Recent Developments in Genomic Selection for Minor Gene Quantitative Disease Resistance Plant Breeding

Dagnachew Bekele1,2*, Kassahun Tesfaye1, Asnake Fikre4
1Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia; 2Ethiopian Institute of Agricultural Research, Addis Ababa, Ethiopia; 3Ethiopian Biotechnology Institute, Addis Ababa, Ethiopia; 4International Crop Research Institute for Semi-Arid Tropics, Addis Ababa, Ethiopia

ABSTRACT
To speed up the development of improved crop varieties, genomics assisted plant breeding is becoming an important tool. With traditional breeding and marker assisted selection, there have been several achievements in breeding for diseases resistance. Most research for disease resistance has been focused on major disease resistance genes which are highly effective although very vulnerable to breakdown with rapid changes in pathogenic races. In contrast, breeding for minor gene quantitative resistance can produce more durable plant varieties although it is very slow and challenging breeding. As the genetic architecture of the plant disease resistance shifts from single major R genes to many minor quantitative genes, the most appropriate approach for molecular plant breeding is genomic selection (GS) than marker assisted selection or conventional breeding. With the advent of new genomic tools, GS has emerged as one of the most important approaches for predicting genotype performance to improve genetically complex quantitative traits. Consequently, GS helps to accelerate the rate of genetic gain in breeding by using whole genome sequence data to predict the breeding value of offspring. GS breeding for quantitative resistance will therefore necessitate whole genome prediction models and selection methodology as implemented for classical complex traits. With the implementation of GS for yield and other economically important traits, whole genome marker profiles are available for the entire set of breeding lines, enabling genomic selection for disease resistance with no additional direct cost. Therefore, recent developments in GS including a two stream GS + de novo GWAS models (GS+) and GS for combined highest level of quantitative resistance with R genes (QR +R gene) individuals are expected to further advance disease resistance plant breeding and briefly discussed.

Keywords: Breeding; Genomics selection; Diseases; Resistance; Quantitative resistance

INTRODUCTION
Conventional plant breeding strategies for selecting best and resistant plant genotypes for release depending on the phenotypic trait have proven to be of limited success. To develop improved varieties, any breeding program must have an efficient strategy for evaluating candidate genotypes for high yield, disease resistance, good agronomic performance, and better end use qualities [1]. Plant breeding for disease resistance varies depending on whether the resistance is considered to be qualitative or quantitative. The genetic architecture of disease resistance is closely tied to whether the resistance is quantitative or qualitative, and hence both the phenotypic and molecular breeding approaches must be matched accordingly. Qualitative disease resistance conditioned by single, major genes that does not have a complex genetic architecture. It is more suitable for identifying and mapping single resistance genes of large effect [2]. In contrast, quantitative disease resistance can be approached using whole genome prediction models developed for quantitative traits [2]. Genomic selection (GS) is an ideal approach for highly polygenic complex traits with lower heritability and a complex genetic architecture that are controlled by thousands of genes each with very small individual effects [3].

Recent advances in sequencing technologies helped to apply effective genotyping methods such as genotyping by sequencing (GBS) in different crop species with large number of markers for GS [1,2,4,6]. GS is highly advanced and accurate technique in estimating breeding values for complex quantitative traits compared to the traditional phenotypic selection or marker assisted selection (MAS). GS has more power to capture small effect loci that would be missed by MAS [7]. Besides, the use of dense genome wide markers increases the chance of markers being...
in linkage disequilibrium (LD) with QTL influencing the trait of interest [8], and determines to some extent how well genetic relationship and genetic architecture are captured by the genomic selection model [9].

In GS, breeding program representative individual are genotyped with dense, genome wide markers, such as SNPs and phenotyped for traits of interest. This set of individuals are called a training population which will be used to create a genomic prediction model [2,10-12]. This model estimates the sum of the additive genetic effects of the genome wide alleles on the trait of individuals and known as genomic estimated breeding values (GEBVs). In selection cycles, non phenotyped breeding materials are genotyped, with the same set of markers as training population. Then, appropriate prediction model is used to predict their GEBVs for the trait of interest. One of the most important features of GS is all selected markers need to be significant and used for prediction modeling [5-8].

