We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,400
Open access books available

117,000
International authors and editors

130M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Molecular Breeding of Cotton

Yuksel Bolek, Khezir Hayat, Adem Bardak and Muhammad Tehseen Azhar

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/64593

Abstract

Molecular characterization provides comprehensive information about the extent of genetic diversity, it assists for the development of an effective, highly accurate, and rapid marker-assisted cotton breeding program. Due to one of the world’s leading fiber crops, molecular studies of cotton are being explored widely by cotton researchers. Cotton provides raw material to the textile industry among other products. Limitations in conventional breeding program for genetic improvement are due to the complexity and limited knowledge on economically important traits. The use of molecular markers for the detection and exploitation of DNA polymorphism is one of the most significant developments in molecular genetics. In the present scenario, cotton molecular breeding has become a reliable source through the study and exploitation of its genetic diversity and due to better understanding of the cotton genomes using the next-generation sequencing technologies. Cotton breeders should utilize genomics in breeding programs for effective selection of best parents for agronomic and fiber-related traits, as well as for the development of resistance against biotic and abiotic stresses. The genomic research work could be based upon genotyping using DNA markers, quantitative trait loci mapping, genome-wide associations, and next-generation sequencing. The objective of this chapter is to describe evolution as well as utilization of various molecular markers and review the contribution of marker-assisted selection (MAS) in cotton breeding.

Keywords: cotton, DNA markers, genotyping by sequencing (GBS), genome-wide association studies (GWAS), marker-assisted selection (MAS)
1. Introduction

Plant breeders select those plants, which looks phenotypically more promising due to the presence of desirable traits. Most of the traits are controlled by polygenes with complex nonallelic quantitative effects and environmental interactions. In most cases, despite the fact that biometrical genetics reveals the presence of additive or non-additive effects on loci involved in the inheritance of quantitative trait, a specific locus may not be detected [1]. Tightly linked loci with desired trait can support plant breeding program by rapid introgression of quantitative trait loci (QTL) using associated molecular markers [2]. Genomic region having genes of interest for a particular trait is designated as QTL (Quantitative Trait Loc). QTL analysis involves partitioning of genetic variation in single component. So, DNA-based molecular markers provide a tool to plant breeders for the selection of desirable plants based on genotype instead of phenotype.

The expression of gene(s) individually their interaction with the climatic factors and agronomic measures can determine the cultivar adaptability [3]. Selection of new plant varieties with the desirable traits under given environmental conditions and cultural practices is the fundamental basis of plant breeding [4], genetic variability produced in germplasm as a result of selection, which alter the inheritance pattern of the traits, is quite useful to screen and select the cultivars for required traits. New cultivars have been developed by exploiting genotypes with enormous variation [5]. Rapid changes are needed in agricultural production, and biologically diverse as well as low-input novel farming systems must be developed and employed. There is also a need for new crop varieties that are (1) fitting-in to global climate change in the present era, (2) adapted to biodiverse farming systems, and finally (3) giving more products to farmers and eventually to consumers.

Cotton (*Gossypium* spp.) is one of the most intensively cultivated species grown in more than 80 countries in varying climatic conditions [6]. Globally, cotton is the ultimate source of fiber for industry and provides oil to diet [7]. Being utmost fiber manufacturing crop and the third contributor to oilseed production; China, India and USA are top contributors for fiber [8]. *Gossypium* genus is divided into eight genomes (A-G and K) and comprised of 45 diploids and five allotetraploid species which are found in the arid and semi-arid regions of Africa, Central and South America, Galapagos, Indian subcontinent, Australia, Arabia, and Hawaii [9–11].

At the beginning of the 20th century, scientists discovered that Mendelian factors controlling inheritance are organized in linear order on chromosomes. It was shown that genes could be inherited individually or in combination with other genes. The individual fragments flanking within a defined interval are known as molecular DNA markers [12]. Precise DNA portion with a known position on the chromosome [13], or a measurable trait that is associated with variation in DNA sequence [14, 15] or a difference may act as a genetic marker if it identifies characteristics of an individual.

Markers are broadly divided into three classes: (1) morphological markers, which themselves have phenotypic traits meaning the morphological and physiological features of plants are used to understand the genetic variation. Although morphological features may be indicative
of the phenotype, they are also highly affected by environmental factors and growth practices; (2) biochemical markers, including isozymes, which involve allelic variants of proteins/enzymes; (3) molecular markers, manifest mutations in heredity material such as DNAs and RNAs [16–19].

Polymorphism of molecular markers shows differentiation of homozygotes and heterozygotes [20]. Thottappilly et al. [21] refer to molecular markers as naturally occurring polymorphism, which include the proteins and nucleic acids that indicate certain differences. The use of molecular markers in plant breeding is called marker-assisted selection, often referred as MAS or marker-assisted breeding (MAB) (Figure 1) [4, 22].

![Figure 1. Marker-assisted scheme](http://dx.doi.org/10.5772/64593)

In traditional plant breeding, traits are selected depending on the phenotype, which is highly affected by the climatic factors. This approach makes the breeding a slow, expensive, and challenging process [23–25]. Practical advantages of using genetic markers, potential values of linkage maps, and exploiting for direct selection in plant breeding were begun to be studied about the 1930s [26]. Molecular markers are essential for mapping the genes of interest, MAS/MAB, and cloning of genes using mapping-based cloning strategies [27]. In addition, the use of molecular markers includes gene introgression through backcrossing, germplasm characterization, and phylogenetic analysis [28]. It has been observed that MAS is more efficient than conventional breeding techniques [4, 29, 30]. Selection based on genotypic structure through employment of molecular markers in the field crops [31] has laid the foundation of MAS [32,
Many biological and medical science applications and studies, including genetic diversity, molecular tagging of economic traits, and procurement of heritable diseases have successfully utilized molecular markers [2, 34–37]. Thus far, molecular markers have been exploited in rice [38], wheat [39], maize [40, 41], and barley [42, 43]. However, MAS has achieved the desired goals in cotton with limited success due to a genetic bottleneck through historic domestication and limited polymorphism in cultivar germplasm [44–47].

About 145 morphological markers are reported in cotton so far, but they have low utility in variety development because of incapability to assemble diverse markers in a genotype [48]. Isozymes produced through allelic variants are considered more authentic but not widely used due to their differential expression in different growth stages. For improving productivity and other key quantitative traits, cotton genetic markers have more value than morphological or isozyme markers [48]. DNA markers have become handy and effective tools for plant breeders because their expression is not necessary for their detection [49]. In order to enhance the benefits through molecular markers, vast developments have been made in ‘omics’, which, in turn, allowed the use of these markers in diverse ways for genetic studies instead of using them solely for phylogenetic studies [50]. Obtaining pure DNA plays a major role for the development of molecular markers in cotton [51–53]; genetic analysis has many drawbacks due to the presence of phenolic compounds, which affect quality of DNA and protein during tissues grinding [51].

Polygenic traits are mostly affected by the climatic conditions and show discrete variability after hybridization. Recombination frequency allows investigators to differentiate genes on linkage map by relative distance between a generation and their parents. The main hindrance for QTL mapping of agronomic traits is related to a large number of genes involved in phenotypic expression and their interaction with the environment [54]. As number of genes affects the trait phenotypically, it is desired that more loci should be evaluated for QTL determination, and the screening of individuals should be done at multiple locations/environments to maximize the use of QTLs. MAB uses QTLs to pyramid favorable alleles and break linkage groups for tagging QTLs of interest [55–57]. In recent years, conventional plant breeders started to use MAS for the identification of traits with high heritability such as disease resistance, as well as the yield of major field crops [57]. However, yield-related components have low heritability, which is a major challenge for the utilization of MAS [56, 58]. MAS is being employed for the identification of transgressive segregants. Transgressive segregation is the production of plants in F2 generation that are superior to both parents for one or more traits. Transgressive breeding aims at improving yield or contributing to yield-related traits through transgressive segregation [59–61]. Several QTLs have been identified for seed cotton yield, fiber quality, plant architecture, resistance to diseases such as bacterial blight and *Verticillium* wilt [57], resistance to pests like root knot nematode, and flowering date [62] as well as for abiotic stresses (drought, salt tolerance) [55, 63].

