Genetic Diversity, QTL Mapping, and Marker-Assisted Selection Technology in Cotton (*Gossypium* spp.)

Fakhriddin N. Kushanov[1,2]*, Ozod S. Turaev[1], Dilrabo K. Ernazarova[1,2], Bunyod M. Gapparov[1], Barno B. Oripova[1,2], Mukhlisa K. Kudratova[1], Feruza U. Rafieva[1], Kuvandik K. Khalikov[1], Doston Sh. Erjigitov[1], Mukhammad T. Khidirov[1], Madina D. Kholova[1], Naim N. Khusenov[3], Roza S. Amanboyeva[2], Sukumar Saha[4], John Z. Yu[5] and Ibrokhim Y. Abdurakhmonov[3]

[1]Institute of Genetics and Plant Experimental Biology, Academy of Sciences of the Republic of Uzbekistan, Tashkent, Uzbekistan, [2]Department of Biology, National University of Uzbekistan, Tashkent, Uzbekistan, [3]Center of Genomics and Bioinformatics, Academy of Sciences of the Republic of Uzbekistan, Tashkent, Uzbekistan, [4]Crop Science Research Laboratory, USDA-ARS, Washington, DC, United States, [5]Southern Plains Agricultural Research Center, USDA-ARS, Washington, DC, United States

Cotton genetic resources contain diverse economically important traits that can be used widely in breeding approaches to create of high-yielding elite cultivars with superior fiber quality and adapted to biotic and abiotic stresses. Nevertheless, the creation of new cultivars using conventional breeding methods is limited by the cost and proved to be time consuming process, also requires a space to make field observations and measurements. Decoding genomes of cotton species greatly facilitated generating large-scale high-throughput DNA markers and identification of QTLs that allows confirmation of candidate genes, and use them in marker-assisted selection (MAS)-based breeding programs. With the advances of quantitative trait loci (QTL) mapping and genome-wide-association study approaches, DNA markers associated with valuable traits significantly accelerate breeding processes by replacing the selection with a phenotype to the selection at the DNA or gene level. In this review, we discuss the evolution and genetic diversity of cotton *Gossypium* genus, molecular markers and their types, genetic mapping and QTL analysis, application, and perspectives of MAS-based approaches in cotton breeding.

Keywords: cotton, genetic diversity, DNA markers, QTL mapping, GWAS, marker-assisted selection

INTRODUCTION

Cotton is one of the oldest cultivated crop plants and it is grown as the main source of raw materials for the textile industry. More than 103 million tons of textile fibers were consumed in 2019 and cotton fiber had a market share of approximately 24% in 2020 over 26 million tons of cotton was produced worldwide (ICAC, 2021). Cotton is valued for its fiber quality in the global market and it determines price of the fiber. Cotton fiber faces a grave challenge by a chemically produced (synthetic) fiber. Synthetic fibers currently control over 75% of the global market share

in textile fiber consumption (ICAC, 2021). Therefore, competition from synthetics has increased textile industry demands for cotton fiber with high quality and superior spinning performance. However, one of the serious impediments to improve the cotton fibers is the narrow genetic base in Upland cotton. In this regard, there is a constant need to introduce genetic diversity into the new varieties with excellent fiber quality and high yield potentials (Abdurakhmonov et al., 2008, 2009; Yu et al., 2014).

Creating new varieties using traditional breeding methods, specifically, the introduction of genes of desirable traits to the elite cotton from a donor source to the elite is very laborious and requires at least 10 years of hard work (Abdurakhmonov et al., 2009). At the same time, there is often a negative correlation between desired traits, such as fiber quality and fiber yield or resistance to abiotic environmental stress factors, which often prevent the breeder from the effective selection and breeding (Griffith and Crowfoot, 1934; Nicholson, 1960; Muthukumaran, 2016).

Similar problems can be solved by introducing modern biotechnological approaches based on the use of molecular markers in breeding programs (Abdurakhmonov et al., 2008, 2009). Modern breeding programs for the accelerated generation of new varieties provide in-depth study of breeding material at both the phenotypic and genotypic levels. DNA markers are commonly referred to as molecular markers, although previously widely used isozymes and other marker systems were based on protein polymorphism. With the introduction of DNA marker technology into the practice of plant breeding, new opportunities have emerged for studying genetic diversity, identifying and improving economically useful crop traits (Preetha and Raveendren, 2008). The advent of molecular marker technology provides breeders with powerful new tools for identifying complex quantitative traits. Moreover, DNA marker technology allows breeders to increase efficiency and reduce costs and time to create new varieties and hybrids compared to traditional breeding methods. A large number of DNA markers and genes controlling resistance to biotic and abiotic stresses, yield, and quality traits have been identified and mapped for many crop species in recent years (Zhang et al., 2013; Han et al., 2019; Wang et al., 2019).

Molecular markers gave a great opportunity to improve the efficiency and precision of crop improvement programs *via* marker-assisted selection (Collard and Mackill, 2008). The use of DNA markers in plant breeding is called marker-assisted selection (MAS) and it is a component of the molecular breeding approach (Collard and Mackill, 2008). MAS technology allows conducting the selection at any stage of plant growth and development. In short, the development of the MAS technology aimed at the selection of crops led to high achievements in genomics, which became a vital part of agricultural science.

TAXONOMY, EVOLUTION, AND GENETIC DIVERSITY OF COTTON *GOSSYPIUM* GENUS

The *Gossypium* L. (the cotton genus) has a long history of taxonomic and evolutionary study. The *Gossypium* genus,

belonging to the tribe *Hibisceae* (*Malvaceae* family), includes approximately 46 diploid and 7 allotetraploid species (Fryxell, 1979; Wendel and Albert, 1992; Wendel et al., 2009; Paterson et al., 2012; Wendel and Grover, 2015). The diploid ($n = x = 13$) species of cotton is classified into eight (A-G and K), and the tetraploid ($n = 2x = 26$) species into one (AD) cytogenetic group (Brubaker et al., 1999; Wendel and Cronn, 2003). Mainly 4 species are cultivated in around 90 cotton producing countries of the world: *G. hirsutum* L., which occupies more than 90% of the total area, followed by *Gossypium barbadense* L., approximately 8% and only 1% two diploid species – *G. arboreum* L. and *G. herbaceum* L. The genome size of diploid cottons varies from about 880 to 2,500 Mb, as well as the tetraploid cotton genome has an estimated size of 3,000 Mb (Hendrix and Stewart, 2005).

Based on their origin, diploid cotton species are divided into two types: African-Asians and Australian (Blenda et al., 2012). *G. arboreum* L. and *G. herbaceum* L, having twisted fiber, were originally grown on the Asian continent. Subsequently, as a result of hybridization between diploid A-genomic (Asian) and D-genomic (Mexican) representatives, which occurred about 1.5 million years ago, formed five allotetraploid species of cotton (Yu et al., 2013).

Upland cotton (*G. hirsutum*) is widely cultivated, industrial cotton among all species of the *Gossypium* genus (Iqbal et al., 2001). The origin of this species is considered Guatemala, but it is distributed throughout Central America and in Caribbean countries. According to Mauer (1954), there are four groups of subspecies of *G. hirsutum*: *mexicanum, punctatum, paniculatum, and euhirsutum*. These four groups of subspecies include several wild races, such as *yucatanense*, *richmondi*, *latifolium*, *palmeri*, *morilli*, *purpurascens,* and their variety samples, as well as a number of cultivated samples (Adams et al., 2004).

Gossypium barbadense (Egyptian, Sea Island or Pima cotton) is widely distributed throughout most of South America, Southern Mesoamerica, and the Caribbean (Rathore et al., 2006). This species of *Gossypium* genus initially sprouted on ribbed coastal islands and in the valleys of the United States and was named Sea Island cotton. Then, Sea Island cotton was introduced into the Nile Valley of Egypt, where it is widely cultivated for the production of long and thin fibers (Abdalla et al., 2001).

The remaining three tetraploid species (AD_3–AD_5) are common in other regions. So, for example, *G. mustelinum* Miers ex Watt widely distributed in Northeast Brazil (Wendel et al., 1994), *G. darwinii* Watt is an endemic of the Galapagos Islands (Wendel and Percy, 1990), and *G. tomentosum* Nutt ex Seem is endemic to the Hawaiian Islands (Hawkins et al., 2005). All of them are truly wild species (Westengen et al., 2005). Genetic diversity represents the existence of various variants of biological forms or the degree of morphological and physiological features of organisms within populations (often called traits), which are essential for biological individuals both for a positive response to a rapid change in the environment and for their survival. The lack of genetic diversity or its narrowness in various types of crops creates the potential threat

to plant productivity due to the vulnerability of genetically homogeneous varieties to new biotic and abiotic stresses. Consequently, the broad genetic diversity of crops has the potential to protect them from new diseases, pests, and unexpected global environmental changes (Abdurakhmonov et al., 2007).

Thus, the genus *Gossypium*, covering large geographical and ecological niches, has a wide amplitude of morpho-biological and genetic diversity, preserved in centers of origin of cotton *in situ*, in collections of germplasm of cotton *ex situ*, as well as in materials of breeding programs throughout the world. These resources can be successfully used in cotton breeding programs to transfer economically valuable traits from wild species to the cultivated genotypes in order to create promising competitive varieties.

IMPORTANCE OF GENETIC DIVERSITY FOR COTTON IMPROVEMENT

Cotton faces various problems in production and marketing, such as competition from synthetic fiber, wide variability from year to year in yield, and plus new requirements for fiber quality due to technological changes in the textile industry (Perkins et al., 1984; Esbroeck et al., 1999). A longer fiber, like that of low-yielding cotton species *G. barbadense*, is genetically stronger, thinner, and more uniform than a shorter fiber of the widely sown, early-growing and high-yielding cotton *G. hirsutum* (Perkins et al., 1984). Changing these fiber properties in medium fiber cotton is a big challenge facing cotton breeding programs around the world (Esbroeck et al., 1999).

Since *G. hirsutum* is the most widely cultivated species in the world, due to its high yield, early maturity, and unpretentiousness of cultivation, much research has been devoted to the analysis of its genetic diversity (Abdurakhmonov et al., 2008, 2009; Blenda et al., 2012). At the same time, less attention is paid to the study of *G. barbadense*, the second most cultivated cotton species, since the varieties of this species have lower yields and weaker indicators of other economically important traits compared to the *G. hirsutum* species (Wendel and Percy, 1990; Fryxell, 1992; Yu et al., 2013). Cotton researchers constantly carry out selection measures for crossing these two species, with the goal of transferring the superior fiber quality components specific only to *G. barbadense*, to the cultivated varieties of *G. hirsutum* (Fryxell, 1992). It should be noted through interspecific hybridization between *G. barbadense* and *G. hirsutum* species, the desired alleles for most QTLs associated with fiber quality are transmitted from *G. barbadense* (Lacape et al., 2005). Moreover, scientists proved that *G. hirsutum*, in turn, can also contribute to the improvement of fiber length, strength, and micronaire (Lacape et al., 2005; Yu et al., 2013). This discovery confirms the assumption that in generations of interspecific hybrids having a mosaic genome, best gene allele combinations can be achieved (Tanksley and Nelson, 1996). In most cases, the use of

interspecific crossing of *G. barbadense* with *G. hirsutum* with classical breeding methods to improve fiber quality traits, such as length, strength, and micronaire in Upland cotton, did not lead to the expected stable introgression. According to the study published by the researchers (Tanksley and Nelson, 1996; Lacape et al., 2005; Yu et al., 2013, 2014), the solution to these problems at the genetic level requires knowledge of broader variation in the cotton germplasm. However, according to the review by Zhang et al. (2014), new introgression lines with high yields and fiber quality were developed as a result of interspecific hybridization between *G. hirsutum* and *G. barbadense* (Zhang, 2011).

Narrow genetic diversity can be caused by the intensive use of one or several closely related genotypes in breeding programs (Iqbal et al., 2001) or the consequence of a "genetic bottleneck" at the historical transformation of wild plants into cultural forms, which led to the subsequent distribution of a limited number of genotypes (Abdurakhmonov et al., 2007). The productivity, viability, and success of cotton breeding, like many other crops, also depend on the diversity of the gene pool (Esbroeck et al., 1999). According to the authors (Iqbal et al., 2001), the existing and projected problems of the world cotton breeding programs related to the narrowness of the genetic base of the germplasm arise because of the complexity of the tasks and the lack of new genomic approaches for mobilizing beneficial genetic variations from various exotic cotton species of the *Gossypium* genus into breeding varieties.

MOLECULAR MARKERS AND THEIR TYPES

Genetic markers are any observable inherited traits that differ individually from one another. For many years, they have been used to characterize genetic diversity for crop improvement. This tool is particularly useful in analyzing complex quantitative traits. Genetic markers can be divided into three main types: morphological (or phenotypic), cytological, and molecular markers (**Figure 1**). Morphological markers represent the actual polymorphism of the phenotype and they are identified easily, quickly, and most importantly with minimum laboratory equipment (Nadeem et al., 2018). The physical maps based on morphological and cytological markers lay a foundation for genetic linkage analysis using molecular methods (Jiang, 2013). However, direct use of cytological markers has been very limited in genetic mapping and plant breeding (Jiang, 2013; Ahmad et al., 2017; Nadeem et al., 2018). Nevertheless, the number of informative morphological markers is very small (Eagles et al., 2001). The low occurrence rate and many other deficiencies did not allow morphological markers to enter widely into selection practice.