**METHODS**

**Breeding for disease resistance**

Breeding methods for disease resistance vary depending on whether the resistance is considered to be qualitative or quantitative. Qualitative plant disease resistance is resistance controlled by a single resistance (R) gene recognizing a virulence factors in a classic gene for gene mechanism. With molecular markers for R genes, direct selection for disease resistance can be implemented in the breeding programs. There has been significant effort on the identification of markers linked with major genes and mapping quantitative trait loci (QTLs) for disease resistance [2]. Several hundreds of R genes have been mapped across important crop plants, including rice [13,14], wheat [15-17], maize [18-20], soybean [21,22] and potato [23,24] as well as numerous other crop pathogens [25,26].

Nevertheless, in applied plant breeding there are relatively very few examples of large scale implementation of MAS for disease resistance [2]. As reported by Mielaner & Korzun [2] in wheat and barley breeding, the lack of markers applied in commercial breeding for disease resistance could be due to having few diagnostic markers. Besides, few monogenic resistances are durable, and only a few QTLs with high effects have been successfully transferred into elite breeding material [27].

Currently marker assisted selection has failed significantly to improve polygenic trait [28,29]. Quantitative resistance is considered to be more durable than qualitative disease resistance. Unlike resistance based on R genes, quantitative resistance generally does not appear to be race specific [30,31]. With race specific resistance, the prevalent pathogen race is a component of the environment, leading to greater observed G × E interaction. By comparison, minor gene resistance that has no or minimal gene for gene interaction leads to much less G × E interaction. In breeding, yield stability is the ideal target, especially because good performance across years is desired. Stability in resistance, resistance with minimal G × E, is important for achieving yield stability, particularly in areas prone to epidemics.

Breeding for race nonspecific minor gene resistance is one way to minimize G × E of resistance. In addition to minimizing G × E, there are also quantitative genetic and genomic prediction models that can help improve breeding efficiency when G × E is present, as long as there is some genetic correlation between environments [29,32]. Therefore, as the genetic architecture of resistance shifts from single major R genes to a diffused architecture of many minor genes, the best approach for molecular plant breeding is shifting from marker assisted selection to genomic selection. With the implementation of GS for yield and other economically important traits, whole genome marker profiles are available for the entire set of breeding lines, enabling genomic selection for disease resistance with no additional direct cost.

**Major steps and advantages of GS**

Genomic selection showed great promise to strongly increase the rate of genetic improvement in plant breeding programs [8]. It allows a comparative larger gain from selection by estimating all marker effects simultaneously and subsequent selection of genetically superior individuals based on their genomic estimated breeding value (GEBV) [33], instead of using a few significant markers as in classical marker assisted selection.

As illustrated in Figure 1, GS uses a training population of individuals that have been both genotyped and phenotyped to develop a model that takes genotypic data from a candidate population of untested breeding materials and produces genomic estimated breeding values (GEBVs). These GEBVs say nothing of the function of the underlying genes but they are the ideal selection criterion. In the plant breeding context, untested breeding materials would belong to a broader population defined as a crop market class or the breeding program as a whole [1,11].

Besides, very recently a two stream GS breeding scheme was developed [34] in which unutilized germplasm is systematically incorporated into a GS breeding pipeline again to test and predict the presence of new, highly effective allele combinations (Figure 2). In stream 1: Several pre-breeding materials with many favorable alleles from exotic germplasm are sequentially introduced into adapted germplasm. After many cycle of backcrossing and recombination to break linkages after the initial F1 cross between un-adapted and adapted material, selection of individuals from breeding population 1 is performed using a combination of GS + de novo GWAS models (GS+), in which the exotic QTL are fit as fixed effects and phenotype. The training population GS would be a subset of breeding population 1, that is, a fraction of breeding population 1 would be both genotyped and phenotyped, while the rest of breeding population 1 would be genotyped only.