There is a gap between discovery of useful genes and QTLs, and their utilization in breeding programs. To date, few examples are reported [55, 63] for the successful release of genotypes developed by MAS, and they have shown significant contribution to yield improvement. High-throughput, high-density genome-profiling tools enable the rapid and low-cost of crop
genome in a precise and high-resolution manner. Identification of molecular variants in DNA sequence opens opportunities for plant scientists [55]. The potential exists in plant breeding for efficient use of next-generation sequencing (NGS) that also has revolutionized the plant genomics [55]. Markers can be analyzed across the genomes simply and accurately, with high-throughput. Increased number of next generation sequencing allows conducting genome-wide association studies (GWAS) [63]. It is thanks to the developments in knowledge of useful genetic diversity and QTLs, advances in sequencing, genotyping, and bioinformatics approaches that rapid, high-throughput molecular marker discovery methods have been enabled.

Day-by-day developments of new, specific markers, and trait determination tools makes molecular markers important in understanding the genomic variability and diversity within and among species. In this chapter, we discuss about the applications and types of molecular markers, next-generation sequencing, and role of molecular breeding in development of plants with improved economical traits in cotton.

2. Breeding for polygenic traits

Economically important traits such as nutritive value, earliness, agronomic traits, resistance etc. can be improved through MAS [64, 65]. Polygenic mapping allows breeders to estimate and assess the hereditary pattern of the traits governed by many genes found throughout genome; Ultimately it leads to efficient utilization of these traits for molecular breeding. Highly saturated genetic maps in a high population index permits to observe the impact of many regions of genomes on a single trait value. Paterson [66] revealed that sharing of homologues during crossing over is the basis of QTLs. The regions of the genomes connected to the traits of economic value are QTLs [67]. Association of a marker’s genotypic value to a phenotype is the basis of QTL mapping. Recombination frequency is used to evaluate the relative distance among markers in the linkage map. It is assumed that markers at or lower recombination ratio of 50% are considered as unlinked found either on homologues or alternative loci while the markers which are tightly connected will be transferred to offspring more often than the unlinked markers [67].

Reinisch et al. [68] developed the pioneer genetic map of cotton during 1994. Although large number linkage maps have been constructed since then due to abundance of several DNA markers, it is still needed to determine reliable QTLs from breeding perspective. Yu et al. [69] screened genotypes by simple sequence repeats (SSRs) to map loci connected to fiber quality and lint yield in a backcross inbred line and developed a pioneer genetic map using BIL within allotetraploid cultivated cotton species. Map consisted of 392 highly cosegregated loci covering 2895 cM length and having mean interlocus distance 7.4 cM. As a whole, 39 QTLs were directly connected to yield components and 28 were associated to fiber quality.

Altaf et al. [70] explored F2 population developed from three different species of Gossypium for identification of evolutionary relationship among these species by linkage map. Eleven linkage groups were constructed having 521.7 cM map size in cotton genome and relative distance of
16.8 cM was found among markers through screening randomly amplified polymorphic DNAs (RAPDs) and amplified fragment length polymorphisms (AFLPs). Jiang et al. [71] utilized F2 population developed from *G. hirsutum* × *G. barbadense*, and produced a restriction fragment length polymorphism (RFLP) genetic map having 3767 cM length; 27 linkage groups with distance of 14.4 cM among loci.

Shappley et al. [72] used F2:3 families derived from HS-46 and MARCABUCAAG-1-8896 genotypes and constructed genetic map by using 120 RFLPs which spanned to 865 cM and arranged in 31 linkage groups. Fifty one linked groups were developed through a map constructed with RFLP and RAPD markers [73] spanning to 6663 cM including 332 AFLPs, 91 RAPDs and three morphological markers. Khan et al. [74] studied comparison for ploidy level to diploid ancestors and tetraploid cotton with RAPD markers. 119 F2:3 families developed from MD5678ne × Prema and utilized RFLPs for genetic map. Seventeen linkage panels were distributed on 700.7 cM map having mean distance of 7–8 cM among the markers [75].

RFLP, AFLP, and SSRs were screened in a backcrossed breeding population derived from crosses of [(*G. hirsutum* cv. Guazunchoz × *G. barbadense* cv. VH8-4602) × *G. hirsutum* cv. Guazunchoz] [76]. Linkage map covered 4400 cM of genome and consist of 888 loci arranged on 26 and 11, long- and short-linkage groups, respectively. EST-SSRs from *G. arboreum* were used for linkage map construction in backcross inbred line [(TM1 × Hai7124) × TM1] [77]. Map spans to 5644.3 cM with mean interlocus distance of 9.0 cM. As a whole, 111 loci were detected with these 99 EST-SSRs incorporated into backbone map including 511 SSR loci. These EST-SSRs will be useful in MAS for improving fiber quality.

Mei et al. [78] developed interspecific population among *G. hirsutum* L. cv. Acala-44 and *G. barbadense* L. cv. Pima S-7 and published genetic map, which covers 3287 cM of the genome. They used AFLPs, RFLPs, and SSRs and have; identified total 392 loci being 333, 12, and 47 markers, respectively. They were able to identify high repetitive DNA and heterochromatin in D-genome and relative distance among mapped loci in A-genomes that were also compared to homologous in D-genome [79].

Two hundred and thirty-three linked loci were mapped in backcross population of [(*G. hirsutum* cv. Guazunchoz × *G. barbadense* VH8-4602) × *G. hirsutum* cv. Guazunchoz] by using 204 SSR markers, which produced 261 polymorphic bands [80]. Linkage map was published by adding 233 loci to already developed map [76] covering 5519 cM genome and having mean inter-related marker distance of 4.8 cM and consisting of 1160 loci. Nuygun et al. [80] applied STS markers for developing linkage maps that will fasten the genomics era by using diploid and tetraploid (AtDt) genomes. The genetic map consists of 2584 loci having Inter-locus marker distance of 1.72 cM and 763 loci intervals depending on 2007 probes from allotetraploid genome while 763 loci at relative marker distance of 1.96 cM intervals identified by 662 probes in D-genome. All desired homologous chromosome pairs were observed owing to locus repetition. Moreover, number of chromosomal variations including number of inversions and reciprocal translocations were observed.

Wang et al. [81] applied microsatellites for identifying QTLs related to fiber quality in RIL population. The genetic map was published with two common QTLs for lint percentage and
The results were in accordance to earlier studies and can be utilized in marker-assisted breeding. Lin et al. [82], screened SRAP, SSR, and RAPDs, have constructed linkage map, and a mean relevant distance was 9.08 cM among markers and total length of the map was 5141.8 cM. Park et al. [83] published the pioneer linkage map by applying EST-SSRs in RIL population derived from (G. hirsutum TM1 × G. barbadense Pima) for fiber. The linkage having about 27% genome coverage, covering 1277 cM genome and having 193 loci of those 121 newly mapped for fiber traits.