Until recent advances in molecular genetics, breeders have improved both qualitative and quantitative hereditary traits by traditional breeding methods based on evaluation and selection for phenotypic variation, which are resource-intensive
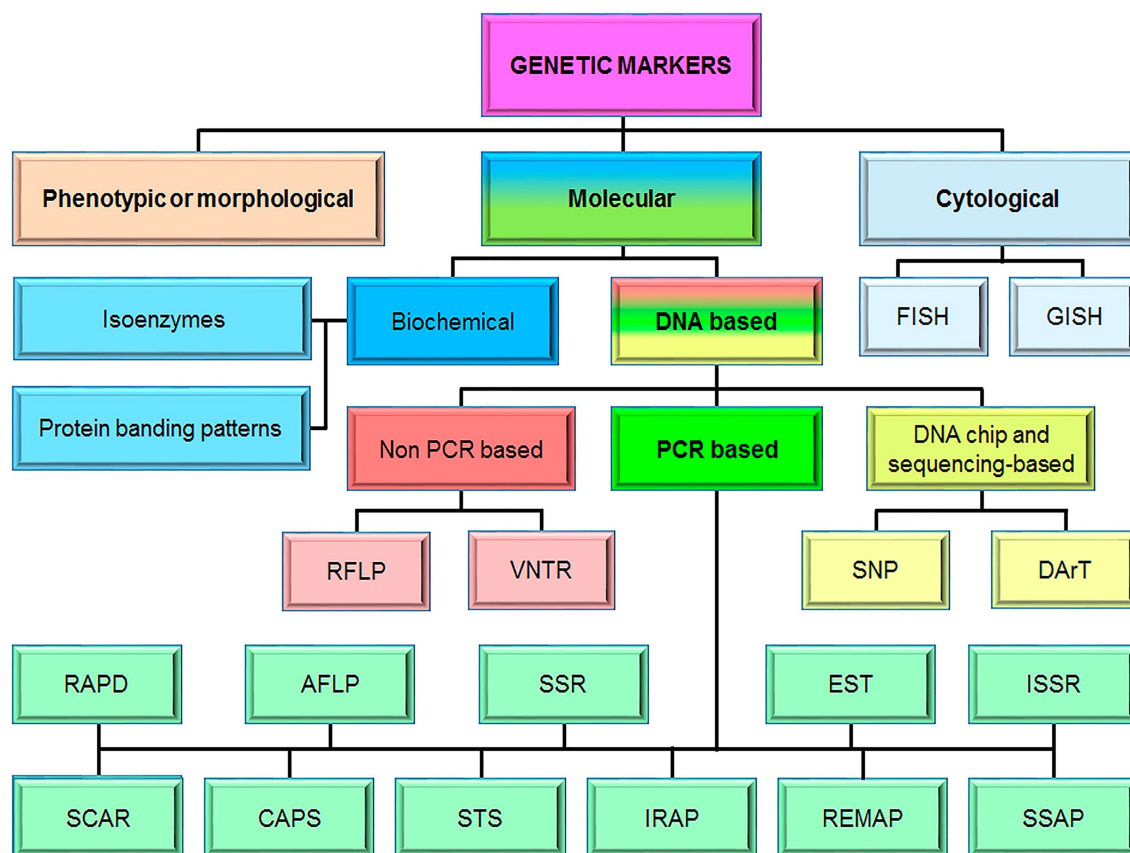
FIGURE 1 | The classification of genetic markers. FISH, fluorescence *in situ* hybridization; GISH, genome *in situ* hybridization, RFLP, restriction fragment length polymorphism; VNTR, variable number tandem repeat; SNP, single-nucleotide polymorphism; DArT, diversity arrays technology; RAPD, random amplified polymorphic DNA; AFLP, amplified fragment length polymorphism; SSRs, simple-sequence repeats; EST, expressed sequence tag; ISSR, inter simple-sequence repeat; SCAR, sequence-characterized amplified region; CAPS, cleaved amplified polymorphic sequences; STS, sequence-tagged sites; IRAP, inter-retrotransposon amplified polymorphism; REMAP, retrotransposon-microsatellite amplified polymorphism, and SSAPs, sequence-specific amplification polymorphisms.

(Said et al., 2015a). Currently, two main types of molecular markers, biochemical and DNA markers are available for genetic studies (Lander and Botstein, 1989; Jansen, 1994). It should be noted that the first molecular markers were created based on the analysis of protein polymorphism. However, the possibilities of biochemical markers are limited by the low level of protein polymorphism in populations, restrictions in the choice of biological material, and the time of its collection (Kumar et al., 2009).

DNA-based molecular markers are genetic markers that are analyzed at the DNA level. This marker system plays a huge role in the study of gene inheritance and their allelic status. Such markers are used to analyze genetic polymorphism and phylogenetic relationships between species, populations, and individuals, as well as to identify diagnostic markers that are closely linked to the genes controlling the economically valuable traits of crops (van Ooijen, 1992). An important tool for finding DNA markers is linkage mapping, which allows to combine phenotypic data and DNA polymorphism data. Currently, there are many different types of DNA markers, and their numbers are constantly increasing with the

achievement of modern technologies and knowledge of individual genes and genomes of plants in general. So, DNA markers can be divided into three main groups: markers based on hybridization (or non-PCR based), markers based on PCR, and markers based on DNA chips (Singh and Singh, 2015). The PCR-based marker system is very popular and more widely utilized among these groups of DNA markers (Reiter, 2001). Moreover, it is PCR-based DNA markers that are widely introduced into the plant selection process. One of the most informative types of PCR-based DNA markers is microsatellite or SSR (simple-sequence repeat) markers, as they are widely distributed in the genome and have a high level of polymorphism (Vieira et al., 2016).

Thus, molecular or DNA markers are genetic tools that allow plant breeders to perform a variety of tasks. Especially, DNA markers play an important role in the study of genetic polymorphism, inheritance of genes, and their allelic state, in phylogenetic analysis, as well as identification of QTLs that are closely linked with genes controlling the economically valuable traits of plants (Bruford et al., 2003; Mittal and Dubey, 2009).

GENETIC MAPPING AND QTL ANALYSIS FOR AGRONOMICALLY AND ECONOMICALLY VALUABLE TRAITS IN COTTON

Description of Genetic Mapping Approaches

One of the main tasks of molecular markers is the mapping of genes and quantitative trait loci (QTLs; Jansen, 1993). The theory of QTL mapping was first described by Karl Sax (1923) when he observed segregation of seed weight associated with segregation for a seed coat color marker in *Phaseolus vulgaris* L. (Sax, 1923). He noted that one gene controlling seed color should be associated with one or more polygenes controlling seed size. Since the development of the first molecular markers (Grodzicker et al., 1974), a large amount of theoretical and practical results and methodological developments describing the stages of molecular genetic mapping have been accumulated (Lander and Botstein, 1989; van Ooijen, 1992; Jansen, 1994). The concept of genetic linkage mapping is based on the study of the genome of an organism by DNA markers (Goldstein and Weale, 2001; Jannink and Walsh, 2002; Oraguzie et al., 2007), determining the relative position of these markers on linkage groups and determining their genetic association with QTLs (van Ooijen, 1992). Genetic mapping is mainly accomplished in two ways; the first, traditional method, the so-called linkage analysis or QTL mapping (Jansen, 1994) has already become a classic method. This mapping study is carried out using experimental (biparental) populations of F_n generations (Lewis, 2002), backcross (BC_n), recombinant inbred lines (RIL), and/or doubled haploid lines (DHL).

The second method used in the construction of modern genetic maps of plants is the analysis of linkage disequilibrium (LD) and association mapping (Goldstein and Weale, 2001; Jannink and Walsh, 2002). LD mapping uses different lines from natural populations or germplasm collections (Asíns, 2002; Glazier et al., 2002). Thus, when mapping, individuals are divided into genetic classes for each DNA marker used (van Ooijen, 1992). Next, the values and variations of the parameters are calculated and compared between the classes. The identified polymorphism between genetic classes provides information about the relationship of the marker used with the phenotype of interest, and its connection with the QTL (Xu et al., 2017a). After detecting genes those regulate quantitative traits, positional mapping is used based on statistical data analysis.

Currently, a rich arsenal of QTL-mapping methods has been created, which implements various approaches (Wan et al., 2009). The developed methods are based on the well-known principles of parametric and nonparametric analysis of linkage, as well as new approaches using the analysis of components of dispersion (Zhang and Gai, 2008), analysis of associations (Jia et al., 2014), and multipoint mapping (Liu and Muse, 2005). The development of statistical methods follows the path of increasing their power and stability of methods to inaccuracies of genetic models and the incompleteness of empirical data (Price, 2006).

QTL Analysis for Fiber Quality, Stress and Disease Resistance, and Some Morphological Traits in Cotton

Since the development of molecular mapping technology, researchers have created hundreds of genetic maps and identified many QTLs associated with economically valuable traits (Saranga et al., 2001; Gao et al., 2003; Du et al., 2004; Abdurakhmonov et al., 2005; Bolek et al., 2005; Lacape et al., 2005; Yin et al., 2006; Abdurakhmonov et al., 2007; Wang et al., 2007; Qin and Zhang, 2008; Yang et al., 2008; Zan et al., 2008; Goicoechea et al., 2009; Jiang et al., 2009; Wan et al., 2009; Said et al., 2013, 2015a; Tang et al., 2014; Zhang et al., 2015; Kushanov et al., 2016; Diouf et al., 2018; Zhao et al., 2018). As a primary goal, the cotton research community has set QTL mapping with molecular markers associated with fiber yield, quality, and yield traits (Wang et al., 2007; Abdurakhmonov et al., 2008; Abdurakhmonov et al., 2009; Said et al., 2013; Yu et al., 2013; Tang et al., 2014). Some QTLs related to environmental stress resistance, such as drought (Saranga et al., 2001), as well as loci associated with the formation and morphology of stems and leaves (Said et al., 2013), chlorophyll content (Qin and Zhang, 2008), natural leaf defoliation (Abdurakhmonov et al., 2005), and fertility restoration genes (Zhao et al., 2018), are also mapped. Currently, 4,892 QTLs identified either in the populations of *G. hirsutum* or *G. hirsutum* × *G. barbadense* and presented in 156 publications are available in the Cotton QTLdb database[1] (Said et al., 2015b).

The QTL analysis of fiber quality traits of cotton began about 25–30 years ago (Kloth, 1993, 1995). Today, single-nucleotide polymorphism (SNP) markers-based genome-wide association study (GWAS) is widely used for identifying genomic regions that attending to control economically important traits in both natural and experimental populations of cotton. Gapare et al. (2017) have conducted GWAS on natural cotton populations to identify genetic contributions to the fiber quality, plant architecture, and stomatal conductance traits (Gapare et al., 2017). They have used Illumina CottonSNP63 K SNP array for genotyping. The results of analysis showed that 17 and 50 significant SNPs associated for fiber length and micronaire, respectively. Sun et al. (2018) performed a GWAS of fiber quality traits of 719 diverse accessions of upland cotton using Cotton 63 K Illumina Infinium SNP array (Sun et al., 2018). Germplasm resources were screened using more than 10.5 thousand polymorphic SNPs distributed in 26 chromosomes, and 46 significant SNP markers related to fiber quality traits were identified. These important SNPs are distributed on 15 chromosomes and are involved in 612 unique candidate genes, many of which are associated with polysaccharide biosynthesis, signal transmission, and protein translocation. In addition, scientists have identified 163 and 120 fiber genes related to length and strength, respectively. Ma et al. (2018) have identified more than 3.6 million SNPs by re-sequencing 419 cotton accessions and conducted GWAS of 13 fiber-related traits (Ma et al., 2018). More than 7.3 thousand SNPs were associated

---

[1]http://www2.cottonqtldb.org

with fiber quality traits and covered 4,820 genes; more fiber-related genes were determined in D subgenome than in the A subgenome. Liu et al. (2020) have identified 42 SNPs and 31 QTLs significantly associated with five fiber quality traits (Liu et al., 2020). Twenty-five QTLs are the same as QTLs identified in previous studies, and six novel QTLs were firstly identified in their work. In these QTL regions, 822 genes were determined as well two pleiotropic SNPs associated with fiber elongation, strength, length, uniformity, and strength were identified.

Besides, Fang et al. (2014) have identified 131 fiber QTLs and 37 QTL clusters on experimental mapping population using 2,132 polymorphic SSR markers out of 15,538 SSRs (Fang et al., 2014). Two QTL clusters were determined on chromosomes 7 and 16. Comparison of 131 QTLs showed that 77 were identified in the previously studies, and 54 novel QTLs. Recently, Islam et al. (2016) used an Upland cotton multi-parent advanced generation inter-cross (MAGIC) population, developed through random mating of 11 diverse cultivars for five generations, in a molecular map of SNP markers associated with fiber traits from four environments (Islam et al., 2016). They used a high-throughput genotyping approach of Genotyping-by-Sequencing (GBS) developing about 6,071 SNP markers and 223 microsatellite markers of 547 recombinant inbred lines (RILs) of the MAGIC population. They used a GWAS using a mixed linear model to identify markers significantly associated with fiber QTLs. They discovered one QTL cluster associated with four fiber quality traits [short fiber content (SFC), strength (STR), length (UHM), and uniformity (UI)] on chromosome A07. They further identified several candidate genes related to fiber quality attributes in this region. Gene expression and amino acid substitution analysis suggested that regeneration of bulb biogenesis 1 (GhRBB1_ A07) gene is a candidate for superior fiber quality in Upland cotton. The DNA marker CFBid0004 designed from an 18 bp deletion in the coding sequence of GhRBB1_A07 in Acala Ultima is associated with the improved fiber quality in the MAGIC RILs and 105 additional commercial Upland cotton cultivars. Using GBS technology and a MAGIC population enabled more precise fiber QTL mapping in Upland cotton. Normally Acala Upland cotton lines carried superior fiber quality traits compared to other Upland cotton. Thyssen et al. (2018) have identified identify seven highly significant fiber quality loci associated with six major cotton fiber quality traits in a MAGIC population using GWAS and whole-genome sequencing (Thyssen et al., 2018). At these loci, they found 14 genes with non-synonymous SNPs. Sarfraz et al. (2021) have conducted subsequent genome-wide predictions along with association analyses that uncovered a set of highly significant key SNPs related to agronomic and fiber quality traits (Sarfraz et al., 2021). The integration of a GWAS with RNA-sequence analysis yielded 275 candidate genes near the key SNPs. The main part of candidate genes is associated with fiber micronaire and lint percentage. As well, 54 putative candidate genes were identified in association with the heterosis of quoted traits.

