. Annu Rev Phytopathol 9(1):275–296

Foolad M (1999) Comparison of salt tolerance during seed germination and vegetative growth in tomato by QTL mapping. Genome 42(4):727–734

Gowda M, Das B, Makumbi D, Babu R, Semagn K, Mahuku G, Olsen MS, Bright JM, Beyene Y, Prasanna BM (2015) Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theor Appl Genet 128(10):1957–1968

Hammond-Kosack KE, Jones JD (1997) Plant disease resistance genes. Annu Rev Plant Biol 48(1):575–607

Ishikawa A, Tanaka H, Nakai M, Asahi T (2003) Deletion of a chaperonin 60β gene leads to cell death in the Arabidopsis lesion initiation 1 mutant. Plant Cell Physiol 44(3):255–261

Johal GS, Briggs SP (1992) Reductase activity encoded by the HM1 disease resistance gene in maize. Science 258(5084):985–987

Jones JD, Dangl JL (2006) The plant immune system. Nature 444(7117):323

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178(3):1709–1723

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43(2):163

Levings CS, Siedow JN (1992) Molecular basis of disease susceptibility in the Texas cytoplasm of maize. In: 10 years plant molecular biology. Springer, Dordrecht, pp 135–147

Li X, Wang Z, Gao S, Shi H, Zhang S, George M, Li M, Xie C (2008) Analysis of QTL for resistance to head smut (Sporisorium reiliana) in maize. Field Crop Res 106(2):148–155

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12(2):e1005767

Lu X, Brewbaker J (1999) Molecular mapping of QTLs conferring resistance to Sphacelotheca reiliana (Kühn) Clint. Maize Genetics Cooperation Newsletter (73)

Lübberstedt T, Xia X, Tan G, Liu X, Melchinger A (1999) QTL mapping of resistance to Sporisorium reiliana in maize. Theor Appl Genet 99(3–4):593–598

Mammadov J, Sun X, Gao Y, Ochsenfeld C, Bakker E, Ren R, Flora J, Wang X, Kumpatla S, Meyer D (2015) Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (Zea mays L.). BMC Genomics 16(1):916

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262(5138):1432–1436

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C (2009) Genetic properties of the maize nested association mapping population. Science 325(5941):737–740

Nimchuk Z, Eulgem T, Holt Iii BF, Dangl JL (2003) Recognition and response in the plant immune system. Annu Rev Genet 37(1):579–609

Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18(2):83–90

Olukolu BA, Negeri A, Dhawan R, Venkata BP, Sharma P, Garg A, Gachomo E, Marla S, Chu K, Hasan A (2013) A connected set of genes associated with programmed cell death implicated in controlling the hypersensitive response in maize. Genetics 193(2):609–620

Park KJ, Sa KJ, Kim BW, Koh H-J, Lee JK (2014) Genetic mapping and QTL analysis for yield and agronomic traits with an F2: 3 population derived from a waxy corn× sweet corn cross. Genes Genomics 36(2):179–189

Parlevliet JE (2002) Durability of resistance against fungal, bacterial and viral pathogens; present situation. Euphytica 124(2):147–156

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335(6192):721

Pérez Brito D, Jeffers D, González de León D, Khairallah M, Cortés C, Velázquez C, Azpíroz S, Srinivasan G (2001) QTL mapping of Fusarium moniliforme ear rot resistance in highland maize. Agrociencia, Mexico, 35(2)

Perkins J, Pedersen W (1987) Disease development and yield losses associated with northern leaf blight on corn. Plant Dis 71(10):940–943

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci 14(1):21–29

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci 108(17):6893–6898

Pratt RC, Gordon SG (2006) Breeding for resistance to maize foliar pathogens. Plant Breed Rev 27:119

Price AH (2006) Believe it or not, QTLs are accurate! Trends Plant Sci 11(5):213–216

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38(8):904

Pring DR, Lonsdale DM (1989) Cytoplasmic male sterility and maternal inheritance of disease susceptibility in maize. Annu Rev Phytopathol 27(1):483–502

Quarrie S, Gulli M, Calestani C, Steed A, Marmiroli N (1994) Location of a gene regulating drought-induced abscisic acid production on the long arm of chromosome 5A of wheat. Theor Appl Genet 89(6):794–800

Ray J, Yu L, McCouch S, Champoux M, Wang G, Nguyen H (1996) Mapping quantitative trait loci associated with root penetration ability in rice (Oryza sativa L.). Theor Appl Genet 92(6):627–636

Rodríguez-Gacio MDC, Iglesias-Fernández R, Carbonero P, Matilla ÁJ (2012) Softening-up mannan-rich cell walls. J Exp Bot 63(11):3976–3988

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14(6):R55

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics 8(6):552–560

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44(7):825

Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, Kaeppler SM (2011) Genome-wide atlas of transcription during maize development. Plant J 66(4):553–563

Simcox KD, Bennetzen JL (1993) The use of molecular markers to study Setosphaeria turcica resistance in maize. Phytopathology 83(12):1326–1330

Singh B, Singh AK (2015) Marker-assisted plant breeding: principles and practices. Springer

Song W-Y, Wang G-L, Chen L-L, Kim H-S, Pi L-Y, Holsten T, Gardner J, Wang B, Zhai W-X, Zhu L-H (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270(5243):1804–1806

Tan Y, Sun M, Xing Y, Hua J, Sun X, Zhang Q, Corke H (2001) Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. Theor Appl Genet 103(6–7):1037–1045

Technow F, Bürger A, Melchinger AE (2013) Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. G3 3(2):197–203

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES IV (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28(3):286

Tian Y, Zhang H, Xu P, Chen X, Liao Y, Han B, Chen X, Fu X, Wu X (2015) Genetic mapping of a QTL controlling leaf width and grain number in rice. Euphytica 202(1):1–11

Ullstrup A (1972) The impacts of the southern corn leaf blight epidemics of 1970-1971. Annu Rev Phytopathol 10(1):37–50

Wang J, Levy M, Dunkle LD (1998) Sibling species of Cercospora associated with gray leaf spot of maize. Phytopathology 88(12):1269–1275

Wang S, Basten C, Zeng Z (2007) Windows QTL cartographer 2.5. Department of statistics. North Carolina state university, Raleigh

Wang M, Yan J, Zhao J, Song W, Zhang X, Xiao Y, Zheng Y (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Sci 196:125–131

Wisser RJ, Balint-Kurti PJ, Nelson RJ (2006) The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathology 96(2):120–129

Xiao W, Zhao J, Fan S, Li L, Dai J, Xu M (2007) Mapping of genome-wide resistance gene analogs (RGAs) in maize (Zea mays L.). Theor Appl Genet 115(4):501–508

Yin X, Wang Q, Yang J, Jin D, Wang F, Wang B, Zhang J (2003) Fine mapping of the Ht2 (Helminthosporium turcicum resistance 2) gene in maize. Chin Sci Bull 48(2):165–169

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38(2):203

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178(1):539–551

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42(4):355

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1(1):5–20

Zila CT, Samayoa LF, Santiago R, Butrón A, Holland JB (2013) A genome-wide association study reveals genes associated with Fusarium ear rot resistance in a maize core diversity panel. G3 3(11):2095–2104