Researchers [84, 85] have used SSRs and AFLPs for determining oligonucleotides that is a good source for pyramiding of genes for marker-assisted selection. The mapping population developed by crossing parents having diversity for drought. Highly favorable environment was used; dryland and irrigated regimes for screening of genotypes. Quantitative trait loci mapped on different loci including one QTL (BNL1693) for seed cotton production on chromosomes 1 and 15 and two additional QTLs (BNL1153 and BNL2884) on chromosome 6. Moreover, chromosomes 6, 14, and 25 having BNL2884, BNL3259, and BNL1153 marker-associated QTLs found for osmotic pressure for drought in highly uniform lines. Researchers also revealed that NAU2715 and NAU2954 can be used as marker for relative water contents while relative water contents with NAU2954 will contribute a lot to drought tolerance in cotton. SSRs were analyzed for establishing genetic diversity and QTLs [86]. F$_2$ population of crosses (7235 × TM 1), (HS 427-10 × TM-1), and (PD 6992 × SM 3) utilized for assigning QTLs for fiber traits in the three different linkage maps which span to 666.7, 557.8, and 588 cM, respectively, with number of mapped loci with difference of 86, 56, and 73 [86].

He et al. [87] screened RAPDs, Retrotransposon-microsatellite amplified polymorphism (REMAP), SSRs, and sequence-related amplified polymorphisms in hybrids of G. hirsutum L. cv. Handan 208 and G. barbadense L. cv. Pima 90 for construction of linkage map. As a whole, 1029 loci were mapped on 26 chromosomes; map spans to 5472.3 cM with mean Inter-locus distance of 5.32 cM. Saleem et al. [85] determined two QTLs related to drought tolerance in F$_2$ progeny developed from diverse parents by applying SSRs and EST-SSRs. The progeny screened with parents for osmotic pressure using hydroponic culture.

Abdurakhmonov et al. [88] revealed that chromosomes 12, 18, 23, and 26 having QTLs controlling lint percentage by applying SSRs and EST-SSRs in a RIL population. Four QTLs for lint index, eight for seed index, 11 for lint yield, four for seed cotton yield, nine for number of seeds per boll, three for fiber strength, five for fiber length, and eight for fiber fineness were determined in F$_2$ population (G. hirsutum L. cv. Handan 208 × G. barbadense L. cv. Pima 90) [87].

SSRs were used to screen F$_2$ progeny for nematode resistance [89], and researchers identified gene “GB713” that control resistance, and could be used for reniform nematode resistance. They found two QTLs located on chromosome 21 having 168.6 cM on the genetic map while other QTL was located on chromosome 18. Morphological traits of RIL populations developed by hybridization of G. hirsutum and G. barbadense [90]. QTLs governing the plant architecture including plant height, number of primary and secondary branches were screened. Researchers found that angle of branch, angle of fruits, plant height, leaf size, main fruiting, etc. were governed by a single QTL. Infestation of disease is a severe problem in cotton, e.g., Xanthomonas
oxysporum [91], root knot nematode [89, 92–94], Verticillium [57], and cotton leaf curl disease (CLCuD) [95, 96] that warn the cotton scientist to find natural resistance sources and their urgent exploitation by using MAS.

3. DNA makers in cotton

Several types of molecular markers are available for characterization of germplasm of crop plants (Table 1). The amount of variation prevailing in the germplasm helps to maintain genetic conservation [98]. Availability of vast genomic database provides opportunity to develop enormous markers for detection of genetic variation [99, 100]. According to Weising et al. [101], these molecular markers must be (1) highly polymorphic, (2) codominant, (3) evenly distributed in a genome, (4) without pleiotropic effects, (5) easy to handle and fast assayed, (6) low cost and reproducible.

The cost of production of a marker is directly related to marker technique in use, polymorphic nature, and efficiency [102]. Polymorphic markers are divided into three types: (1) hybridization-based, (2) polymerase chain reaction (PCR) based, and (3) DNA sequence based markers [103].

3.1. Hybridization-based markers

Hybridization is occurred to the fragments of genomic DNAs produced by restriction endonucleases with various lengths among individuals. These types of markers are called “hybridization-based markers.”

3.1.1. Restriction fragment length polymorphism

Restriction fragment length polymorphism (RFLP) is a type of hybridization-based marker in plant genome and initially used for detection of polymorphism in a DNA sequence for gene mapping during the 1975s [31]. Nucleotide sequences of 4, 5, 6, or 8 bp, called restriction sites, are recognized by restriction endonucleases [104]. Digestion of DNA with restriction enzymes results in fragments whose number and size can vary among individuals, populations, and even within species.

Many scientists developed genetic mapping during the 1975s populations of cottons that were analyzed by using RFLP. Domestication of G. hirsutum was investigated with RFLPs [105]; they have revealed that Yucatan is the wild ancestor of upland cotton. Wright et al. [106] used RFLP for MAS and evaluated resistance allele for bacterial blight resistance. Hybridization carried out with probes for microsatellite sequences to yield a variable number of tandem repeats (VNTR) and allow oligonucleotide fingerprinting [107]. Joint map was constructed by using F_{2,3} populations derived from different intra-hirsutum accessions [108]. Two hundred and eighty-four polymorphic markers and 49 linked pairs were observed on the map. The genetic map spanned to 1502.6 cM having 5.3 cM distance between markers. RFLPs have played a
significant part for omics studies [109]. Low level of polymorphism, costly chemicals, and more
time for analysis limit RFLP use in MAS [104].

<table>
<thead>
<tr>
<th>S.N.</th>
<th>Feature</th>
<th>RFLP</th>
<th>RAPD</th>
<th>AFLP</th>
<th>SSRs</th>
<th>SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DNA require (µg)</td>
<td>10</td>
<td>0.02</td>
<td>0.5–1.0</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>PCR based data</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>DNA quality</td>
<td>High</td>
<td>High</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>4</td>
<td>No of polymorphic loci analyzed</td>
<td>1–3</td>
<td>1.50–50</td>
<td>20–100</td>
<td>1–3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Type of polymorphism</td>
<td>Single base change, insertion deletion</td>
<td>Single base change, insertion deletion</td>
<td>Single base change, insertion deletion</td>
<td>Change in repeat length</td>
<td>Single base change, insertion deletion</td>
</tr>
<tr>
<td>6</td>
<td>Dominance</td>
<td>Codominant</td>
<td>Dominant</td>
<td>Dominant/ codominant</td>
<td>Codominant</td>
<td>Codominant</td>
</tr>
<tr>
<td>7</td>
<td>Reproducibility</td>
<td>High</td>
<td>Unreliable</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>8</td>
<td>Ease of use and development</td>
<td>Not easy</td>
<td>Easy</td>
<td>Easy</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>9</td>
<td>Automation</td>
<td>Low</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>10</td>
<td>Cost per analysis</td>
<td>High</td>
<td>Low</td>
<td>Moderate</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>11</td>
<td>Developmental cost</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>12</td>
<td>Need for sequence data</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>Accuracy</td>
<td>Very high</td>
<td>Very low</td>
<td>Medium</td>
<td>High</td>
<td>Very high</td>
</tr>
<tr>
<td>14</td>
<td>Radioactivity detection</td>
<td>Usually yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>Genomic abundance</td>
<td>High</td>
<td>Very high</td>
<td>Very high</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>16</td>
<td>Part of genome surveyed</td>
<td>Low copy coding regions</td>
<td>Whole genome</td>
<td>Whole genome</td>
<td>Whole genome</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Level of polymorphism</td>
<td>Low</td>
<td>Low to moderate</td>
<td>Low to moderate</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>18</td>
<td>Inheritance</td>
<td>Codominant</td>
<td>Dominant</td>
<td>Dominant</td>
<td>Codominant</td>
<td>Codominant</td>
</tr>
<tr>
<td>19</td>
<td>Detection of alleles</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>Utility for genetic mapping</td>
<td>Species specific</td>
<td>Cross specific</td>
<td>Cross specific</td>
<td>Species specific</td>
<td>Species specific</td>
</tr>
<tr>
<td>21</td>
<td>Utility in marker-assisted selection</td>
<td>Moderate</td>
<td>Low to moderate</td>
<td>Low to moderate</td>
<td>High</td>
<td>Low to moderate</td>
</tr>
<tr>
<td>22</td>
<td>Cost and labor involved in generation</td>
<td>High</td>
<td>Low moderate</td>
<td>Low moderate</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