Xu et al. (2020) have carried out the study to identify candidate genes related to fiber quality traits through the integration of meta-QTL, significant SNP, and transcriptomic data (Xu et al., 2020). Scientists have used fiber quality traits associated 884 QTLs from 12 studies for meta-QTL analysis based on reference genome TM-1. As a result of meta-analysis, 74 meta-QTLs were identified, in particular 19 meta-QTLs for fiber length, 18 meta-QTLs for fiber strength, 11 meta-QTLs for fiber uniformity, 11 meta-QTLs for fiber elongation, and 15 meta-QTLs for micronaire. As well as with 8,589 significant SNPs associated with fiber quality traits gathered from 15 studies, 297 candidate genes were determined in the meta-QTL intervals, 20 of which showed high expression levels specifically in the developing fibers.

Cotton is mainly grown in the regions which are affected by abiotic stresses, such as drought and salt (Abdelraheem et al., 2019). There is an urgent need to study of genetic bases of abiotic stress resistance and to improve drought resistance of cotton (Li et al., 2020c). Recently, Abdelraheem et al. (2019, 2021) have identified drought and salinity stress resistance-related QTLs using SNP markers on an inter-cross mapping population (Abdelraheem et al., 2019, 2021). A total of 20 QTL were determined for drought tolerance and 23 QTL for salt tolerance out of 473,516 polymorphic SNPs. Nine QTL identified were in common between drought and salt tolerance, indicating a general genetic basis for both traits. Li et al. (2020a) have studied the genetic architecture for drought resistance in cotton using phenomics-based GWAS analysis (Li et al., 2020a). In their study, scientists have used an automatic phenotyping platform to examine drought stress tolerance at the seedling stage, across a natural population of upland cotton accessions. The phenomics data allowed to identify 390 genetic loci and drought tolerance-related genes by GWAS. Zhu et al. (2020) have conducted GWAS using 57,413 high-quality SNPs in 316 G. hirsutum accessions that grown under four salt conditions over 2 years and identified a total of 42, 91 and 25 stable QTLs for single boll weight, lint percentage, and boll number per plant, respectively (Zhu et al., 2020).

At the same time, great progress was achieved in the QTL mapping, determining resistance to Verticillium wilt (VW) and Fusarium wilt (FW). Most studies on the mapping of resistance to this pathogen have been conducted in germplasm accessions and diverse mapping populations. For instance, Li et al. (2017) have conducted a study to examine the genetic architecture of cotton Verticillium wilt disease resistance (Li et al., 2017). They performed a GWAS in 299 cotton accessions and 85,630 SNPs detected using the specific-locus amplified fragment sequencing (SLAF-seq) approach and were detected a total of 17 significant SNPs. Haplotype block structure analysis predicted 22 candidate genes for VW resistance. Zhang et al. (2019) have carried out GWAS analysis using 473,516 SNPs/Indels in 550 recombinant inbred lines (RILs) of multi-parent advanced generation inter-cross (MAGIC) population (Zhang et al., 2019). Consequently, a significant QTL for FW resistance on chromosome c14 was identified. An interesting aspect is that a major resistance gene (B12) for bacterial blight resistance and one QTL for Verticillium wilt resistance were also identified within the QTL region in this MAGIC population. Another group of scientists has conducted a GWAS using high-density genotyping with the CottonSNP63K array

and identified a total of 15 and 13 QTL for VW and FW resistances were anchored by 30 and 56 significant SNP markers, respectively (Abdelraheem et al., 2020). Similar studies were also conducted by Bardak et al. (2021) to discover the genetic markers associated with the Verticillium wilt disease in a Worldwide Collection of Cotton (*Gossypium hirsutum* L.; Bardak et al., 2021). Through the association mapping analysis, common SNP markers were obtained using 4,730 SNP alleles. As a result, 23 markers were associated with defoliating (PYDV6 isolate) pathotype, 21 markers with non-defoliating (Vd11 isolate) pathotype, 10 QTL with Disease Severity Index (DSI) of the leaves at the 50–60% boll opening period, and 8 markers were associated with DSI in the stem section.

Also, some reports have been published on QTL and/or gene mapping of flowering time using diverse marker technology. Researchers have identified more than 30 candidate genes that are involved in various flowering processes in Upland cotton (Lai et al., 2011). In addition, Zhu and Kuraparthy (2014) managed to localize the photoperiod-sensitive locus Gb_Ppd1 and several closely related SSR markers on the cotton chromosome 25. Guo et al. (2008) presented the mapping of flowering-time QTL, assessed by node of first fruiting branches in cotton (Guo et al., 2008). Using more than four thousand SSR markers, researchers identified about 60 loci, associated with early maturing traits of cotton (Li et al., 2013b). Recently, using 212 SSR and 3 cleaved amplified polymorphic sequence (CAPS) markers, 6 QTLs were identified that directly associated with flowering time and photoperiodic flowering in the F_2 population, whereas 7 QTLs were discovered in F_3 generation (Kushanov et al., 2017). Li et al. (2020b) have reported a cotton genome variation map that is generated by the re-sequencing of 436 cotton accessions (Li et al., 2020b). Whole-genome scans for sweep regions identified 357 putative selection sweeps covering 112 Mb of the upland cotton genome, containing 5,184 genes. These genes were functionally associated with flowering-time control, hormone catabolism, aging, and defense response adaptations to climate changes.

Some QTLs related to the formation and morphology of stems and leaves (Said et al., 2013), chlorophyll content (Qin and Zhang, 2008), natural leaf defoliation (Abdurakhmonov et al., 2005), and fertility restoration genes (Zhao et al., 2018) are also mapped.

Although traditional QTL mapping based on the biparental crossing is still an important method for identifying desired genes/loci in plant chromosomes, it has nevertheless become a kind of "modern classical analysis" method. The disadvantage of this method is its low resolution only allowing for the evaluation of a few alleles over a rather long period of analysis. At the same time, markers detected during QTL mapping and specific for some lines may not be specific for other populations or germplasm of a given cultivation.

At the same time, another problem remains topical – to obtain genotypes that are not only resistant to diseases, but at the same time having high yield and superior fiber quality. One of the ways to solve this problem is the interspecific crossing between *G. barbadense* and *G. hirsutum* varieties,

which differ in the indicated characteristics. However, these attempts are not always successful due to the problems associated with sterility, cytological impairments, and distorted segregation arising from interspecific crosses. Thus, QTL mapping is one of the powerful methods for improving agricultural crops, which allows using the marker-assisted selection technology to introgress the genes of interest from donor lines to breeding material.

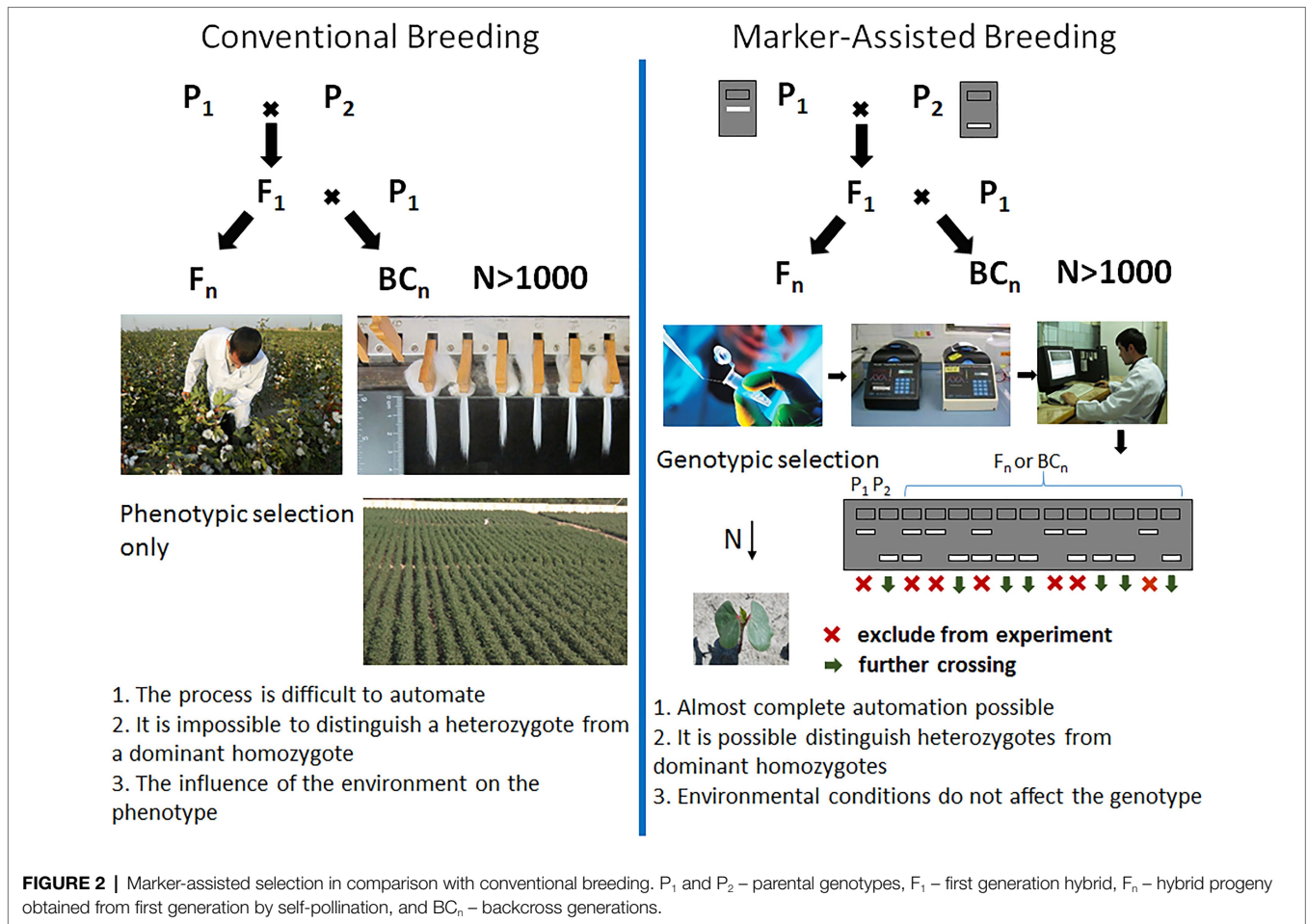# APPLICATION AND PERSPECTIVES OF MAS IN COTTON BREEDING

Application of MAS

The concept of using linked genes has arisen to follow the inheritance of genes that control other traits. It was launched in 1961 by Thoday, who made the first attempt to map and characterize all polygenes that affect the line using monogenic markers (Thoday, 1961; Mutschler et al., 1987). When he worked with morphological markers, the main practical limitation was availability of few suitable markers are available.

With the advent of DNA marker technology and the QTL-mapping approach, the possibilities of breeding for crop improvement have increased significantly. The use of marker-assisted breeding revolutionized the process of creating crop varieties, reducing field trials at an early stage of breeding, and reducing the time required by almost half (**Figure 2**). At the same time, DNA markers associated with traits of interest allow breeders to accurately select individuals based on genotype. This approach is very useful in cases where the trait of interest is complex and time consuming to assess. Moreover, the desired alleles in wild relatives with a low phenotype can also be identified with DNA markers. Such transgressive loci of wild species can be selected and used to create new varieties with a more desirable phenotype, introducing useful variations in crops.

To carry out marker-assisted selection, a large number of polymorphic markers must be identified by analyzing the whole genome. It is necessary to evaluate the marker informativity between the parents used in the crossing, and this can be used to assess the segregating population for the absence or presence of this genetic marker. The benefits of genetic selection can be maximized by increasing the genetic pool or population size so that individuals with an unusual genotype can be identified. At the same time, an increase in the number of markers used proportionally increases the reliability to assess the genome structure. In order to use molecular selection in large-scale breeding programs, it is necessary to introduce automated technologies.

Thus, the basic principle of MAS technology is to identify a tight linkage between the marker and the gene controlling the trait, and subsequently using this association for practical purposes to create new varieties and breeding lines. After the association between the marker and trait has been identified, the creation of new genotypes is carried out using traditional breeding methods, such as hybridization, backcrossing, self-pollination, and selection (**Figure 3**). Because of using MAS technology, a breeder can get rid of the problem of transfer

FIGURE 2 | Marker-assisted selection in comparison with conventional breeding. P_1 and P_2 – parental genotypes, F_1 – first generation hybrid, F_n – hybrid progeny obtained from first generation by self-pollination, and BC_n – backcross generations.

undesirable genes from the donor, which often occurs during the crossing.