Stream 2 continues the process of further refining and improving existing elite materials. Adapted materials from breeding population 1 are crossed into breeding population 2 where they are further refined using GS + de novo GWAS models, where the fixed effects would include valuable QTL identified based on GWAS performed in Breeding Population 2, the exotic QTL from Stream 1, or any other large effect QTL a breeder might normally target for trait improvement. Output from Stream 2 can be advanced toward variety release or fed back into stream 1 to serve as parents for further crossing and population development. This approach helps the breeders to learn directly from data on new and diverse germplasm and make rapid genetic gain.

The main factors that affect the accuracy of GS include the heritability of the trait, the rate of linkage disequilibrium decay, the marker density, and the number of individuals in the training population [9]. When LD decays more rapidly, then a greater number of markers and individuals for model training are needed.
When predicting across populations, as in the case where previous breeding candidates are used for model training to predict new selection candidates, the relationship between the model training population and the selection candidates is important [35].

GS is a promising method for exploiting molecular genetic markers to design novel breeding programs and to develop new markers based models for genetic evaluation. In plant breeding, it provides opportunities to increase genetic gain of complex traits per unit time and cost. Therefore, for complex quantitative traits, GS provides higher selection accuracy in reduced time giving an accurate and good genetic gain per unit time as expected as compared with conventional phenotypic selection and MAS for complex traits. Besides, the fundamental difference between MAS and GS impacting the effectiveness of these two selection tools is scale [36]. MAS is limited in its ability to predict breeding values as it concentrates on a small number of QTLs that are tagged by markers with well-defined associations. In contrast, GS uses a dense set of markers from across the entire genome, assuring that all QTL are in LD with at least one SNP marker (Figure 3).

**Statistical models for GS**

There are a variety of statistical models used to estimate breeding

![Figure 1: Steps for predicting genotypic value in GS.](image)

![Figure 2: New a two stream GS breeding scheme [34].](image)
values in GS. Based on the crop species, trait considered, and breeding population design, the choice of statistical models used in GS has been shown to have a significant effect on prediction accuracy [37,38]. Very interestingly, for breeding crops like rice and wheat with large effect QTL, statistical models that incorporate a select number of molecular markers as fixed effects have been shown to contribute to improved prediction accuracy [34]. Two of the most commonly used models for purely quantitative traits are Genomic best linear unbiased prediction (G-BLUP) and ridge regression BLUP (RR-BLUP) [2]. G-BLUP is a mixed linear

Figure 3: Next-generation sequencing (NGS) based marker technologies and high throughput phenotyping (HTP) accelerates GS based plant breeding [37].

Figure 4: A GS scheme for combining R gene resistance (+R) and high level of Quantitative Resistance (QR) [2].
model, with individuals as random effects, and the covariance among individuals is assumed to be proportional to the genomic relationship matrix estimated with genome-wide markers.

G-BLUP is a modification of the conventional BLUP model [39], which uses pedigree relationships rather than genomic relationships. When using G-BLUP for prediction, both the model-training and validation individuals are included in the relationship matrix, but only the model training individuals have phenotypic data. RR-BLUP is also a mixed linear model, but markers are considered random effects [40]. Covariance between markers is considered to be zero, and the marker variance is assumed to be the total genetic variance divided by the number of markers. This assumes that variance is equal for all markers, which enables many more marker effects to be estimated than there are phenotypic records [2]. Other suitable models for traits that fall between quantitative and qualitative inheritance are the Bayesian models Bayes-A, Bayes-B [8], Bayes-Cπ [41], and Bayesian LASSO [42]. With Bayes-A, each marker is assumed to have a unique variance. Bayes B is an extension of Bayes-A and allows some markers to have no effect.

Successes of GS in disease resistance breeding

In the past few years, there have been a number of research investigations that clearly demonstrated the application of many whole genome prediction models and GS approaches for disease resistance plant breeding (Table 1). These research efforts have confirmed the effectiveness of different GS models to capture and predict the genetic variation for disease resistance, particularly quantitative disease resistance [43-55].