Table 1. Salient features of various molecular markers [97].
Ulloa et al. [110] published genetic maps by using intraspecific populations developed from parents having diverse genetic background. Fifteen linkage groups were used for designating the chromosomes. Earlier mapped data was used for construction of map by observing the deficiency analysis of the probes. QTLs were determined for fiber and yield traits by using this map. As a whole 63 QTLs were found in A subgenome at five different loci and 29 QTLs observed at 3-loci of D-subgenome. First genetic map spans to 117 cM produced 26 QTLs with 54 RFLPs while second map produced 19 QTLs with 27 RFLPs, and spanned to 77.6 cM. It was revealed that these maps will serve as map-based cloning for fiber quality.

3.2. PCR-based markers

PCR-based markers, i.e., RAPD [111–113], AFLP [114-116], microsatellites (SSRs) [117-119], and inter-simple sequence repeats (ISSRs) [120-121] represent major class of markers in cotton genomics due to their high utility and exploitation. Below are the major advantages of PCR techniques as compared to hybridization-based methods: (1) low amount of DNA used for genotyping; (2) capacity to amplify fragments from frozen cells; (3) high polymorphism that enables to generate many genetic markers within a short time; and (4) ability to screen many genes simultaneously either for direct collection of data or provide opportunity to collect information prior to submit for nucleotide sequencing [109].

The comparison of different aspects of generally used molecular markers is given in (Table 1) and brief description of these three classes of molecular markers is described below with reference to cotton genetics.

3.2.1. Amplified fragment length polymorphism

Amplified fragment length polymorphism (AFLP) relies on the restricted sequences and PCR amplification. Initially, genomic DNA is digested by a restriction enzyme and resulting fragments are ligated with adapters to both ends. Then, the adapter and restriction site sequences are selectively amplified; only the fragments whose ends are complementary to 3’ ends of selective primers are amplified resulting in small sequences. Finally, a gel is run for the separation of amplified fragments and it is visualized by fluorescence [34]. The focal point of this methodology relies on the magnification of endonuclease restricted fragments through PCR.

The important advantages of using of AFLP markers is that they exist in large numbers in genomes, they have a great reproducibility due to high PCR annealing temperatures, and less cost per marker basis [104]. In addition to reliability and reproducibility [116], there is no need of DNA sequence for analysis. In contrast to RFLPs and microsatellites, enormous polymorphic loci can be investigated by having single oligonucleotide pair running a single gel through AFLPs [122]. For digestion; partially degraded DNA and good quality DNA can be utilized, but care should be taken that isolated genomic DNA should be free of chemicals that interferes with polymerase chain reaction.

Lacape et al. [97] initially developed RILs population by introgression of Guazuncho 2 (G. hirsutum) and VH8-4602 (G. barbadense), and constructed a genetic map with 800 markers.
(AFLP, RFLP, and SSR) loci. AFLPs and RAPDs were used for development of linkage map in cotton [123]. Three hundred and seven SSR markers and 72 AFLP oligonucleotides were used for the development of genetic map in F$_2$ population which derived from intra-hirsutum hybridization. The map consisted of 27 linkage groups and it has 21, 72 cM distance between the markers [114]. Map saturation in various genotypes of cottons was analyzed [115].

AFLPs were screened in a backcross population developed from intra-hirsutum cultivars for agronomic traits and fiber quality enhancement [124]. They found 50 AFLPs associated with the fiber quality traits and few for other; further evaluated that E1M1-106, E1M4-153; E1M3-168, E6M3-266 for lint yield and lint percentage, respectively can be used in future for MAS [124]. AFLPs were used for introgression among *G. hirsutum* and *G. tomentosum* being close relative to Upland cotton [125]. Through analysis, species-specific [11, 16] AFLP markers were selected from *G. hirsutum* and *G. tomentosum*, respectively for assessing *G. hirsutum* relatedness. These species-specific AFLP markers would be useful for detecting gene flow between *G. hirsutum* and *G. tomentosum*.

Jixang et al. [124] revealed genetic diversity in a germplasm by using AFLPs. A range of 0.1–0.34 estimates of genetic diversity were found among the genotypes, and showed that genotypes having significant variation in the gene stock include AU 5367, Acala 1517-99, and LA 05307025.

### 3.2.2. Random amplified polymorphic DNA

Randomly amplified polymorphic DNA (RAPD) relies on use of short and random primers to amplify random portions of genome [126]. Such markers have found to be widespread in population genetic studies whose characterizations of genetic diversity and divergence within and among populations are based on assumptions of Hardy-Weinberg equilibrium and selective neutrality of the markers is employed [127]. Ultimate success of RAPDs is shown in the increase of molecular markers which require small amount of DNA and no need for sequencing, except of having all prerequisites for PCR conditions [126]. DNA fragments having sequence of about 10 bp are amplified with artificial primers by using PCR [128]. RAPDs are being used vigorously for profiling of genotype of important field crops, also for mapping for certain traits in addition to biotic and abiotic stresses. For such studies, RAPD primers show polymorphism and should be free from palindromic sequences and should have minimum 40% GC contents in the fragments [113].

Many scientists have explored RAPDs in cotton for studying different aspects like phylogenetic studies, genetic diversity, and CLCuV disease screening [111, 112, 128]. R-6592 and UBC607500 [113, 129] male sterility and fertility restorer traits can be improved by using RAPDs. Lan et al. [130] applied RAPDs for mapping fertility genes that is of immense value in cotton and tagged fertility restorer gene *R*-6592 which may be utilized for productivity enhancement. Lan et al. [130] conducted phylogenetic studies in cotton and argued that this procedure is helpful and reliable for introgression of desirable traits. RAPDs were used in cotton for comparing cotton cultivars resistance to jassids, mites, and aphids [131]. DNA finger printing, mapping and genetic diversity has been studied in cotton through RAPDs [132–134].
Noormohammadi et al. [135] screened F$_2$ population of Upland cotton and Opal variety by using 10 homo-primers and seven hetero-primers out of 26 RAPDs and found 261 reproducible bands, with an average of 4.18 [261 bands/17 primers = 15 bands/primer] bands per primer and 22% polymorphism for analyzing genetic resemblance in agronomic traits with 45 (Upland) and 80% (Opal) polymorphism, respectively. By applying agarose gel, multilocus genotyping can be carried out by staining with ethidium bromide and this facility is available in every lab working on molecular breeding [136]. RAPDs are often laboratory dependent and require immense care to design protocols for getting polymorphism. Several factors have been reported to influence the reproducibility of RAPD results such as quantity of template DNA, buffers of polymerase, concentration of magnesium chloride, primer to template ratio, annealing temperature, type or source of DNA polymerase, and brand of thermal cycler [137]. RAPDs also fail to discriminate between homozygotes and heterozygotes and complication of expressing Mendelian ratio of loci [138].