Marker-assisted selection (MAS) technology refers to any form of molecular selection that uses genetic markers to crop improvement (Choudhary et al., 2008; Sebastian et al., 2012). Basically, depending on the goal of the research, MAS is used for the following tasks: (i) Evaluation of the purity and identity of the varietal material and the assessment of the genetic diversity of modern varieties (Williams et al., 1993; Sonah et al., 2011), (ii) Introgression of genes/QTL loci in various MAS schemes (Mackay, 2001; Hospital, 2009), and (iii) Combining several genes/loci of QTL donor lines into one genotype and, thus, the creation of new lines that have several useful traits (Visscher et al., 1996; Hospital, 2009).

In order to effectively use molecular breeding and DNA markers, several strategies for MAS technology have been proposed. However, four main breeding schemes are widely used for crop improvement in practice: (i) Marker-assisted Backcross Selection (MABS), (ii) Marker-assisted Recurrent Selection (MARS), (iii) Marker-assisted Gene Pyramiding (MAGP), and (iv) Genomic Selection (GS; Frisch et al., 1999; Ribaut and Betrán, 1999; Witcombe and Hash, 2000; Collard and Mackill, 2008; Ribaut et al., 2010; Hayes et al., 2013; Varshney et al., 2013; Jiang, 2015; Gokidi et al., 2016; Xu et al., 2017b). All

these strategies of molecular selection can be defined as the use of molecular genetic markers, in combination with information of linkage maps and sequenced genomes, to improve the desired traits in plants based on genetic analysis. Among all these schemes, genomic selection is the most popular and widespread method. A relatively new direction, but already a very active area of research in plant and animal breeding – genomic selection, also called Genome-Wide Selection, opens up new exciting prospects for the development of molecular selection for crop improvement (Hayes et al., 2013; Varshney et al., 2013).

MAS-Based Approaches

Several advanced molecular breeding approaches are used in the creation of crop variety, such as marker-assisted backcrossing (MABC), marker-assisted gene pyramiding, marker-assisted recurrent selection (MARS), and genomic selection (GS). These approaches can help accelerate breeding processes with the early and direct selection of desirable individual plants in the DNA level which are resulting time and resource savings.

Marker-Assisted Backcross Selection

The backcross method has been widely used in conventional breeding since the beginning of the last century for the

FIGURE 3 | The main procedures of marker-assisted selection technology.
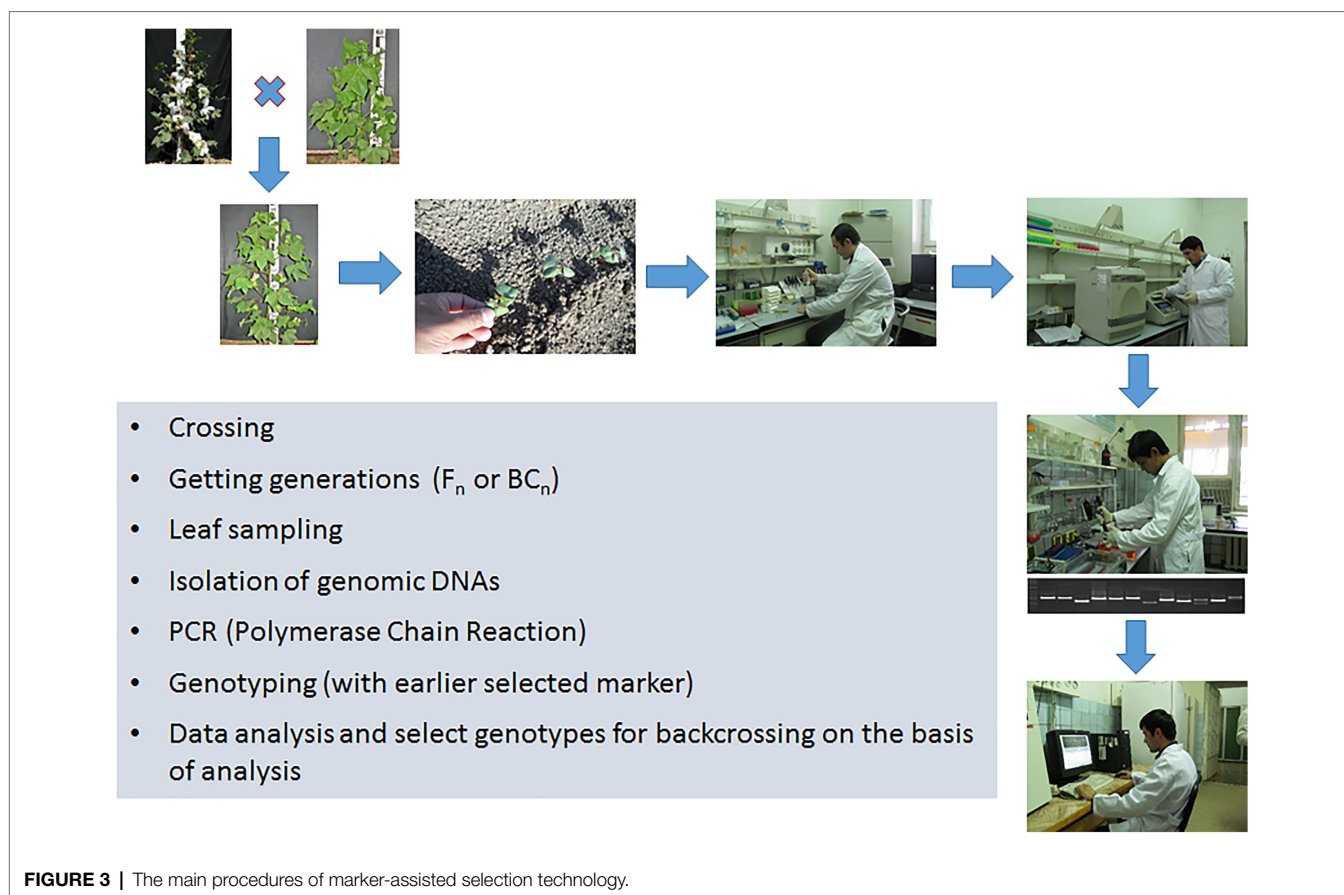
introgression of one or more genes from a donor to an elite variety (Collard and Mackill, 2008). However, the use of DNA markers in backcross programs in combination with phenotypic selection significantly accelerates the production of breeding material (Frisch et al., 1999; Xu et al., 2017b).

Marker-assisted backcross selection (MABS) is one of the simplest and most promising approaches of MAS technology (Acquaah, 2012). The main goal of MABS is to apply markers to select for target QTL, minimize the length of the donor genome segment containing a target QTL, and accelerate the recovery of the recurrent parent genome (**Figure 4**). According to Holland (2004), the method has three main levels of breeding, in which markers may be used in backcross selection (Holland, 2004): (i) Foreground selection or selecting of target loci, (ii) recombinant selection or selecting backcross progeny with the target gene, and (iii) background selection or selecting backcross progeny with background markers. These three levels are used in one or another combination in backcross breeding programs for gene introgression.

It is known that for the introgression of one dominant gene it is necessary to carry out at least six backcrosses (Acquaah, 2012), so that in the end the content of the genome of the recurrent parent would be 99% only in theory. With conventional backcrossing, it takes a minimum of five to six generations. The use of DNA markers allows reducing the number of

required backcrossing to four and reducing the amount of genetic material transferred with the "target" locus.

MABS has been applied in several important crops, including maize, rice, wheat, barley, common bean, soybean, pear millet, potato, and tomato. For example, the integration of the *Bt* transgene into diverse corn genetic backgrounds has been obtained by using this approach of MAS in maize (Gassmann et al., 2011). MABS strategy was used for rice improvement (Bishwas et al., 2016; Das et al., 2017). According to Rambabu et al. (2015), leaf blast resistance gene was introgressed into variety "Swarna" (Rambabu et al., 2015). In addition, MABS has been used for the effective introgression of favorable alleles from the wild germplasm into elite cultivars; MABS has been used in other crops.

In turn, Li et al. (2013a) based on the results of such research, as well as using the MABS approach, were able to successfully introgress wilt resistance QTL from *G. barbadense* to *G. hirsutum* (Li et al., 2013a). Also, MABS has been initiated in Uzbekistan to improve important fiber traits of cotton. Association mapping has been applied for the identification of QTLs associated with fiber quality, and the selection of donor lines with superior quality using diverse sets of Uzbek cotton germplasm (Abdurakhmonov et al., 2008, 2009). Twenty-six germplasm accessions, as donor lines and more than 10 varieties, as the recipient parents were selected for QTL mobilization through MABS.

FIGURE 4 | The scheme of marker-assisted backcross selection.
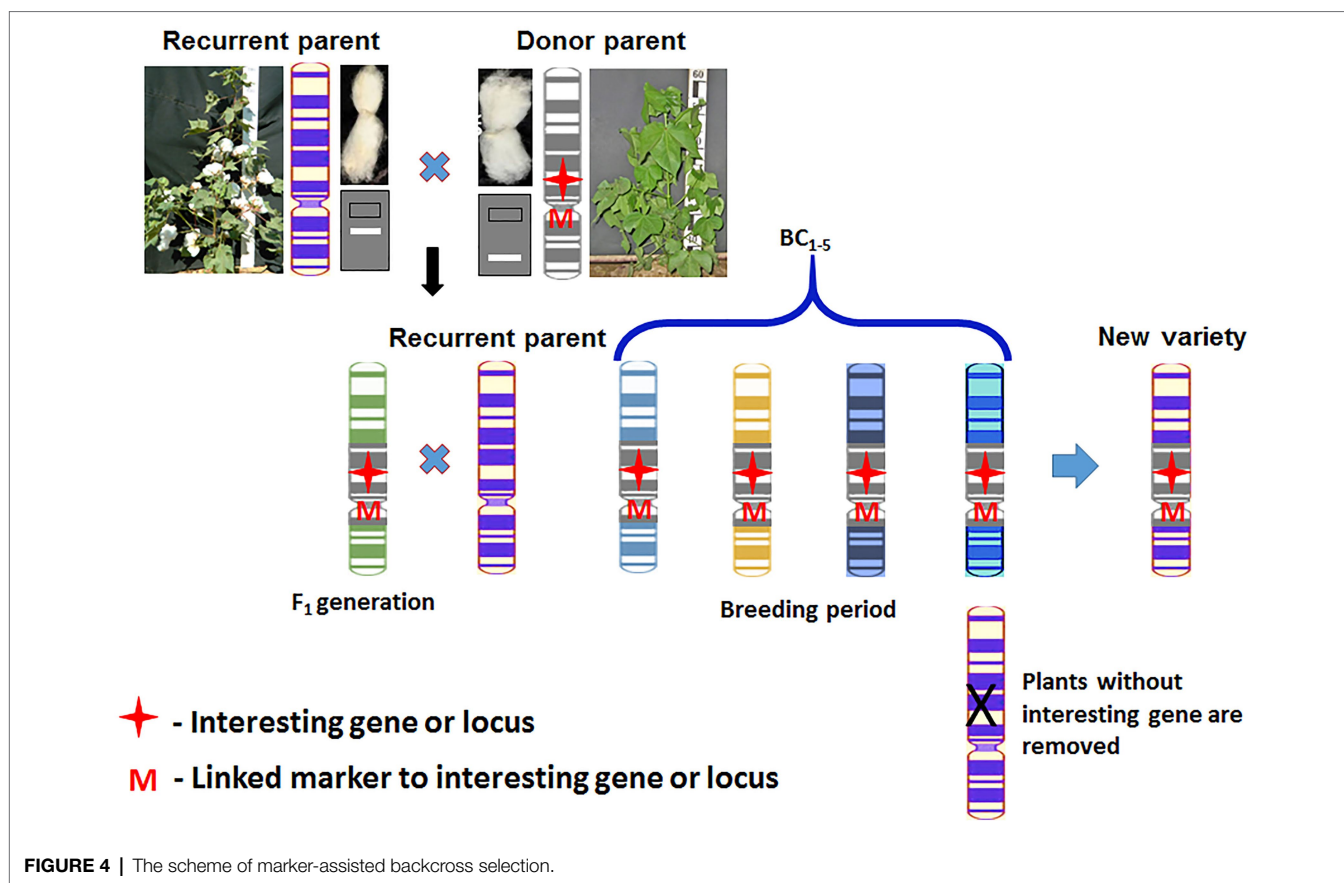
As a result, new varieties "Ravnaq-1" and "Ravnaq-2" were developed (Darmanov et al., 2015). Both varieties possess higher fiber strength and improved length. "Ravnaq-1" has improved fiber strength (37 g/tex) and staple length (38 mm) compared to its recurrent parent "Andijan-35" which has 32 g/tex fiber strength and 35 mm staple length.

Marker-Assisted Recurrent Selection

According to Ribaut et al. (2010), the goal of the marker-assisted recurrent selection (MARS) scheme is the identification and selection of several regions on the genome engaged in the expression of complex quantitative traits to assemble within a single cross or across related populations (Ribaut et al., 2010). Utilization of markers to pyramid for multiple genes or QTLs is more difficult and in this situation, the recurrent selection is a potential approach for the improvement of polygenic traits (Ceballos et al., 2015). As noted by Jiang (2013), the MARS strategy of MAS selection performs genotypic selection and intercrossing in the same crop season for one breeding cycle (Jiang, 2013).

MARS program has been successfully applied to improve important agronomic traits in maize. As described by Beyene et al. (2016), this strategy showed excellent results than the conventional selection in the studies to develop improved drought tolerance germplasm (Beyene et al., 2016). Recently, another research group under the maize improvement program

has improved grain yield based on a biparental population using the SNP marker system in MARS (Bankole et al., 2017). However, very little information is available on the practical application of the MAS strategy to improve cotton. According to Yi et al. (2004), cotton bollworm (*Helicoverpa armigera*) resistance of *G. hirsutum* has been increased using MARS (Yi et al., 2004).