One of the most studied plant disease for the application of GS models in disease resistance plant breeding is rusts in wheat. The major rust pathogens of wheat are stem rust, yellow/stripe rust, and leaf brown rust [2]. Wheat rusts disease resistance can be either qualitative or quantitative, with several race specific R genes [56]. The R genes usually detected in the seedling stage. However, quantitative disease resistance loci generally confer resistance only in the adult plants, as a result it is referred as adult plant quantitative disease resistance (APR). APR has been shown to vary to some degree across environments, which could be due to race, temperature, or other unknown environmental factors.

<table>
<thead>
<tr>
<th>Crop</th>
<th>Disease type</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>Rust</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>Septoria tritici blotch</td>
<td>[44]</td>
</tr>
<tr>
<td>Barley</td>
<td>FHB*</td>
<td>[45]</td>
</tr>
<tr>
<td>Wheat</td>
<td>Stem rust</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>FHB*</td>
<td>[46]</td>
</tr>
<tr>
<td>Wheat</td>
<td>FHB*</td>
<td>[47]</td>
</tr>
<tr>
<td>Wheat</td>
<td>Stem rust</td>
<td>[48]</td>
</tr>
<tr>
<td>Wheat</td>
<td>FHB*</td>
<td>[49]</td>
</tr>
<tr>
<td>Maize</td>
<td>NCLB**</td>
<td></td>
</tr>
<tr>
<td>Maize</td>
<td>Gibberella ear rot</td>
<td>[50]</td>
</tr>
<tr>
<td>Cassava</td>
<td>Cassava mosaic disease</td>
<td></td>
</tr>
<tr>
<td>Cassava</td>
<td>Cassava anthracnose disease</td>
<td>[52]</td>
</tr>
<tr>
<td>Barley</td>
<td>FHB*</td>
<td>[53]</td>
</tr>
<tr>
<td>Wheat</td>
<td>Stem rust</td>
<td>[54]</td>
</tr>
<tr>
<td>Wheat</td>
<td>Yellow rust</td>
<td>[54]</td>
</tr>
<tr>
<td>Wheat</td>
<td>FHB*</td>
<td>[55]</td>
</tr>
</tbody>
</table>

Though disease resistance breeding efforts using major R genes for wheat rust have produced highly resistant varieties, their resistance breaks very shortly [2]. Therefore, breeding strategy for minor gene resistance highly preferred to generate varieties with durable disease resistance. Quantitative disease resistance is race nonspecific and sometimes effective against more than one rust species [57].

Previous studies on the application of GS for wheat rust disease resistance reported moderate to high prediction accuracies ranging from 0.3 to 0.8, and were able to predict both within and across environments with comparable accuracy [54,55]. Besides it is also indicated that GS can be relatively with small training population [46,48].

Besides, Fusarium head blight (FHB) is a serious disease in different parts of the world every year causing a very high yield loss and reduction of grain quality [58,59]. Different previous studies proven that resistance to FHB in wheat is quantitatively inherited [60,61] and that genetic variation for FHB resistance is predominantly additive [62,63] showing accumulation of resistance genes possible. The first report on GS models for FHB resistance in wheat found that various prediction models have high accuracy to be useful in breeding [55]. Besides, various studies evaluated RR-BLUP as well as a nonlinear and a variable selection model on a diverse set of breeding lines for FHB and confirmed different traits associated with FHB resistance, could be predicted with moderate to high accuracy [47].

One of the major challenges with GS for disease resistance is that many disease resistances are highly heritable, which makes phenotypic selection hard to beat in both per-cycle and per unit genetic gain [46,47]. In order for GS to outperform phenotypic selection for quantitative disease resistance, it may be necessary to increase the selection intensity in addition to decreasing cycle time. For low-heritability traits, using genotype in addition to phenotype can substantially improve selection accuracy [64,65].

Recent advances in disease resistance breeding

The main advantage of using GS for durable disease resistance plant breeding is, quantitative resistance can be selected in the presence major R genes [2]. Several R genes confer a very high level of resistance and can severely mask the effect of other resistance
QTLs in the background. In different plant breeding populations, both quantitative resistance loci and R genes are present together. In this case the ideal genotypes are those individuals with at least one effective R gene along with a high level of quantitative resistance. Therefore, this would lead to a very high level of disease resistance provided that the R gene is effective, and when the R gene is rendered ineffective, quantitative resistance in the background would still provide a good level of protection against yield loss.