3.2.3. Intersimple sequence repeat

Modifications of microsatellites, which utilize microsatellites-complementary primers, overcome the need for flanking fragment information [139]. Polymorphism is revealed among simple sequence repeat (SSR) markers by using primer (16–25 bp) adjacent to a single SSR and annealing occur at either ends [139]. ISSR utilizes microsatellites as oligonucleotides in a PCR reaction to amplify inter simple sequence repeats for desired DNA. ISSRs utilize SSRs repeats dinucleotide, trinucleotide, and tetranucleotide as oligonucleotide [140]. Usually, ISSR primers have substantial fragments contrary to RAPD primers, enabling elevated annealing temperature, which produce highly polymorphic bands as compared to RAPDs [120, 141]. The amplified products can be separated by agarose and polyacrylamide gel due to longer length ranged from 200 to 2000 bp [139].

ISSR markers have been vastly used in cotton improvement, phylogenetic study and for mapping of germplasm [120, 121]. Parkihya et al. [142] studied genetic diversity among cotton genotypes by using nine ISSR oligonucleotides and detected 86 bands of which 54 bands exhibited polymorphism of 62.79% having mean of six bands per primer. The PIC ranged from 0.8616 to 0.9090 and genetic similarity ranged from 0.60 to 0.917. Phylogenetic relation was revealed in 21 cotton genotypes by using 12 intersimple sequence repeat primers and observed 49.6% reproducibility [143].

Genetic diversity was studied in cotton with 10 ISSR and showed 88.5% polymorphism [144]. Liu and Wendel [145] showed that ISSR can be designed with low cost. Genetic diversity observed in geneepool comprised of wild species and elite lines through SSRs and ISSR [146]. They observed 173 alleles having mean 3.93 alleles per locus by analyzing 39 SSRs and 5 ISSR markers which produced 89.6% reproducible bands. Among genotypes variation ranged from 0.04 to 0.58 while in diploid and tetraploid species it was 0.23–0.57%. Similar to RAPD, there may be some fragments with the same mobility originate from non-homologous regions [120].
3.2.4. Sequence characterized amplified region

RAPDs have more demerits of polymorphism as compared to other PCR-based markers which are used for analyzing a large number of individuals with low cost. This problem was overcome by using sequence characterized amplified region. PCR assay uses couple of distinct oligonucleotides for DNA sequence at a specific locus [147]; oligonucleotides might be having a high-copy, dispersed fragment within polymorphic loci. After sequencing the two ends of the two reproducible DNA fragments, one can develop two SCAR markers. SCAR 431vgo can be used in MAS program for screening genotypes with fiber strength. By using SCAR, codominance is produced [148].

These markers have been used in genetic analysis and used for molecular breeding [149, 150]. Extended sequence specificity of primers in SCARs results in higher reproducibility than RAPDs [151]. SCAR is widely used among researchers for mapping studies within closely related species [152]. SCARs are more authentic for MAS after conversion of DNA markers. SCAR markers are cost effective and highly polymorphic which make them suitable to be used for evaluating large number of mapping populations in cotton [152]. QTLs for leaf traits were observed [153].

3.2.5. Sequence-tagged site

Oslen et al. [154] developed sequence tag sites (STS) through observing impact of the PCR on human genome research, and argued that single-copy DNA sequences of known map location could serve as markers for genetic and physical mapping of genes along the chromosome. STS marker allows the utilization of PCR with specific primers which produces one oligonucleotide connected to the trait of interest. In order to utilize STS for molecular breeding, RFLP, AFLP, and RAPDs are usually converted into STS [155]. Thus, in a broad sense, STS include the markers such as microsatellites (SSRs), SCARs, and ISSRs mentioned above. Backcross breeding population was developed [(B416R × Ark8518) × Ark8518] and used for identification of STS markers related to fertility [155]. Tetraploid and diploid species were involved and artificial hybrids created by colchispolly.

RAPDs such as UBC1471400, UBC607500, UBC979700, and UBC169800 loci were associated to productivity restoration, and it was verified that UBC607500 is having enormous value for pyramiding genes to be used in molecular breeding [129]. Linkage maps were developed by using STS for diploid and tetraploid (A1, D1) Gossypium genomes [156]. Genetic map composed of 763 loci at 1.96 cM (approximately 500 kb) intervals detected by 662 probes (D), and 2584 loci at 1.72 cM (approximately 600 kb) intervals based on 2007 probes (A1, D1).

Several cotton breeders have used STS markers for identification of male restorer parental lines for hybrid cotton [129] who mapped cotton genotypes by using backcross inbred lines (BILs) and RIL populations with informative primers, and detected 21 and 7 polymorphic STS markers in BILs and RIL populations, respectively. Twelve STS markers were mapped in BIL population, and four of them were located along with resistance gene analog-amplified fragment length polymorphism (RGA-AFLP) markers on the same chromosome. Im-
Importantly, two were mapped on chromosome 4, flanking two main-effect QTLs, which were previously detected. These STS markers should be useful for high-throughput genotyping, gene mapping, and MAS for disease resistance including *Verticillium* wilt resistance in cotton.

3.2.6. Simple sequence repeats

Tandem repeats composed of several to over hundred repeats of one to four nucleotide motifs are found in all eukaryotic genomes. These repeats are designated as (AAAC)_n, here “n” represents number of tandem repeats. The flanking sequences of simple sequence repeats (SSRs) are used for the development of oligonucleotides [118]. Tandem repeats induce variability, which evolve polymorphism of different size due to slipped strand arise because of mispairing occurs during DNA replication [118], variation in size of PCR amplification/products induce polymorphism which can be separated by electrophoresis.

Kinship studies are conducted by employing SSR markers assess the extent of variation [119]. Vos et al. [157] used agarose and polyacrylamide gel for the identification of SSRs having codominance nature like AFLP. Akkaya et al. [158] stated that genetic mapping is on fast track due to the use of SSRs in self-pollinated crops where these markers are of great interest for breeders [159, 160]. SSRs are mostly codominant, and are indeed excellent for studying of population genetics and mapping [161–163]. The use of fluorescent primers in combination with automatic capillary or gel-based DNA sequencers has got its way in most advanced laboratories, and SSRs are also shown to be excellent markers for fluorescent techniques and high-throughput analysis.

Derived from trispecies hybridization that can be segregated for natural leaf defoliation trait. This RIL population screened with microsatellite markers, JESPR-13, JESPR-153, and JESPR178 tandem repeats were found to be highly associated to leaf defoliation trait value [162]. It was found that JESPR178 is closely linked to this trait in cotton. It has an immense importance that gene pyramiding can be accomplished for molecular breeding [164]. QTLs were tagged using SSRs in the nematode resistance RIL population developed via introgression from *G. barbadense* [165]. In that study, a single marker analysis identified four major QTLs located on chromosomes 3, 4, 11, and 17 were identified to account for 8.0–12.3% of the phenotypic variance.

Fiber length was increased up to 12–20% in cotton by using microsatellites in a population derived from interspecific hybridization and loci were discovered for marker-assisted selection [166]. Twenty-three chromosomes were analyzed by SSRs and found on an average relative distance of 4.9 cM [167]. Researchers [168–170] have utilized SSR markers for studying genetic diversity in cotton and observed limited genetic variations. Reddy et al. [171] used SSR-enriched genomic libraries and identified 300 SSR markers. Multinational Seed Company has reported more than 1200 SSRs [172].

Abdurakhmonov et al. [173] conducted genome-wide association mapping based on linkage disequilibrium (LD), scanning Upland germplasm consisting of 334 *G. hirsutum* accessions collected from Uzbek, Latin American, and Australian ecotypes. Screening under different
climates and by using mixed linear model involving population kinship and population structure 12–22 SSR markers were associated with fiber length, fiber strength, fiber fineness and six other fiber quality traits. Mei et al. [78] have reported 145 SSR polymorphic markers for yield and yield components in cotton by screening germplasm of 358 upland cotton varieties. Cao et al. [174] assessed genetic stock fiber quality and reported 97 polymorphic SSR markers by using LD. 