Marker-Assisted Gene Pyramiding

To create new varieties, using traditional breeding techniques is time consuming, labor-intensive, and can be costly (Moreau et al., 2004). In particular, it is difficult to develop a large number of populations, advancing the hybrids up to F9/F10 generations, the difficulty of the selection processes due to the negative effects of environmental factors on the appearance of morpho-biological traits. In most cases, it is necessary to wait until the last stage of plant ontogenesis to make a selection according to the trait of interest and complexity of combining significant genes in a single genotype (Gupta and Varshney, 2005). In such cases, the selection process has been proven to last for 20–25 years in practice.

Watson and Singh (1952) were first introduced Gene pyramiding conception. According to Allard (1999), pyramiding multiple genes is achieved by crossing parental lines with complementary desirable genes and selecting the desired

recombinants from among the progeny population. MAS-based gene pyramiding (MAGP) method is combining at the same time multiple genes/QTLs together into a single genotype using several trait-associated DNA markers. Since the development of this technology, a unique chance has appeared not only to speed up the selection process and reduce costs but also to direct efforts to create varieties with multilateral resistance through gene pyramiding technology. Using traditional breeding methods, it is extremely difficult or impossible to implement this process in the early generations (Sheikh et al., 2017). Nowadays, this technology is considered also as the main acceptable strategy for developing new varieties of crops.

MAGP application had been reported in cotton, wheat, rice, tomato and pepper, etc. The most widespread application of the gene pyramiding method has been for combining multiple disease resistance genes in order to develop durable disease resistance (Ashkani et al., 2015). In recent years, reports have appeared on the application of this technology in cotton. Researchers used the MAGP strategy to combine the major QTL traits of fiber quality and wilt resistance from different donors into one genotype in several commercial cotton varieties in order not only to expand the genetic base of the developed MAS lines but also to ensure their genetic stability. Guo et al. (2005) successfully introduced the effect of pyramiding QTLs for strong fiber strength and transgene cryIA in cotton (Guo et al., 2005). They developed insect-resistant and high-yielding new cotton variety with superior fiber quality.

Genomic Selection

Full-genomic sequencing has become available thanks to the development of modern sequencing platforms – next-generation sequencing (NGS) technology. The advantage is the analysis of a large number of markers and the possibility of identifying new genetic variants. NGS technology has made it possible to speed up and cheapen the determination of the complete genome sequence of organisms (Pareek et al., 2011; Zhang et al., 2011). It becomes possible to simultaneously evaluate thousands of genes in organisms, tissues, and cells (sequencing of transcriptomes) and analyze the regulation of their activity. To date, about 300 genomes of different plant species have been sequenced, and this number is increasing each year. The introduction of methods of high-performance genotyping of agricultural organisms opened the way for the application of a new method of selection.

Genomic selection (GS) is considered a novel strategy of MAS for plant breeding, based on the analysis of a large number of DNA markers evenly distributed throughout the genome. The term "genomic selection" was first introduced by Haley and Visscher in 1998 (Haley and Visscher, 1998). Three years later, Meuwissen et al. (2001) developed and presented a GS methodology, as progress in MAS technology for the study of quantitative traits (Meuwissen et al., 2001). In plant breeding, GS has become more productive due to a large number of single-nucleotide polymorphisms (SNPs) detected by sequencing the crop genome. Currently, the full-genome SNP chips have been developed for several types of crops for automatic analysis of DNA polymorphism. According to Meuwissen et al. (2001), the genomic selection also proposes the prediction model based on the genotypic and phenotypic data of the reference population. The reference population is used to receive genomic estimated breeding values (GEBVs) for all individuals of a breeding population based on their genomic profile.

Especially with the advent of the GS method, significant changes have occurred in the evaluation of breeding value in world livestock breeding (Jonas and de Koning, 2015). The accumulation of fundamental knowledge in these areas allowed sequencing the genomes of the main types of agricultural animals – cattle, pigs, and sheep, and carrying out the genotyping of animals by DNA markers. It should be noted that the greatest success in the practical application of GS was noted for Holstein dairy cattle. Naturally, the development of GS methods brought a lot of success also in plant breeding (Wang et al., 2018). An example is a recently published work of Heffner et al. (2010), in which the genetic gain of the GS method in maize breeding is higher than that of its pedigree MAS technology (Heffner et al., 2010). Xu et al. (2014) also confirm that the genomic prediction served to select potential hybrids from recombinant inbred lines (RIL) of rice (Xu et al., 2014). Daetwyler et al. (2014) in their research on wheat rust resistance, applied genomic best linear unbiased prediction (GBLUP), and a Bayesian regression method to predict resistance to leaf, stem, and strip rust (Daetwyler et al., 2014). Gapare et al. (2018) accommodating genotype x environment interaction (GxE) based on a population of 215 breeding lines of tetraploid cotton *G. hirsutum* identified potential breeding lines for fiber length and strength (Gapare et al., 2018). Hulse-Kemp et al. (2015) have developed the CottonSNP63K intraspecific SNPs for use within the Upland cotton (*Gossypium hirsutum* L.) cultivars and interspecific SNPs for use with crosses several cotton species with *G. hirsutum* L. (Hulse-Kemp et al., 2015). Wang et al. (2016) have used CottonSNP70K Chip to detect SNP in four salt tolerance and four salt-sensitive cotton varieties. SNP variation of the same seedling pre- and after-salt stress in different varieties was screened and polymorphic SNP and SNP related to salt tolerance were obtained (Wang et al., 2016). As well, Cai et al. (2017) developed a CottonSNP80K array that plays an important role in germplasm genotyping, variety verification, functional genomics studies, and molecular breeding in cotton by selecting from the re-sequencing data of 100 cotton cultivars (Cai et al., 2017). Above mentioned, SNP arrays are valuable new resources for molecular breeding approaches, such as marker-assisted selection (MAS), marker-assisted gene pyramiding (MAGP), and genomic selection (GS).

Thus, GS is a powerful tool for use in the molecular breeding of crops and is more efficient than MAS for improving complex traits with low heritability (Jannink et al., 2010). This method allows breeders to select new breeding material based on genetic potential. That is, the best hybrid of the breeding population can be selected only on the basis of a simple DNA test instead of waiting for 2–3 years of field data. GS also improves the

options for selecting several traits at the same time. The major obstacle for the wide dissemination of this method in the selection of crops is the presence of one of the key stages to analyze SNPs, more precisely the high cost of genotyping.

FUTURE PERSPECTIVES

Genetic diversity is essential to the genetic progress of cotton breeding. The level of genetic diversity is low in *G. hirsutum*, especially among agriculturally elite types, as revealed in many previous studies (Gutiérrez et al., 2002; Abdurakhmonov et al., 2009, 2012). Chromosome substitution lines (CSL) have been used to overcome the problem of interspecific introgression using conventional breeding methods (Saha et al., 2013, 2017; Jenkins et al., 2017a,b). Chinese scientists used a different approach than CSL to introgress alleles from *G. barbadense* into *G. hirsutum,* where they have developed chromosome segment introgression lines using TM-1, *G. hirsutum*, as the recipient and a high-quality *G. barbadense* line "Hai7124" as the donor (Wang et al., 2012). After four cycles of the MAS breeding program using markers specific to the donor line Hai7124, they developed 174 lines containing 298 introgressed segments with 86 lines having single introgressed segments. "The total length of introgressed segments covered 2948.7 cM with an average segment length of 16.7 cM and represented 83.3% of the tetraploid cotton genome" (Wang et al., 2012). These lines were used in the genetic dissection of the complex fiber quality traits, with 43 additive QTLs and six epistatic QTLs associated with fiber quality traits in a molecular map (Wang et al., 2012). Jenkins et al. (2018) crossed 18 *G. barbadense* CSL to three Upland cotton cultivars and developed a random mated population for the cotton breeding program (Jenkins et al., 2018). After five cycles of random mating using a mixer of pollens from individual CSL followed by one generation of self-pollination to increase the seed supply to develop the random mated population with improved genetic diversity. They used 139 *G. barbadense* chromosome-specific SSR markers to assess a random sample of 96 plants for introgression. They detected 121 of 139 marker loci among the 96 plants. The number of *G. barbadense* alleles ranged from 10 to 28 in each individual plant. They also discovered that the individual plants among the 96 plants had marker loci from 6 to 14 different chromosomes or chromosome arms. However, results on the identity by descent showed little relatedness among plants and no population structure was indicated by a heat map. Using CSL, they were able to develop a mostly Upland random mated population with considerable introgression of *G. barbadense* alleles which would be useful for the cotton breeding program.

Recently, very cost-effective high-throughput sequencing technologies open up a new paradigm in the molecular cotton breeding programs using RNA-seq technologies. High-throughput sequencing technologies are used for RNA-seq experiments to generate cDNA sequences derived from the total RNA molecules followed by library construction and massively parallel deep sequencing to quantify the abundance level of relative changes of the individual transcripts at a specific stage of development or under specific treatment conditions. The application of the RNA-seq tool to associate changes in gene expression from high-throughput results of transcriptomics with low background noise to associate with important traits shows great potential in a future cotton breeding program. Recently, Naoumkina et al. (2019) used RNA-seq analysis in a GWA study in a MAGIC population. RNA-seq analysis of the longest and shortest fiber length RILs from D-11ref and D-11alt populations detected 949 significantly differentially expressed genes (DEGs; Naoumkina et al., 2019). Gene set enrichment analysis identified that different functional categories of genes were overrepresented during fiber elongation between four selected RILs. They discovered that 12 genes possessing non-synonymous SNPs were significantly associated with the fiber length. They also detected that in close proximity to fiber length QTL on chromosome D11, an auxin-responsive GH3 gene with a significantly downregulated expression level in one of the longest fiber length RILs suggesting that it could play a role in the regulation of fiber-cell elongation.

Transcriptome-wide association studies (TWAS) are a powerful strategy that integrate GWAS and gene expression datasets for identification of gene-trait associations (Wainberg et al., 2019; Bhattacharya et al., 2021). Recently, Li et al. (2020b) performed a fiber transcriptome analysis by sequencing of natural *G. hirsutum* population with 251 accessions and identified 15,330 expression quantitative trait loci (eQTL) associated with 9,282 genes (Li et al., 2020b). Analysis of eQTL and GWAS data uncovered molecular regulation of cotton fiber development and revealed the genetic basis of cell wall synthesis during fiber-cell elongation.

The results of above mentioned studies using the RNA-seq and TWAS tools provided an insight into the molecular aspects of genetic variation and fiber development as well recommended the potential sources for MAS and genetic manipulation technologies, such as CRISPR, in the future cotton improvement programs.

CONCLUSION

Cotton is the most important source of natural fiber worldwide. The negative impacts of the natural environment like water scarcity, soil salinization, diverse insect pests, and diseases cause serious damage to cotton productivity and fiber quality. It is imperative to create new cultivars with high yield, superior fiber quality, and resistance to the biotic and abiotic stresses through the use of diverse germplasm resources including wild cotton species and utilize high-throughput technologies. The use of genetic diversity of cotton species and populations in genetic mapping of quantitative traits allows to identify genome-wide informative DNA markers or genes and to determine potential breeding donors with desirable traits. Marker-assisted selection-based molecular breeding approaches could be helpful in pyramiding multiple genes/QTLs linked with resistance, quality, and yield components into a single genotype (Dormatey et al., 2020).

Progress in this area will be further increased by taking the information generated through "omics" studies (Boopathi, 2020). Furthermore, as stated above, involving innovative approaches, combining diverse resources and enhance the capacities for increasing marker-assisted selection in cotton ultimately result in developing cotton cultivars with improved quality and productivity.

AUTHOR CONTRIBUTIONS

FK, OT, and DE wrote the first draft of manuscript. SS, JY, and IA provided the critical comments and edited the manuscript. All authors listed have made substantial, direct, and intellectual contributions to the work, and approved the manuscript.

REFERENCES

Abdalla, A. M., Reddy, O. U. K., El-Zik, K. M., and Pepper, A. E. (2001). Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor. Appl. Genet.* 102, 222–229. doi: 10.1007/s001220051639

Abdelraheem, A., Elassbli, H., Zhu, Y., Kuraparthy, V., Hinze, L., Wedegaertner, T., et al. (2020). A genome-wide association study uncovers consistent quantitative trait loci for resistance to Verticillium wilt and Fusarium wilt race 4 in the US upland cotton. *Theor. Appl. Genet.* 133, 563–577. doi: 10.1007/s00122-019-03487-x

Abdelraheem, A., Esmaeili, N., O'Connell, M., and Zhang, J. (2019). Progress and perspective on drought and salt stress tolerance in cotton. *Ind. Crop. Prod.* 130, 118–129. doi: 10.1016/j.indcrop.2018.12.070

Abdelraheem, A., Thyssen, G. N., Fang, D. D., Jenkins, J. N., McCarty, J. C., Wedegaertner, T., et al. (2021). GWAS reveals consistent QTL for drought and salt tolerance in a MAGIC population of 550 lines derived from intermating of 11 upland cotton (*Gossypium hirsutum*) parents. *Mol. Gen. Genomics.* 296, 119–129. doi: 10.1007/s00438-020-01733-2

Abdurakhmonov, I. Y., Abdullaev, A. A., Saha, S., Buriev, Z. T., Arslanov, D., Kuryazov, Z., et al. (2005). Simple sequence repeat marker associated with a natural leaf defoliation trait in tetraploid cotton. *J. Hered.* 96, 644–653. doi: 10.1093/jhered/esi097

Abdurakhmonov, I. Y., Buriev, Z. T., Shermatov, S. E., Abdullaev, A. A., Urmonov, K., Kushanov, F., et al. (2012). "Chapter 16: Genetic diversity in *Gossypium* genus," in *Genetic Diversity in Plants*. ed. M. Çalışkan (Rijeka: IntechOpen), 313–338.