The major advantage of genomics assisted plant breeding for disease resistance includes, it provides very good opportunities for combined selection of both R gene and quantitative resistance. The process mainly begins with crossing two parents that segregate for a race specific R gene based resistance to generate a population of selection candidates as illustrated in Figure 4. The selection candidates are then genotyped with WGS or GBS for genome wide markers. The selection candidates without effective R genes are then phenotyped for quantitative disease resistance. The phenotypes and genotypes are used to train genomic prediction model (GS model), which is then used to predict the level of quantitative resistance in the individuals with the effective R genes [2]. The individuals that combine the highest level of quantitative resistance with R genes (QR + R-gene candidates) are then selected for further advancement.

RESULTS AND DISCUSSION

To speed up the development of improved crop varieties to feed an ever increasing world population, genomics assisted breeding is an important tool in any breeding program in the world. With traditional plant breeding and marker assisted selection, there has been several successes in breeding for diseases resistance. Most research works for disease resistance has been focused on major disease resistance genes which are highly effective although very vulnerable to breakdown with rapid changes in pathogenic races. There were several attempts in constructing genetic maps for important disease resistance genes and identifying several markers for MAS [2,10,11]. However, for many diseases, MAS quickly becomes very complex and intractable in an applied breeding program [2].

In contrast, breeding for minor gene quantitative resistance can produce more durable plant varieties although it is very slow and challenging breeding. To develop improved varieties with more durable disease resistance, the most appropriate approach is genomic selection (GS) than marker assisted selection or conventional breeding. The primary objective for implementing GS in plant breeding program is selection for high yield to reduce the time and costs associated with yield testing. As a result, GS is currently broadly accepted as an efficient method to improve genetically complex quantitative traits. Besides, recent developments in different sequencing technologies created excellent opportunities to apply different genotyping techniques including whole genome sequencing, genotyping by sequencing (GBS) in different crop species yielding an appropriate large number of markers for genomic. Therefore, GS helps to speed up the rate of genetic gain in breeding by using whole genome data to predict the breeding value of offspring. GS breeding for quantitative resistance will therefore necessitate whole genome prediction models and selection methodology as implemented for classical complex traits such as yield. With the application of GS for yield and other several economically important traits, whole genome marker profiles will be available for the entire set of breeding lines, enabling genomic selection for disease resistance with no additional direct cost.

GS has a great potential to improve plant breeding progress through increased selection intensity and decreased cycle time. In several studies, field based high throughput phenotyping highly demonstrated potential to measure different phenotypic traits faster and more accurately [66]. The combined power of genomics and phenomics is highly expected to lead to new eras in plant breeding and functional genomics [66,67]. Besides, current advances in disease resistance plant breeding through applications of a two stream GS breeding scheme using a combination of GS + de novo GWAS models (GS+) and genomic prediction model that combine the highest level of quantitative disease resistance with R genes (QR + R-gene candidates) are expected to further advance the implementation GS in disease resistance plant breeding.

CONCLUSION

Currently, there is a strong confidence and foundation for application of GS breeding for disease resistance plan breeding. Several studies successfully demonstrated the great potential of the current genomewide prediction models efficiency to predict and select for quantitative disease resistance. With the implementation of GS for yield and other economically important traits, whole genome marker profiles will be available for the entire set of breeding lines, enabling genomic selection for disease resistance with no additional direct cost. Generally, genomic selection can increase breeding progress through increased selection intensity and decreased cycle time. Toward this major objective, the combined power of genomics and high throughput phenotyping technologies are expected to lead GS to new eras in disease resistance plant breeding and functional genomics studies. Besides, recent developments in disease resistance breeding, through a two stream GS breeding scheme using a combination of GS + de novo GWAS models (GS+) and GS model that combine the highest level of quantitative disease resistance with R genes (QR+R-gene) will further revolutionize the implementation GS in disease resistance plant breeding.

REFERENCES