Bolek et al. [57] used SSR markers for verticilium resistance in cotton by using F₂ population; 255 SSRs were screened over bulks constituted by 10 resistant, and 10 susceptible progenies. QTLs were tagged by using 60 polymorphic markers. Genetic map produced 11 linkage groups having 15.17 cM inter-locus distance and spanning 331 cM.

Backcross inbred lines used [175] for observing the genetic variation of 446 SSR markers having relative mean distance of 10 cM interspecific linkage map and also detected 58 QTLs related to fiber quality and yield components. By using SSRs, genetic markers associated to cotton earliness were determined in progeny developed from intra-hirsutum hybrids [117], and these markers correspond to bud to flower duration and flower to boll period.

Earliness in cotton can be induced by the introgression of QTLs located near to the SSR markers such as BNL1044, DPL0209, NAU1004a, NAU 5046, NAU6078, and TMB0481 [117]. F₂ progeny was developed within G. hirsutum for utilizing gene pyramiding in marker-assisted breeding to enhance fiber quality and agronomic traits of economical value by using SSRs, SRAPs, EST-SSRs, and SSCP-SNPs [176].

Textile industry entirely depends upon fiber with good quality. Marker-assisted selection allows developing a cultivar having good fiber quality. There are many SSRs which can be used for fostering the breeding program; for example, lint percentage can be approved by using TMB0471 and MGHES-31, TMB0366, BNL3590, BNL1395, BNL1672, BNL1694, JESPER101, JESPR204, NAU3308, BNL1672, NAU3308; NAU4024 [168, 177–180]. Genetic base can be broadened for span length by using BNL 1395, DC40182, NAU2980, BNL2752, NAU2985, NAU1167, NAU1200, NAU2277 [50, 123, 177, 179]. BNL1122, BNL1317, BNL3145, BNL1521, CIR307, CGR6164, CGR6683, GH454, BNL3463, JESPER153, DC40182, NAU 1037, SHIN-0463, TEMB1618, NAU3736, NAU445, NAU780, NAU1102, NAU1197, NAU1322, NAU1369, JESPR218, TMD05 can be applied for fiber strength enhancement [52, 92, 173, 174, 176, 181–188].

3.2.7. Single nucleotide polymorphism

Single nucleotide polymorphisms (SNPs) manifest alteration in single base. SNPs are the most frequent occurring variability in the individuals which is found in each 1000 bases [189]. These are changes in bases from transitions (C/T or G/A) to transversions (C/G, A/T, C/A, or T/G) while insertions and deletions also induce SNPs which show single base changes. SNPs show useful allelic variations and have been markers of choice in various genetic studies [190]. Rapid progress in high-through put sequencing has allowed discovering SNPs in complex genomes.
with economic value by using genotyping by sequencing [191]. Frelichowski et al. [167] revealed that reproducibility is a major hindrance for using large number of markers developed in *G. arboreum*, *G. raimondii* and *G. hirsutum* [192]. In combination with genome and expressed sequence tags (ESTs) in model plant species [193], the efficiency of Sanger sequencing has been improved to accelerate the identification of variations at single base pair resolution [194].

Genotyping in plant sciences is progressing rapidly because SNPs for observing variation in a specific locus are utilized. Moreover, availability of enormous SNPs due to insertions-deletions and whole genome genomic studies is laying the corner stone for next generation sequencing [195]. Developed genomic databases and SNPs information allow evolving SNPs to an influential research for related relatives. Owing to most common type of DNA polymorphism, SNPs are also flexible in the selection of SNP variants at target loci, and they provide the option to choose from a large number of genome-wide loci when selecting sets of informative markers for specific germplasm pools [196]. Breeding programs comprised of genomic estimated breeding values are highly favored for whole genome techniques additional to targeted loci [197–199].

Economically important traits from breeding perspective are also investigated through genome-wide sequencing [200], saturated mapping of polygenic traits [201], and by using LD-mapping [202]. An et al. [203] studied the expression of R2R3-MYB transcription factors where few are expressed during fiber initiation and elongation. They observed phylogenetic relation among R2R3-MYB genes and published a map by using SNPs in Upland cotton. QTLs were mapped in population derived from intra-hirsutum and interspecific (*G. hirsutum* × *G. barbadense*) [204]. Researchers [204] have collected all published QTL data. QTLs were identified for seed, yield and fiber quality by using two populations through meta-analysis. QTLs connected to biotic and abiotic stressed were also detected. However, the development of high-throughput genotyping platforms for large numbers (thousands to millions) of SNPs has proved to be relatively time-consuming and costly.

Deynze et al. [205] reported more than 200 loci in *G. hirsutum* breeding germplasm, which were genetically mapped on mapping population derived from TM-1 and 3-79. Genepool comprised of 24-acccessions derived from 8-parental lines of mapping populations of six cotton species and 16 promising cotton strains used for genotyping. As a whole more than 1000 SNPs were polymorphic among *G. hirsutum* and *G. barbadense* were developed from 270 loci and 290 indels from 92 loci. Roche 454 pyrosequencing platform in four allotetraploid cottons using reduced representation library (RRL) have helped to map a large number of SNPs [206]. The conversion rate of SNPs using KASPar assay was about 35.8%. Three hundred and sixty-seven SNP markers were used for linkage map construction, which spanned to 1688 cm. High-resolution maps can be formed rapidly by utilizing parallel sequencing methods to determine the reads in resequencing. Pacific Biosciences technologies used for long reads [207] while Illumina and Ion Torrent are applied in sequencing for getting short reads [196].
4. Mapping populations

The group of plants, which is used for screening of molecular markers and segregated for the trait of interest, is designated as “mapping population.” From commercial point of view, such populations are developed from within species and can be also developed between different species for creating desirable variation. Polymorphism is compulsory in the progenitors for required trait [208]. The exchange of chromosome fragments during crossing over produces recombination, which provides the basis for developing linkage maps [59]. Populations are required for creating genetic maps in order to locate the quantitative trait loci from economical point of view.

Mapping populations can be exemplified by F2, backcross, recombinant inbred lines (RILs), doubled haploid lines (DHL), F2-derived F3 (F2:F3) populations, and near-isogenic lines (NILs). F1 is produced by selfing two parents, having extreme properties for trait of interest that show a significant polymorphism for whichever type of loci are scored. Mostly, this population is used for genetic mapping as it requires less time for development. However, there are some drawbacks for this population, most important of which is the fact that it is not stable. Qualitative and quantitative traits in cotton have been mapped by using F2 [70, 74, 81, 86].

Backcross (BC) population is developed by crossing a genotype with an elite cultivar, which is deficient for a single gene or QTL [67]. A concept of BC population was developed in 1922 and widely applied in plant breeding programs till 1960 [209]. Backcross population has been used for linkage mapping in cotton for improving various traits [129, 155–156].

Near isogenic lines (NILs) can be developed either by using selfing until purity is achieved; for all traits with wide variation of the trait of interest among NILs or by hybridizing the donor parent to the F1 plants and choosing the desired trait [63]. NILs are of high importance for genetic studies as they are stable like RILs. Researchers [210] used NILs for observing QTLs related to yield and drought related traits. They evaluated that NILs can be used for evaluating drought and can be used for MAS. Essenberg et al. [116] developed NILs in cotton and mapped bacterial blight resistance. They revealed that lines having Acala-44 in their parentage are showing dominance to bacterial blight.