Abdurakhmonov, I. Y., Kohel, R. J., Yu, J. Z., Pepper, A. E., Abdullaev, A. A., Kushanov, F. N., et al. (2008). Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics* 92, 478–487. doi: 10.1016/j.ygeno.2008.07.013

Abdurakhmonov, I. Y., Kushanov, F. N., Djaniqulov, F., Buriev, Z. T., Pepper, A. E., Fayzieva, N., et al. (2007). The role of induced mutation in conversion of photoperiod dependence in cotton. *J. Hered.* 98, 258–266. doi: 10.1093/jhered/esm007

Abdurakhmonov, I. Y., Saha, S., Jenkins, J. N., Buriev, Z. T., Shermatov, S. E., Scheffler, B. E., et al. (2009). Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica* 136, 401–417. doi: 10.1007/s10709-008-9337-8

Acquaah, G. (2012). "Chapter 22: Marker assisted selection," in *Principles of Plant Genetics Breeding. 2nd Edn*. ed. G. Acquaah (USA: Wiley-Blackwell, John Wiley & Sons, Ltd Publication), 424–435.

Adams, K. L., Percifield, R., and Wendel, J. F. (2004). Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* 168, 2217–2226. doi: 10.1534/genetics.104.033522

Ahmad, F., Akram, A., Farman, K., Abbas, T., Bibi, A., Khalid, S., et al. (2017). Molecular markers and marker assisted plant breeding: current status and their applications in agricultural development. *J. Environ. Agric. Sci.* 11, 35–50.

Allard, R. W. (1999). *Principles of Plant Breeding. 2nd Edn* New York, Toronto: John Wiley & Sons, Inc 264 p.

Ashkani, S., Rafii, M. Y., Shabanimofrad, M., Miah, G., Sahebi, M., Azizi, P., et al. (2015). Molecular breeding strategy and challenges towards improvement of blast disease resistance in rice crop. *Front. Plant Sci.* 6:886. doi: 10.3389/fpls.2015.00886

Asíns, M. J. (2002). Present and future of quantitative trait locus analysis in plant breeding. *Plant Breed.* 121, 281–291. doi: 10.1046/j.1439-0523.2002.730285.x

Bankole, F., Menkir, A., Olaoye, G., Crossa, J., Hearne, S., Unachukwu, N., et al. (2017). Genetic gains in yield and yield related traits under drought stress and favorable environments in a maize population improved using marker assisted recurrent selection. *Front. Plant Sci.* 8:808. doi: 10.3389/fpls.2017.00808

Bardak, A., Çelik, S., Erdoğan, O., Ekinci, R., and Dumlupinar, Z. (2021). Association mapping of Verticillium wilt disease in a worldwide collection of cotton (*Gossypium hirsutum* L.). *Plants.* 10:306. doi: 10.3390/plants10020306

Beyene, Y., Semagn, K., Crossa, J., Mugo, S., Atlin, G. N., Tarekegne, A., et al. (2016). Improving maize grain yield under drought stress and non-stress environments in sub-saharan africa using marker-assisted recurrent selection. *Crop Sci.* 56, 344–353. doi: 10.2135/cropsci2015.02.0135

Bhattacharya, A., Li, Y., and Love, M. I. (2021). MOSTWAS: multi-omic strategies for transcriptome-wide association studies. *PLoS Genet.* 17:e1009398. doi: 10.1371/journal.pgen.1009398

Bishwas, N., Sharma, M., Hasan, A., Akhtar, S., and Sharma, N. (2016). Improvement of rice crop by marker-assisted backcross method. *Int.Res. J. Eng. Technol.* 3, 1851–1858.

Blenda, A., Fang, D. D., Rami, J. F., Garsmeur, O., Luo, F., and Lacape4 J-M., (2012). A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. *PLoS One* 7:e45739. doi: 10.1371/journal.pone.0045739

Bolek, Y., El-Zik, K. M., Pepper, A. E., Bell, A. A., Magill, C. W., Thaxton, P. M., et al. (2005). Mapping of Verticillium wilt resistance genes in cotton. *Plant Sci.* 168, 1581–1590. doi: 10.1016/j.plantsci.2005.02.008

Boopathi, N. Manikanda (2020). *Genetic Mapping and Marker Assisted Selection: Basics, Practice and Benefits*. Singapore: Springer

Brubaker, C. L., Paterson, A. H., and Wendel, J. F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42, 184–203. doi: 10.1139/g98-118

Bruford, M., Bradley, D., and Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* 4, 900–910. doi: 10.1038/nrg1203

Cai, C., Zhu, G., Zhang, T., and Guo, W. (2017). High-density 80 K SNP array is a powerful tool for genotyping *G. hirsutum* accessions and genome analysis. *BMC Genomics* 18:654. doi: 10.1186/s12864-017-4062-2

Ceballos, H., Kawuki, R. S., Gracen, V. E., Yencho, G. C., and Hershey, C. H. (2015). Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. *Theor. Appl. Genet.* 128, 1647–1667. doi: 10.1007/s00122-015-2555-4

Choudhary, K., Choudhary, O. P., and Shekhawat, N. S. (2008). Marker assisted selection: A novel approach for crop improvement. *Am. Eurasian J. Agron.* 1, 26–30.

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos.l Trans. R Soc B: Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170

Daetwyler, H. D., Bansal, U. K., Bariana, H. S., Hayden, M. J., and Hayes, B. J. (2014). Genomic prediction for rust resistance in diverse wheat landraces. *Theor. Appl. Genet.* 127, 1795–1803. doi: 10.1007/s00122-014-2341-8

Darmanov, M.M., Makamov, A.K., Kushanov, F.N., Buriev, Z.T., and Abdurakhmonov, I.Y. (2015). "Marker-assisted selection for cotton." in *Proceedings of the Tashkent international Innovation forum, Section Agriculture*, Uzbekistan, Tashkent, May 19-21, 260–267.

Das, G., Patra, J. K., and Baek, K. H. (2017). Insight into MAS: A molecular tool for development of stress resistant and quality of rice through gene stacking. *Front. Plant Sci.* 8:985. doi: 10.3389/fpls.2017.00985

Diouf, L., Magwanga, R. O., Gong, W., He, S., Pan, Z., Jia, Y. H., et al. (2018). QTL mapping of fiber quality and yield-related traits in an intra-specific upland cotton using genotype by sequencing (GBS). *Int. J. Mol. Sci.* 19:441. doi: 10.3390/ijms19020441

Dormatey, R., Sun, C., Ali, K., Coulter, J., Bi, Z., and Bai, J. P. (2020). Gene pyramiding for sustainable crop improvement against biotic and abiotic stresses. *Agronomy* 10:1255. doi: 10.3390/agronomy10091255

Du, W. S., Du, X. M., Ma, Z., and Y., (2004). Studies on SSR markers of resistance gene of Verticillium wilt in cotton. *J. Northwest Sci. Tech. Univ. Agric. Forest.* 32, 20–24.

Eagles, H. A., Bariana, H. S., Ogbonnaya, F. C., Rebetzke, G. J., Hollamby, G. J., Henry, R. J., et al. (2001). Implementation of markers in Australian wheat breeding. *Aust. J. Agric. Res.* 52, 1349–1356. doi: 10.1071/AR01067

Esbroeck, G. A., Bowman, D., May, O. L., and Calhoun, D. S. (1999). Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. *Crop Sci.* 39, 323–328. doi: 10.2135/cropsci1999.0011183X003900020003x

Fang, D. D., Jenkins, J. N., Deng, D. D., McCarty, J. C., Li, P., and Wu, J. (2014). Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in upland cotton (Gossypium hirsutum L.). *BMC Genomics* 15:397. doi: 10.1186/1471-2164-15-397

Frisch, M., Bohn, M., and Melchinger, A. E. (1999). Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci.* 39, 1295–1301. doi: 10.2135/cropsci1999.3951295x

Fryxell, P.A. (1979): *The Natural History of the Cotton Tribe (Malvaceae, Tribe Gossypieae).* USA, Texas A&M University Press: 245.

Fryxell, P. A. (1992). A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheedea* 2, 108–165.

Gao, Y. Q., Nie, Y. C., and Zhang, X. L. (2003). QTL mapping of genes resistant to Verticillium wilt in cotton. *Cotton Sci.* 15, 73–78.

Gapare, W., Conaty, W., Zhu, Q. H., Liu, S., Stiller, W., Llewellyn, D., et al. (2017). Genome-wide association study of yield components and fibre quality traits in a cotton germplasm diversity panel. *Euphytica* 213:66. doi: 10.1007/s10681-017-1855-y

Gapare, W., Liu, S., and Conaty, W., Q.-H. Zhu, Gillespie, V., Llewellyn, D., Stiller, W., and Wilson, I. (2018). Historical datasets support genomic selection models for the prediction of cotton fiber quality phenotypes across multiple environments, *G3*, 8:1721–1732, doi: 10.1534/g3.118.200140.

Gassmann, A. J., Petzold-Maxwell, J. L., Keweshan, R. S., and Dunbar, M. W. (2011). Field-evolved resistance to Bt maize by Western corn rootworm. *PLoS One* 6:e22629. doi: 10.1371/journal.pone.0022629

Glazier, A. M., Nadeau, J. H., and Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science* 298, 2345–2349. doi: 10.1126/science.1076641

Goicoechea, N. (2009). To what extent are soil amendments useful to control Verticillium wilt?. *Pest Manag. Sci.* 65, 831–839. doi: 10.1002/ps.1774

Gokidi, Y., Bhanu, A. N., and Singh, M. N. (2016). Marker assisted recurrent selection: An overview. *Adv Life Sci.* 5, 6493–6499.

Goldstein, D. B., and Weale, M. E. (2001). Linkage disequilibrium holds the key. *Curr. Biol.* 11, R576–R579. doi: 10.1016/S0960-9822(01)00348-7

Griffith, F. L., and Crowfoot, G. M. (1934). On the early use of cotton in the Nile Valley. *J. Egypt. Archaeol.* 20, 5–12. doi: 10.1177/030751333402000102

Grodzicker, T., Williams, J., and Sharp, P. (1974). Physical mapping of temperature-sensitive mutations of adenovirus. *Cold Spring Harb. Symp. Quant. Biol.* 34, 439–446.

Guo, Y., McCarty, J. C., Jenkins, J. N., and Saha, S. (2008). QTLs for node of first fruiting branch in a cross of an upland cotton, *Gossypium hirsutum* L., cultivar with primitive accession Texas 701. *Euphytica* 163, 113–122. doi: 10.1007/s10681-007-9613-1

Guo, W. Z., Zhang, T. Z., Zhu, X. F., and Pan, J. J. (2005). Modified backcross pyramiding breeding with molecular marker-assisted selection and its applications in cotton. *Acta Agron. Sin.* 31, 963–970.

Gupta, P. K., and Varshney, R. K. (2005). "Cereal genomics: An overview," in *Cereal Genomics*. eds. P. K. Gupta and R. K. Varshney (Dordrecht: Springer Netherlands), 1–18.

Gutiérrez, O. A., Basu, S., Saha, S., Jenkins, J. N., Shoemaker, D. B., Cheatham, C. L., et al. (2002). Genetic distance among selected cotton genotypes and its relationship with F2 performance. *Crop Sci.* 42, 1841–1847. doi: 10.2135/cropsci2002.1841

Haley, C. S., and Visscher, P. M. (1998). Strategies to utilize marker-quantitative trait loci associations. *J. Dairy Sci.* 81, 85–97. doi: 10.3168/jds.S0022-0302(98)70157-2

Han, J., Han, D., Guo, Y., Yan, H., Wei, Z., Tian, Y., et al. (2019). QTL mapping pod dehiscence resistance in soybean (Glycine max L. Merr.) using specifclocus amplied fragment sequencing. *Theor. Appl. Genet.* 132, 2253–2272. doi: 10.1007/s00122-019-03352-x

Hawkins, J. S., Pleasants, J., and Wendel, J. F. (2005). Identification of AFLP markers that discriminate between cultivated cotton and the Hawaiian island endemic, *Gossypium tomentosum* Nuttall ex Seeman. *Genet. Resour. Crop. Evol.* 52, 1069–1078. doi: 10.1007/s10722-004-6115-z

Hayes, B., Cogan, N., Pembleton, L., Michael, G., Junping, W., German, S., et al. (2013). Prospects for genomic selection in forage plant species. *Plant Breed.* 132, 133–143. doi: 10.1111/pbr.12037

Heffner, E. L., Lorenz, A. J., Jannink, J.-L., and Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 50, 1681–1690. doi: 10.2135/cropsci2009.11.0662

Hendrix, B., and Stewart, J. M. (2005). Estimation of the nuclear DNA content of *Gossypium* species. *Ann. Bot.* 95, 789–797. doi: 10.1093/aob/mci078

Holland, J.B. (2004) "Implementation of molecular markers for quantitative traits in breeding programs – challenges and opportunities." in *Proceedings of the 4th International Crop Science Congress*, 26 Sep – 1 Oct 2004, Brisbane, Australia. 1–13.

Hospital, F. (2009). "Chapter 10: Marker-assisted back-cross breeding: a case-study in genotype-building theory," in *Quantitative Genetics, Genomics and Plant Breeding.* ed. M. S. Kang (Wallingford, UK: CABI Publishing), 135–141.

Hulse-Kemp, A. M., Lemm, J., Plieske, J., Ashrafi, H., Buyyarapu, R., Fang, D. D., et al. (2015). Development of a 63K SNP array for cotton and high-density mapping of intra-and inter-specific populations of *Gossypium* spp. *G3* 5, 1187–1209. doi: 10.1534/g3.115.018416

ICAC (2021) *DataPortal*. International Cotton Advisory Committee, ICAC, Washington DC. https://icac.org

Iqbal, M. J., Reddy, O. U. K., El-Zik, K. M., and Pepper, A. E. (2001). A genetic bottleneck in the "evolution under domestication" of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theor. Appl. Genet.* 103, 547–554. doi: 10.1007/PL00002908

Islam, M. S., Thyssen, G. N., Jenkins, J. N., Zeng, L., Delhom, C. D., McCarty, J. C., et al. (2016). A MAGIC population-based genome-wide association study reveals functional association of GhRBB1_A07 gene with superior fiber quality in cotton. *BMC Genomics* 17, 1–17. doi: 10.1186/s12864-016-3249-2

Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: From theory to practice. *Briefings Funct, Genomics Proteomics* 9, 166–177. doi: 10.1093/bfgp/elq001

Jannink, J. L., and Walsh, B. (2002). "Chapter 5: Association mapping in plant populations," in *Quantitative Genetics, Genomics and Plant Breeding.* ed. M. S. Kang (Wallingford, UK: CABI Publishing), 59–68.

Jansen, R. C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics* 135, 205–211. doi: 10.1093/genetics/135.1.205

Jansen, R. C. (1994). Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics* 138, 871–881. doi: 10.1093/genetics/138.3.871

Jenkins, J. N., McCarty, J. C., Campbell, T. B., Hayes, R. W., Wu, J., Saha, S., et al. (2017a). Genetic effects of chromosomes 1, 4, and 18 from three tetraploid *Gossypium* species in topcrosses with five elite cultivars. *Crop Sci.* 57, 1338–1346.

Jenkins, J. N., McCarty, J. C., Campbell, T. B., Hayes, R. W., Wu, J., Saha, S., et al. (2017b). Genotypic comparisons of chromosomes 01, 04, and 18 from three tetraploid species of *Gossypium* in topcrosses with five elite cultivars of *G. hirsutum* L. *Euphytica* 213:107. doi: 10.1007/s10681-017-1895-3

Jenkins, J. N., McCarty, J. C., Deng, D., Geng, L., Hayes, R. W., Jones, D. C., et al. (2018). Introgression of *Gossypium barbadense* L. into upland cotton germplasm RMBUP-C4S1. *Euphytica* 214:118. doi: 10.1007/s10681-018-2200-9

Jia, Y. H., Sun, J. L., Wang, X. W., Zhou, Z. L., Pan, Z. E., He, S. P., et al. (2014). Molecular diversity and association analysis of drought and salt tolerance in *Gossypium hirsutum* L. germplasm. *J. Integr. Agric.* 13, 1845–1853. doi: 10.1016/S2095-3119(13)60668-1

Jiang, G.-L. (2013). "Chapter 3: Molecular markers and marker-assisted breeding in plants," in *Plant Breeding from Laboratories to Fields.* ed. S. B. Andersen (Rijeka: IntechOpen Publisher).

Jiang, G-L. (2015): "Chapter 15: Molecular marker-assisted breeding: a plant breeder's review," in *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools.* eds. J. M., Al-Khayri, S. M., Jain, and D. V., Johnson (Springer International Publishing), 431–472.

Jiang, F., Zhao, J., Zhou, L., Guo, W., and Zhang, T. (2009). Molecular mapping of Verticillium wilt resistance QTL clustered on chromosomes D7 and D9 in upland cotton. *Sci China C Life Sci.* 52, 872–884. doi: 10.1007/s11427-009-0110-8

Jonas, E., and de Koning, D. J. (2015). Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. *Front. Genet.* 6:49. doi: 10.3389/fgene.2015.00049

Kloth, R. H. (1993). New evidence relating the pilose allele and micronaire reading in cotton. *Crop Sci.* 33, 683–687. doi: 10.2135/cropsci1993.0011183X003300040007x

Kloth, R. H. (1995). Quantitative trait loci affecting cotton fiber are linked to the t_1 locus in upland cotton. *Theor. Appl. Genet.* 91, 762–768. doi: 10.1007/BF00220956

Kumar, P., Gupta, V. K., Misra, A. K., Modi, D. R., and Pandey, B. K. (2009). Potential of molecular markers in plant biotechnology. *Plant Omics J.* 2, 141–162.

Kushanov, F. N., Buriev, Z. T., Shermatov, S. E., Turaev, O. S., Norov, T. M., Pepper, A. E., et al. (2017). QTL mapping for flowering-time and photoperiod insensitivity of cotton *Gossypium darwinii* Watt. *PLoS One* 12:e0186240. doi: 10.1371/journal.pone.0186240

Kushanov, F. N., Pepper, A. E., Yu, J. Z., Buriev, Z. T., Shermatov, S. E., Saha, S., et al. (2016). Development, genetic mapping and QTL association of cotton *PHYA*, *PHYB*, and *HY5*-specific CAPS and dCAPS markers. *BMC Genet.* 17, 1–11. doi: 10.1186/s12863-016-0448-4

Lacape, J. M., Nguyen, T. B., Courtois, B., Belot, J. L., Giband, M., Gourlot, J. P., et al. (2005). QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* x *Gossypium barbadense* backcross generations. *Crop Sci.* 45, 123–140. doi: 10.2135/cropsci2005.0123a

Lai, D., Li, H., Fan, S., Song, M., Pang, C., Wei, H., et al. (2011). Generation of ESTs for flowering gene discovery and SSR marker development in upland cotton. *PLoS One* 6:e28676. doi: 10.1371/journal.pone.0028676

Lander, E. S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199. doi: 10.1093/genetics/121.1.185

Lewis, C. M. (2002). Genetic association studies: design, analysis and interpretation. *Brief. Bioinform.* 3, 146–153. doi: 10.1093/bib/3.2.146

Li, B., Chen, L., Sun, W., Wu, D., Wang, M., Yu, Y., et al. (2020a). Phenomics-based GWAS analysis reveals the genetic architecture for drought resistance in cotton. *Plant Biotechnol. J.* 18, 2533–2544. doi: 10.1111/pbi.13431

Li, C.-Q., Liu, G.-S., Zhao, H.-H., Wang, L.-J., Zhang, X.-F., Liu, Y., et al. (2013a). Marker-assisted selection of Verticillium wilt resistance in progeny populations of upland cotton derived from mass selection-mass crossing. *Euphytica* 191, 469–480. doi: 10.1007/s10681-013-0917-z

Li, T., Ma, X., Li, N., Zhou, L., Liu, Z., Han, H., et al. (2017). Genome-wide association study discovered candidate genes of Verticillium wilt resistance in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 15, 1520–1532. doi: 10.1111/pbi.12734

Li, C, Wang, X, Dong, N, and Zhao, H., Xia, Z., Wang, X., Wang, Q., Converse R.L., Dong, N., Wang, R. (2013b): QTL analysis for early-maturing traits in cotton using two upland cotton (*Gossypium hirsutum* L.) crosses. *Breed. Sci.*, 63:154–163. doi: 10.1270/jsbbs.63.154.

Li, Z., Wang, P., You, C., Yu, J., Zhang, X., Yan, F., et al. (2020b). Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. *New Phytol.* 226, 1738–1752. doi: 10.1111/nph.16468

Li, L., Zhang, C., Huang, J., Liu, Q., Wei, H., Wang, H., et al. (2020c). Genomic analyses reveal the genetic basis of early maturity and identification of loci and candidate genes in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 19, 109–123. doi: 10.1111/pbi.13446

Liu, K., and Muse, S. V. (2005). PowerMaker: An integrated analysis environment for genetic maker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Liu, W., Song, C., Ren, Z., Zhang, Z., Pei, X., Liu, Y., et al. (2020). Genome-wide association study reveals the genetic basis of fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *BMC Plant Biol.* 20:395. doi: 10.1186/s12870-020-02611-0

Ma, Z., He, S., Wang, X., Sun, J., Zhang, Y., Zhang, G., et al. (2018). Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* 50, 803–813. doi: 10.1038/s41588-018-0119-7

Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annu. Rev. Genet.* 35, 303–339. doi: 10.1146/annurev.genet.35.102401.090633

Mauer, F. (1954): *Origin of Systematics of Cotton*. Tashkent, Izdaelstvo Akademii Nauk Uzbekskoy SSR: 384.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Mittal, N., and Dubey, A. K. (2009). Microsatellite markers- A new practice of DNA based markers in molecular genetics. *Pharmacogn. Rev.* 3, 235–246.

Moreau, L., Charcosset, A., and Gallais, A. (2004). Experimental evaluation of several cycles of marker-assisted selection in maize. *Euphytica* 137, 111–118. doi: 10.1023/B:EUPH.0000040508.01402.21

Muthukumaran, S. (2016). "Tree cotton (*G. arboreum*) in Babylonia," in *Cultural & Material Contacts in the Ancient Near East: Proceedings of the International Workshop*. eds. E. Foietta, C. Ferrandi, E. Quirico, F. Giusto, M. Mortarini and J. Bruno et al., December 1–2, 2014; Torino, Italy 98–105.

Mutschler, M., Tanksley, S., and Rick, C. M. (1987). Linkage maps of the tomato (Lycopersicon esculentum). *Rep. Tomato Genet. Cooperative* 37, 5–34.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Comertpay, G., Yıldız, M., Hatipoğlu, R., et al. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* 32, 261–285. doi: 10.1080/13102818.2017.1400401

Naoumkina, M., Thyssen, G. N., Fang, D. D., Jenkins, J. N., McCarty, J. C., and Florane, C. B. (2019). Genetic and transcriptomic dissection of the fiber length trait from a cotton (*Gossypium hirsutum* L.) MAGIC population. *BMC Genomics* 20, 1–14. doi: 10.1186/s12864-019-5427-5

Nicholson, G. E. (1960). The production, history, uses and relationships of cotton (*Gossypium* spp.) in Ethiopia. *Econ. Bot.* 14, 3–36. doi: 10.1007/BF02859364

Oraguzie, N.C., Wilcox, P.L., Rikkerink, E.H.A., and de Silva, H.N. (2007): Linkage disequilibrium. In: NC Oraguzie, EHA Rikkerink, SE Gardiner and SilvaHN De (ed.) *Association Mapping in Plants*. Springer New York, New York: 11–39.

Pareek, C. S., Smoczynski, R., and Tretyn, A. (2011). Sequencing technologies and genome sequencing. *J. Appl. Genet.* 52, 413–435. doi: 10.1007/s13353-011-0057-x

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427. doi: 10.1038/nature11798

Perkins, H. H., Ethridge, D. E., and Bragg, C. K. (1984). "Fiber," in *Cotton. Agronomy Momograph 24*. eds. R. J. Kohel and C. S. Lewis (Madison, WI: American Society of Agronomy), 605.

Preetha, S., and Raveendren, T. S. (2008). Molecular marker technology in cotton. *Biotechnol. Mol. Biol Rev.* 3, 32–45.

Price, A. H. (2006). Believe it or not, QTLs are accurate! *Trends Plant Sci.* 11, 213–216. doi: 10.1016/j.tplants.2006.03.006

Qin, H. D., and Zhang, T. Z. (2008). QTL mapping of leaf chlorophyll content and photosynthetic rates in cotton. *Cotton Sci.* 20, 394–398.

Rambabu, R., Diego, S., Kumar, S.V., and Kumar, A. (2015) "Marker assisted introgression of blast resistance genes Pi-1 and Pi-2 into popular indica variety Swarna." *International Rice Symposium*, Hyderabad, November 18–20, 2015, 586.

Rathore, K. S., Sunilkumar, G., and Campbell, L. M. (2006). "Cotton (*Gossypium hirsutum* L.)," in *Agrobacterium Protocols*. ed. K. Wang (Totowa, NJ: Humana Press), 267–280.

Reiter, R. (2001). "PCR-based marker systems," in *DNA-Based Markers in Plants*. eds. R. L. Phillips and I. K. Vasil (Dordrecht: Springer Netherlands), 9–29.