Recombinant inbred lines (RIL) are stable and are developed by using single seed descent method from first filial generation. It continues until homozygosity is obtained in the individuals. RILs are permanent and can be screened at multiple locations for desired traits. Each strain is homozygous and stable in the RIL population. Each cycle of selfing results in enhanced recombination frequency and these populations are highly suitable for saturated mapping [129]. Moreover, for genetic mapping in cotton this population has been utilized for various traits including nematode resistance [165], fiber quality improvement [166], and verticillium resistance [175]. One of the drawbacks of this population is long duration for development in which segregation bias can occur due to removal of some genotypes after selfing. Another disadvantage of using these populations is that major QTLs are having a masking effect, and multiple QTLs are having epistatic effects.
Nested association mapping (NAM) [188] is designed for precise identification of QTLs [177]. Economically valued traits related to yield and subsequently to textile sector can be efficiently studied through developing new populations like NAM. NAM populations potentially address the limitations of conventional mapping populations.

5. Some applications of markers in breeding schemes

5.1. Marker-assisted backcrossing

The simplest, most widely used, and the most efficient form of MAS is MAB. In this form, two parents are used for the development; one is “donor parent” having trait of interest for transferring the targeted gene/loci and the other is “recipient parent” which is lacking gene. Parents are hybridized and $F_1$ is developed. Marker-assisted backcrossing relies upon the presence of a molecular marker associated with the trait, instead of targeting the expression of phenotypic value in traditional breeding. $F_1$ is planted for confirming the marker loci at initial stages of development, and pure $F_1$ is hybridized to recurrent parent. Markers are evaluated among individuals at the initial development stages of $BCF_1$ and hybridized to recurrent parent having alternative alleles. $BC_1$ individuals show segregation frequency of $F_1$ population gametes as two genotypes are involved in this population. Highly efficient map is constructed by using this population in contrast to $F_2$ population. This population is mostly used for overcoming hybrid in viability and hybrid breakdown in interspecific crosses [129]. This process is continued until three to four filial generation for stabilizing the marker and its associated trait of interest. MAB population has been utilized for observing traits of interest through quantitative trait loci [115].

5.2. Pedigree selection

Breeding techniques within the two cultivated tetraploid species rely on crossing and selection of traits using pedigree and recurrent selection methods. Promising genotypes having desirable traits can be developed using MAS and can be combined into a pedigree-based selection. Mostly, the efficiency of MAS was investigated using two populations from pedigree selection, and modified backcrossing pyramiding has been developed [211]. The selection efficiency for the fiber strength was greatly increased when QTLs-1 was selected simultaneously with two molecular markers with known genetic distance [211].

5.3. Marker-assisted recurrent selection (MARS)

Molecular markers should be applied for plant improvement in conjunction with the latest breeding methodologies. Marker-assisted recurrent selection offers an opportunity to get maximum output from a recurrent selection [212]; and it is used for introgression of multiple genes.

Quantitative traits can be enhanced efficiently by using MARS, which allows selfing and genotyping within same cropping season in one cycle of selection. The increase in genetic
gain was doubled from MARS in some populations as compared to phenotypic selection [213]. In cotton, resistance to American bollworm was achieved by using marker-assisted recurrent selection; they revealed highly significant differences in individuals studied by MARS for this insect resistance [214].

6. Gene pyramiding for MAS

Gene pyramiding has been widely used for combining number of genes especially disease resistance genes for specific races of a pathogen. Vertical resistance for different strains of pathogens is done by involving multiple strains. It is also done by “molecular breeding” because breeding for resistance is extremely difficult to achieve using conventional methods. Gutiérrez et al. [215] have used this technique for nematode resistance in cotton while [155] they applied sequence tag sites and screened STS markers associated with fertility restoring genes in cotton.

7. Next-generation sequencing (NGS)

SNP genotyping with latest high-throughput sequencing has the potential to speed up the breeding programs [216]. New DNA sequencing technologies have made it possible for the breeders and investigators to perform a genome analysis not only more rapidly but also less expensively [179]. High-throughput bioinformatics assist to identify large number of nucleotides per run [217]. Researchers have developed a lot of NGS methods with success in diverse platforms, which include Roche 454 FLX Titanium [218–220], Illumina MiSeq and HiSeq2500 [221], Ion Torrent PGM [222]. Genomic research contributes immensely to plant and animal sciences thanks to the advances in sequencing techniques [180, 221, 223–228]. The ultimate aim of all these techniques is to discover an authentic marker that could be used for sequencing in MAS with economical benefits [229–230].

Polyploidy is the main hindrance for isolation of useful SNPs in cotton because it produces homeologous and paralogous sequence variants which are combined together in allelic variations among cultivars [231–232]. Two cultivated tetraploids species were screened for the development of genomic SNPs through NGS by using reduced representation library obtained from Roche 454 pyrosequencing [206]. Competitive allele-specific PCR (KASPar) showed 35.8% validity of SNPs and developed the genetic map of *G. hirsutum* via 367 SNP markers which spanned to 1688 cM.

Gore et al. [233] developed a linkage map in a RIL population derived from intra-hirsutum cv. TM-1, and NM24016. The genetic map covered about 50% of the *G. hirsutum* genome which consisted of 429 SSRs and 412 SNPs. They also tagged 10 QTLs related to fiber quality, which provided a unique resource for mapping. Before Affymetrix became commercially available, Gene Chip cotton genome array consisting of 239,777 probe sets that represent 21,485 cotton transcripts has been developed [234, 235]. Sequences from Genome database, dbEST and
RefSeq were used for the development of Chip which promises to be an excellent source for genomics.

8. Genotyping by sequencing

In agricultural sciences, the discovery of reliable and true SNPs is compulsory for knowing about the utilization and importance of particular sequences. Molecular breeding tools can be applied to explore germplasm without available genomic data through genotyping by sequencing (GBS) methodology. GBS permits researchers to analyze complex genomes of polyploid species efficiently at low cost and it has been widely used due to the latest developments in high-throughput sequencing [191]. Reduced representative libraries are developed by using endonucleases [55, 178, 236]. Single nucleotide polymorphism is discovered for genomic studies [237]. Genomic techniques; genome-wide association study (GWAS), genomic diversity, genetic linkage analysis, molecular marker discovery offer to screen genotypes upon genotypic basis for traits of interest through GBS. Genotyping and reproducibility of markers are performed in a single step through GBS and SNPs are developed [238].

GBS-based sequencing data are used for developing genetic map and tagging markers with quantitative traits in populations derived from different ways, i.e., filial generations, RILs, etc. and germplasm collections [218]. GBS approach has been used efficiently for genetic analysis and marker development of rapeseed, lupin, lettuce, switchgrass, soybean, maize, and cotton [38, 219, 222–224, 233, 239].

The merits of GBS over existing marker development methodologies include availability of large number of markers, fast screening of populations composed of more number of individuals, diverse genotyping systems to tackle multiple traits, and more precise SNPs discovery and validity due to availability of high-throughput sequencing data [216]. Recently, GBS approach has been used to identify SNPs in the collections of RILs of wheat and to map various traits useful for breeding programs [55]. It is needed that efforts should be made to develop strategies for getting the benefits of NGS and advanced genotyping from breeder perspectives [196]. GBS protocol of Poland et al. [236] likely is needed to maximize the cost-effective concurrent discovery and genotyping of SNPs within cotton populations.

Although very efficient and productive in terms of achieving the desired goals, there are some drawbacks in GBS as well. GBS has incapability to assign true alleles of each locus in polyploids as compared to other techniques. As exemplified, Huang et al. [178] used RILs and biparental populations for assessing the utility of GBS in hexaploid oat. They observed that data analysis algorithm factors involved in SNP discovery, developed GBS derived loci description by forming two bioinformatics workflow. Its genetic map spans to 45,117 loci, which will be a source of further genetic studies [178].