Ribaut, J.-M., and Betrán, J. (1999). Single large-scale marker-assisted selection (SLS-MAS). *Mol. Breed.* 5, 531–541. doi: 10.1023/A:1009631718036

Ribaut, J.-M., Vicente, M., and Delannay, X. (2010). Molecular breeding in developing countries: challenges and perspectives. *Curr. Opin. Plant Biol.* 13, 213–218. doi: 10.1016/j.pbi.2009.12.011

Saha, S., Wu, J., Jenkins, J. N., McCarty, J. C., Campbell, B. T., Hayes, R. W., et al. (2017). Tri-species shuffling of chromosomes to study the effects on fiber traits using chromosome substitution lines. *Crop Sci.* 57, 1211–1226. doi: 10.2135/cropsci2016.09.0729

Saha, S., Wu, J., Jenkins, J. N., McCarty, J. C., and Stelly, D. M. (2013). Interspecific chromosomal effects on agronomic traits in *Gossypium hirsutum* by AD analysis using intermated *G. barbadense* chromosome substitution lines. *Theor. Appl. Genet.* 126, 109–117. doi: 10.1007/s00122-012-1965-9

Said, J. I., Knapka, J. A., Song, M., and Zhang, J. (2015b). Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum* × *G. barbadense* populations. *Mol. Gen. Genomics.* 290, 1615–1625. doi: 10.1007/s00438-015-1021-y

Said, J. I., Lin, Z., Zhang, X., Song, M., and Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological

traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics* 14, 1–22. doi: 10.1186/1471-2164-14-776

Said, J. I., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D. D., et al. (2015a). A comparative meta-analysis of QTL between intraspecific *Gossypium hirsutum* and interspecific *G. hirsutum × G. barbadense* populations. *Mol. Gen. Genomics.* 290, 1003–1025. doi: 10.1007/s00438-014-0963-9

Saranga, Y., Menz, M., Jiang, C. X., Wright, R. J., Yakir, D., and Paterson, A. H. (2001). Genomic dissection of genotype × environment interactions conferring adaptation of cotton to arid conditions. *Genome Res.* 11, 1988–1995. doi: 10.1101/gr.157201

Sarfraz, Z., Iqbal, M. S., Geng, X., Iqbal, M., Nazir, M., Ahmed, H., et al. (2021). GWAS mediated elucidation of heterosis for metric traits in cotton (*Gossypium hirsutum* L.) across multiple environments. *Front. Plant Sci.* 12:565552. doi: 10.3389/fpls.2021.565552

Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. *Genetics* 8, 552–560. doi: 10.1093/genetics/8.6.552

Sebastian, S., Feng, L., and Kuhlman, L. (2012). Accelerated yield technology™: A platform for marker assisted selection of simple and complex traits. *Desig Soybeans 21st Century Markets* 2012, 297–305. doi: 10.1016/B978-0-9830791-0-1.50020-0

Sheikh, F. A., Dar, Z. A., Sofi, P. A., Lone, A. A., and Shiekh, N. A. (2017). Advanced breeding strategies to mitigate the threat of black stem rust of wheat. *Int. J. Curr. Microbiol. App. Sci.* 6, 1–20. doi: 10.20546/ijcmas.2017.606.001

Singh, B. D., and Singh, A. K. (2015). "Introduction to marker-assisted crop improvement," in *Marker-Assisted Plant Breeding: Principles and Practices.* eds. B. D. Singh and A. K. Singh (India: Springer), 3–16.

Sonah, H., Deshmukh, R. K., Singh, V. P., Gupta, D. K., Singh, N. K., and Sharma, T. R. (2011). Genomic resources in horticultural crops: status, utility and challenges. *Biotechnol. Adv.* 29, 199–209. doi: 10.1016/j.biotechadv.2010.11.002

Sun, Z., Li, H., Zhang, Y., Li, Z., Ke, H., Wu, L., et al. (2018). Identification of SNPs and candidate genes associated with salt tolerance at the seedling stage in cotton (*Gossypium hirsutum* L.). *Front. Plant Sci.* 9:1011. doi: 10.3389/fpls.2018.01011

Tang, S., Teng, Z., Zhai, T., Fang, X., Liu, F., Liu, D., et al. (2014). Construction of genetic map and QTL analysis of fiber quality traits for upland cotton (*Gossypium hirsutum* L.). *Euphytica* 201, 195–213. doi: 10.1007/s10681-014-1189-y

Tanksley, S. D., and Nelson, J. C. (1996). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* 92, 191–203. doi: 10.1007/BF00223376

Thoday, J. M. (1961). Location of polygenes. *Nature* 191, 368–370. doi: 10.1038/191368a0

Thyssen, G. N., Jenkins, J. N., McCarty, J. C., Zeng, L., Campbell, B. T., Delhom, C. D., et al. (2018). Whole genome sequencing of a MAGIC population identified genomic loci and candidate genes for major fiber quality traits in upland cotton (Gossypium hirsutum L.). *Theor. Appl. Genet.* 132, 989–999. doi: 10.1007/s00122-018-3254-8

van Ooijen, J. W. (1992). Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.* 84, 803–811. doi: 10.1007/BF00227388

Varshney, R. K., Mohan, S. M., Gaur, P. M., Gangarao, N. V. P. R., Pandey, M. K., Bohra, A., et al. (2013). Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. *Biotechnol. Adv.* 31, 1120–1134. doi: 10.1016/j.biotechadv.2013.01.001

Vieira, M.L.C., Santini, L., Diniz, A.L., and de Freitas, Carla Munhoz (2016): Microsatellite markers: what they mean and why they are so useful. *Genet. Mol. Biol.,* 39:312–328, doi: 10.1590/1678-4685-GMB-2016-0027.

Visscher, P. M., Haley, C. S., and Thompson, R. (1996). Marker-assisted introgression in backcross breeding programs. *Genetics* 144, 1923–1932. doi: 10.1093/genetics/144.4.1923

Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A., Knowles, D., Golan, D., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* 51, 592–599. doi: 10.1038/s41588-019-0385-z

Wan, Q., Wang, W., Hu, M.-C., Zhang, J., Liu, D.-J., Zheng, J., et al. (2009). Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Mol. Breed.* 24, 49–61. doi: 10.1007/s11032-009-9271-1

Wang, B., Guo, W., Zhu, X., Wu, Y., Huang, N., and Zhang, T. (2007). QTL mapping of yield and yield components for elite hybrid derived-RILs in upland cotton. *Genet. Genomics.* 34, 35–45. doi: 10.1016/S1673-8527(07)60005-8

Wang, X., Lu, X., Wang, J., Wang, D., Yin, Z., Fan, W., et al. (2016). Mining and ANALYSIS of SNP in response to salinity stress in upland cotton (*Gossypium hirsutum* L.). *PLoS One* 11:e0158142. doi: 10.1371/journal.pone.0158142

Wang, X., Xu, Y., Hu, Z., and Xu, C. (2018). Genomic selection methods for crop improvement: current status and prospects. *Crop J.* 6, 330–340. doi: 10.1016/j.cj.2018.03.001

Wang, F., Zhang, J., Chen, Y., Zhang, C., Gong, J., Song, Z., et al. (2019). Identification of candidate genes for key fibre-related QTLs and derivation of favourable alleles in *Gossypium hirsutum* recombinant inbred lines with *G. barbadense* introgressions. *Plant Biotechnol. J.* 18, 707–720. doi: 10.1111/pbi.13237

Wang, P., Zhu, Y., Song, X., Cao, Z., Ding, Y., Liu, B., et al. (2012). Inheritance of long staple fiber quality traits of *Gossypium barbadense* in G. hirsutum background using CSILs. *Theor. Appl. Genet.* 124, 1415–1428. doi: 10.1007/s00122-012-1797-7

Watson, I. A., and Singh, D. S. (1952). The future of rust resistant wheat in Australia. *J. Aust. Inst. Agric. Sci.* 18, 190–197.

Wendel, J. F., and Albert, V. A. (1992). Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Syst. Bot.* 17, 115–143. doi: 10.2307/2419069

Wendel, J., Brubaker, C., Alvarez, I., Cronn, R., and Stewart, J. M. (2009). "Evolution and natural history of the cotton genus," in *Genetics and Genomics of Cotton, Vol. 3.* ed. A. H. Paterson (New York: Springer), 3–22.

Wendel, J., and Cronn, R. (2003). Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78, 139–186. doi: 10.1016/S0065-2113(02)78004-8

Wendel, J., and Grover, C. (2015). "Taxonomy and evolution of the cotton genus, gossypium," in *Cotton, Vol. 57, 2nd Edn.* eds. D. D. Fang and R. G. Percy (Guilford Road, USA: Wiley, John Wiley & Sons, Inc), 25–44.

Wendel, J. F., and Percy, R. G. (1990). Allozyme diversity and introgression in the Galapagos Islands endemic *Gossypium darwinii* and its relationship to continental *G. barbadense.* *Biochem. Syst. Ecol.* 18, 517–528. doi: 10.1016/0305-1978(90)90123-W

Wendel, J. F., Rowley, R., and Stewart, J. M. (1994). Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Syst. Evol.* 192, 49–59. doi: 10.1007/BF00985907

Westengen, O. T., Huamán, Z., and Heun, M. (2005). Genetic diversity and geographic pattern in early south American cotton domestication. *Theor. Appl. Genet.* 110, 392–402. doi: 10.1007/s00122-004-1850-2

Williams, C. E., Wielgus, S. M., Haberlach, G. T., Guenther, C., Kim-Lee, H., and Helgeson, J. P. (1993). RFLP analysis of chromosomal segregation in progeny from an interspecific hexaploid somatic hybrid between *Solanum brevidens* and *Solanum tuberosum.* *Genetics* 135, 1167–1173. doi: 10.1093/genetics/135.4.1167

Witcombe, J. R., and Hash, C. T. (2000). Resistance gene deployment strategies in cereal hybrids using marker-assisted selection: gene pyramiding, three-way hybrids, and synthetic parent populations. *Euphytica* 112, 175–186. doi: 10.1023/A:1003836132603

Xu, Y., Li, P., Yang, Z., and Xu, C. (2017a). Genetic mapping of quantitative trait loci in crops. *Crop J.* 5, 175–184. doi: 10.1016/j.cj.2016.06.003

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., et al. (2017b). Enhancing genetic gain in the era of molecular breeding. *J. Exp. Bot.* 68, 2641–2666. doi: 10.1093/jxb/erx135

Xu, S., Pan, Z., Yin, F., Yang, Q., Lin, Z., Wen, T., et al. (2020). Identification of candidate genes controlling fiber quality traits in upland cotton through integration of meta-QTL, significant SNP and transcriptomic data. *J. Cotton Res.* 3, 34. doi: 10.1186/s42397-020-00075-z

Xu, S., Zhu, D., and Zhang, Q. (2014). Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Natl. Acad. Sci. U. S. A.* 111, 12456–12461. doi: 10.1073/pnas.1413750111

Yang, C., Guo, W., Li, G., Gao, F., Lin, S., and Zhang, T. (2008). QTLs mapping for Verticillium wilt resistance at seedling and maturity stages in *Gossypium barbadense* L. *Plant Sci.* 174, 290–298. doi: 10.1016/j.plantsci.2007.11.016

Yi, C., Guo, W., Zhu, X., Min, L., and Zhang, T. (2004). Pyramiding breeding by marker-assisted recurrent selection in upland cotton: selected effects on resistance to Helicoverpa armigera. *Agric. Sci. China* 3, 330–339.

Yin, J., Guo, W., Yang, L., Liu, L., and Zhang, T. (2006). Physical mapping of the Rf1 fertility-restoring gene to a 100 kb region in cotton. *Theor. Appl. Genet.* 112, 1318–1325. doi: 10.1007/s00122-006-0234-1

Yu, J. Z., Ulloa, M., Hoffman, S. M., Kohel, R. J., Pepper, A. E., Fang, D. D., et al. (2014). Mapping genomic loci for cotton plant architecture, yield components, and fiber properties in an interspecific (*Gossypium hirsutum* L. × *G. barbadense* L.) RIL population. *Mol. Gen. Genomics.* 289, 1347–1367. doi: 10.1007/s00438-014-0930-5

Yu, J., Zhang, K., Li, S., Yu, S., Zhai, H., Wu, M., et al. (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. *Theor. Appl. Genet.* 126, 275–287. doi: 10.1007/s00122-012-1980-x

Zan, W., Gao, F., Liu, H. F., Li, G. Y., Song, W., Luo, C., et al. (2008). Molecular mark of resistance to mapping of QTL. *Xinjiang Agric. Sci.* 45, 805–880.

Zhang, J. F. (2011). Twenty-five years of introgression breeding through interspecific hybridization between *Gossypium hirsutum* and *G. barbadense*. *Proc. Beltwide Cotton Conf.*, 711–716.

Zhang, J., Chiodini, R., Badr, A., and Zhang, G. (2011). The impact of next-generation sequencing on genomics. *J. Genet. Genomics* 38, 95–109. doi: 10.1016/j.jgg.2011.02.003

Zhang, Y.-M., and Gai, J. (2008). Methodologies for segregation analysis and QTL mapping in plants. *Genetica* 136:311. doi: 10.1007/s10709-008-9313-3

Zhang, J., Percy, R. G., and McCarty, J. C. (2014). Introgression genetics and breeding between upland and Pima cotton: a review. *Euphytica* 198, 1–12. doi: 10.1007/s10681-014-1094-4

Zhang, Y., Wang, L., Xin, H., Li, D., Ma, C., Ding, X., et al. (2013). Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol.* 13, 141. doi: 10.1186/1471-2229-13-141

Zhang, J., Yu, J., Pei, W., Li, X., Said, J., Song, M., et al. (2015). Genetic analysis of Verticillium wilt resistance in a backcross inbred line population and a meta-analysis of quantitative trait loci for disease resistance in cotton. *BMC Genomics* 16, 1–13. doi: 10.1186/s12864-015-1682-2

Zhang, J., Zhu, Y., Abdelraheem, A., Zhonghua, T., Thyssen, G., Fang, D., et al. (2019). "A genome-wide association study of Fusarium wilt resistance in a MAGIC population of Upland cotton." in *2019 Beltwide Cotton Conferences, New Orleans, LA*.

Zhao, C., Zhao, G., Geng, Z., Wang, Z., Wang, K., Liu, S., et al. (2018). Physical mapping and candidate gene prediction of fertility restorer gene of cytoplasmic male sterility in cotton. *BMC Genomics* 19, 1–11. doi: 10.1186/s12864-017-4406-y

Zhu, G., Gao, W., Song, X., Sun, F., Hou, S., Liu, N., et al. (2020). Genome-wide association reveals genetic variation of lint yield components under salty field conditions in cotton (*Gossypium hirsutum* L.). *BMC Plant Biol.* 20:23. doi: 10.1186/s12870-019-2187-y

Zhu, L., and Kuraparthy, V. (2014). Molecular genetic mapping of the major effect photoperiod response locus in Pima cotton (*Gossypium barbadense* L.). *Crop Sci.* 54, 2492–2498. doi: 10.2135/cropsci2014.03.0258