Islam et al. [240] used GBS with two different approaches in cultivated cotton germplasm consisting of 11 diverse cultivars and their random-mated RILs. Authors have discovered a large set of polymorphic SNPs with broad applicability. They identified 4441 and 1176
polymorphic SNPs with minor allele frequency of ≥0.1. The utility of developed SNP markers were confirmed using SNPs in 154 Upland cotton accessions with high genetic diversity.

9. Association mapping

Genome-wide association study is used for developing highly saturated maps in cotton germplasm [241]. This technique allows detecting association among various markers and traits through assessing of the genetic diversity of required traits [242]. Linkage disequilibrium-based mapping (LD-mapping) is the advanced tool to study complex traits governed by many genes. LD-mapping has been successfully used in self-pollinated plants [243]. Microsatellites were screened in germplasm consisting of varieties at different locations to tag yield and fiber quality QTLs [202]. QTLs mapped for yield and fiber quality traits will serve as a reliable source to determine the diversity within the species and will contribute a lot in MAS [202]. In contrast to biparental populations, association mapping fosters molecular breeding because a vast genetic diversity is present in germplasms due to diverse sources [173]. Several protocols have been developed including complexity reduction of polymorphic sequences (CRoPS) [244], restriction site associated DNA (RAD) [216], GBS [195], and sequence-based genotyping (SBG) [245, 246] for genome analysis. Of all protocols, LD-mapping is on the top thanks to innovations done for high resolution. Association mapping is an authentic way for molecular tagging as it allows the screening of quantitative traits of value in a precise way [247]. Genome-wide association makes it possible to detect association among various markers and traits.

Abdurakhmonov et al. [46] used LD-mapping in a germplasm collection, which included photoperiodic lines. Simple sequence repeats were used for assessing the extent of LD in cotton and the major fiber quality QTLs were tagged using mixed linear model.

Nested association mapping is also being used for identification of suitable SSRs in a NAM population derived from 20 diverse genotypes of G. hirsutum with Namangan-77. SSR marker screening for development of highly saturated map through NAM F_{2:3} populations for traits of immense value in cotton is underway [188].

10. Public data resources

Sequenced genomic information allows breeders to analyze the genetic variation [248]. Major databases, which serve as a foundation, include CottonGen [58], Comparative Evolutionary Genomics of Cotton [249], National Center for Biotechnology Information [250] for Express sequence tags resource, TropGENE Database [251], the Cotton Diversity Database [252] and BACMan resources at Plant Genome Mapping Laboratory [253]. These resources provide genomic and heredity data of the cotton germplasm, QTLs tagged to loci and highly saturated linkage maps.
11. Targeting-induced local lesions in genomes (TILLINGS)

Phenotypic variation in plant genomes is produced by variation in DNA bases, which can be induced naturally and/or using different chemicals \[254\]. The targeting-induced local lesions in genomes (TILLING) technique allows determining an allelic variation precisely in a single-base pair for the targeted gene. Chemical treatments have been applied to generate SNP mutations. Point mutations, which are useful from breeder’s perspective, can be detected by TILLING and ECOTILLING techniques \[255\]. The mutagens used for induction of point mutation are highly selective and optimal concentration can spontaneously produce single base alternations at a high frequency in TILLING.

Knock-out population is developed by treating the seed with chemicals, inducing change in DNA sequence \[256\]. Auld et al. \[257\] used TILLING in \textit{G. arboreum} and demonstrated the applicability of this technique in cotton. The ultimate success to produce large number of sequence variations of target genome depends upon duration of application, relative capability of ethyl methane sulfonate (EMS), and γ-rays \[53\]. Aslam et al. \[53\] screened three \textit{Gossypium} sp. (\textit{G. hirsutum}, \textit{G. barbadense} and \textit{G. arboreum}) and constructed a kill curve. They observed the impact of different mutagens (EMS and γ-rays) consisting of eight different concentrations of EMS (0.1–0.8\%) and two levels of γ-rays (100–800 Gy). The genotypes of each species were evaluated with morphological parameters emergence and plant height, and yield traits (number of bolls per plant, boll weight, lint yield and lint percentage). For reverse and forward genetics point of view, viable accessions were selected from mutagenized genotypes. They revealed that EMS showed significantly high mutation rate than γ-rays.

There are many software tools which help to observe the bases variation; for instance, the method that determines whether a change occurs in an amino acid hampering codon is named conservation-based SIFT (sorting intolerant from tolerant) \[258\]. Taylor \[259\] described that any alternation of a gene can be detected by PARSESNP (for Project Aligned Related Sequences and Evaluate SNPs \[260\]; graphs show the changes in sequence by using precise co-segregating information, positioning of coding/noncoding regions and reference DNA sequence.

12. Conclusions

Developing reliable markers, which will work in different populations and utilized in the breeding to enhance selection efficiency, is a very important step for breeding. Markers should allow desired genotype selection because of their tight linkage to the trait of interest. On the other hand, emerging technologies like high-throughput marker systems and marker-based selection methodologies have been developed, and are currently being used efficiently in cotton breeding. It is also promising that some economically important traits like fiber quality, yield, \textit{Verticillium} wilt resistance, cotton leaf curl virus, drought tolerance, nematode resistance can be enhanced by using MAS. Genetic diversity can also be evaluated by using DNA markers before starting breeding program. Tremendous efforts have been carried for studying genetic diversity from genotypic and phenotypic aspects in germplasm accessions of cotton. Many
QTLs related to economical traits have been discovered. It is an emerging concern that efforts should be made for the utilization of molecular breeding methodologies to enhance cotton productivity, which can be enhanced through the recent developments in NGS. Moreover, highly saturated maps are useful for determining genetic manipulations from heredity perspectives, and SNPs are the best for this purpose. These markers along with QTLs provide innovative tools in the cotton genomics era.

Author details

Yuksel Bolek¹, Khezir Hayat¹, Adem Bardak¹* and Muhammad Tehseen Azhar²

*Address all correspondence to: adembardak@ksu.edu.tr

1 Faculty of Agriculture, Agricultural Biotechnology Department, Kahramanmaras Sutcu Imam University, Kahramanmaras, Turkey

2 Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan

References


[47] Zehr UB, editor, Cotton, Biotechnology in Agriculture and Forestry. Heidelberg: Springer-Verlag; 2012. DOI: 10.1007/978-3-642-04796-1_5.


[83] Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly D, Kohel RJ, El-Shiny OM, Cantrell RG. Genetic mapping of new cotton fibre loci using EST-derived micro-
satellites in an interspecific recombinant inbred (RIL) cotton population. Molecular Genetics and Genomics, 2005; 274: 428–441. DOI: 10.1007/s00438-005-0037-0.


[93] Wang C, Ulloa M, Roberts PA. Identification and mapping of microsatellite markers linked to a root-knot nematode resistance gene (rkn1) in Acala NemX cotton (Gossypium
hirsutum L.), Theoretical and Applied Genetics, 2006; 112: 770–777. DOI: 10.1007/s00122-005-0183-0.


[121] Bornet B, Muller C, Paulus F, Branchard ML. Highly informative nature of inter simple sequence repeat (ISSR) sequences amplified using tri and tetra-nucleotide primers from DNA of cauliflower (Brassica oleracea var. botrytis L.). Genome, 2002; 45: 890–896. DOI: 10.1139/g02-061.


Pathak D, Bhatia D. Development of advanced mapping populations in cotton. ICAC 2015; XXXIII.


[200] Nicolae DL, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genetics, 2010; 6, e1000888. DOI: 10.1371/journal.pgen.1000888.


[212] Jiang GL, Dong Y, Shi J, Ward RW. QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ9306. II. Resistance to deoxylevinaeol accumulation


